TR-25

SURFACE FITTING OF GROUND WATER TABLE BY MEANS OF LEAST SQUARE APPROACH

SATISH CHANDRA
DIRECTOR

STUDY GROUP

P V SEETHAPATHI

S K SINGH

NATIONAL INSTITUTE OF HYDROLOGY
JAL VIGYAN BHAWAN
ROORKEE-247667(UP)
INDIA

1986-87

CONTENTS

LIST OF TABLES

ABSTRACT

With the advent of high speed computers, the model studies for the analysis of complex groundwater flow problems, have become a common practice. For carrying out such model studies, the spacial and temporal variation of water table levels and aquifer parameters, in a systematic pattern of space coordinate is imperative as an input to the model which are amenable to numerical differentiation. But the water table elevations and aquifer parameters are measured only at few points in space, where an observation well is located. Thus the available historical record of water-table and aquifer parameters are point values from which the spatial distribution in a systematic pattern is required to be assessed. Using the spatial variation of these values the interpolation of water-table levels at nodal points and average aquifer parameters values for a finite difference grid is to be assessed for conducting model studies.

This report deals with such interpolation of water table levels at nodal points, using least square technique to fit a trendsurface or regression on surface coordinates approximating the water table levels. For such surface fitting of appropriate degree, computer program has been developed and a simple test problem was solved.

1.0 INTRODUCTION

The development of irrigation facilities both by surface and subsurface water and construction of major water projects have affected groundwater regime. The excessive rise in groundwater table may have detrimental effects not only to such irrigation projects but may even endanger the existence of the civilization. Hence, it is necessary to have optimal planning of Water Resources in order to maintain the watertable within permissible limits even under critical circumstances. For complex groundwater systems, numerical models are being employed with the advent of fast computers. The aquifer response model enables quantitative determination of the pumping or artificial recharge required to maintain a predicted water table elevation. The spacial and temporal distribution of piezometric levels can be obtained using a spacially distributed models, but, their applicability to physical problems is subjected to problem of extensive data requirement. The primary requirement of data for such models is the watertable level at each nodel points of finite-difference /finite element grid into which space discretization is made. To be more general, for carrying out such model studies of groundwater system the spacial and temperal variation of water-table levels and the aquifer parameters, for a systematic pattern of space coordinates (grid system) is imperative. But normally, the water-table levels and aquifer parameters are measured only at few points in space, where an observation well is situated or test pumping has been done. Thus, the interpolation of these values is necessary at nodal points of a finite-defference/finite element grid to assign initial condition and for subsequent validation in model studies. Interpolation is only tool available because it is not feasible to carry out test pumping and to measure water table elevation at each nodal points or at locations large enough in number to get a batter spacial distribution of these values, due to engineering and economic reasons.

This reports deals with such an interpolation programme for development of water table levels. This is achieved by fitting a trend-surface on space coordinates approximating to the water table level, using least square approach.

2.0 REVIEW

The awareness of conservation and protection of natural resourcess has increased because of their limited availability which led to their optimal utilization. This, in turn, has demanded for their proper assessment and management which require sophisticated model; so that reliable predications can be made of present and future nature/man made impacts on the system. In general, the parameters of water resources systems(e.g.,an underground aquifer)are not directly measurable. Also, because of engineering and economic reasons it may not be feasible to measure directly all the system state variables. Under such circumstances, it is necessary to treat the problem as being one of the distributed parameter identificatioin problems. A considerable amount of research has been conducted dealing with parameter identification of ground water aquifer models, which states two main approaches of parameter-estimation, viz., direct approch and optimization approach. The direct approach treats the parameter as dependent variable in formal boundary value problem. Sagar et al.(25) proposed a method that approximates the dependent variable from measured sample by means of various interpolation algorithms. Fritch and Pinder [21] used the uniqueness of the problem by prescribing the flow at every streamline. Nelson [19] used the energy dissipation method to calculate the unknown properties of the field, assuming known boundary conditions and avilable flow data.

Apart from the above investigators,others who considered the problem from direct method point of view are stallman [29], Sammel [26], Neuman [20], Emseller and De-Marsilly [11], Yeh et al.,[30] and others.

The indirect method which seems to be the most popular approach amongst the researchers of this area because of far less dependence on the necessary experimental data. The indirect method sets up the problem as minimization of a prescribed performance index, which is basically a measure of the difference between the observed output and the model output. This method is known as response error method in system engineering. To formulate the problem in indirect way , various investigators have employed various means,e.g, Slater and Durrer [28] presented a method based on least square technique and linear programming. Chang and Yeh [4] solved the problem using quadratic programming.

The solution of the problem of groundwater parameter identification is based on hydraulic head data. But due to the limitations of measurement techniques, the data usually contain noise that is not necessarily uncorrelated. It is obvious that the results of identification methods are very sensitive to measurement errors in data. Sadeghipour and Yeh [24] studied the ordinary least square technique and generalised least square technique especially designe to reduce the effect of correlated errors and concluded that the results of numerical experiment suggest that generalised least square technique offers a promising approach in effeciently improving the reliability of the estimated parameters.

The functional approximation for spacial variation of peizometric head consists of approximating the true functional relation over space coordinates for peizometric head. This functional relation may either be 'exact', i.e., no residue at the observation points or the least square type, Ralston [22]. Sagar et al. [25] used the spline function approximation which involves passing piecewise continuous polynomial functions through the known functional values at the observation points alongwith the compatibility condition at the interface of the adjacent polynomials. This method is numerically equivalent to fitting by french curves.

Kashyap and Chandra [15] have proposed a numerical scheme for quantitative estimation of parameters related to geohydrological and hydrological characteristics of groundwater aquifer, employing historic data of hydraulic head, rainfall, pumpage, etc. Their scheme is based upon the constrained minimization of sum of squares of residues in Boussinesq equation. The derivatives of hydraulic head were estimated by the least square polynomial approximation.

Mohan Rao [18] critically compared the methods of interpolation of groundwater table as applied to a sub-basin of Agra District (India). From the critical review of the past studies on surface fitting, it is observed that only space coordinates are considered as independent variables upon which the water table level was assumed to depend and in order to use higher degree polynomial, higher order product of space coordinate are also considered as independent variables, the reasonability of which is to be verified.

3.0 PROBLEM DEFINITION

The present problem is to fit a trend surface through the observed values of water-table level at definite points, the space coordinates of which with reference to any origine are known. It is also proposed to examine the validity of taking higher order products of space coordinate as independents variables for water-table level as dependent variable.

4.0 METHODOLOGIES

4.1 General

The surface fitting for groundwater problems requires to fit a smooth surface through the observed hydraulic head at fixed points in space. The interpolation of the hydraulic head values at intermediate points can be obtained by graphical method of drawing watertable contours. This method is especially subjective if the observation points are sparsely distributed.The numerical method of interpolation normally consists of exact functional approximation, e.g., Lograngian interpolation or spline function etc. In these methods a contineous piece of polynomial surface is fitted exactly through the observed head values. The above type of functional approximation exhibit artifical undulations because of the inherrent observational errors in the observed data and also the differentiation of these function may cause serious'noise'problems. The smooth surface can be obtained by least square approximation which makes the use of Weirstrass theorem.

4.2 Exact Fit

The functional approximation of hydraulic head values may be represented by

$$\hat{h}_{ij} = f'_j (x_i, y_i) \qquad \ldots(1)$$

Where

$\hat{h}_{ij}$ is true value of hydraulic head at i^{th} station and j^{th} time period.

(x_i, y_i) is the coordinate of i^{th} station.

But considering the fact that there exists always an observational error associated with each observed hydraulic head value, the above relationship may be written as

$$h_{ij} = f_j (x_i, y_i) + e_{ij} \qquad \ldots(2)$$

Where

h_{ij} is the observed hydraulic head value at i^{th} station and j^{th} time period

e_{ij} is the observation error associated with observed head value at i^{th} station.

Since there exists a large variation in head values at any station for different time steps, it is always better to assume separate functional relations for different time steps, i.e., f_j. Hence, for a particular time step, equation (2) may be written as

$$h_i = f(x_i, y_i) + e_i \qquad \ldots(3)$$

Where

e_i are independent (uncorrelated) error random variables mathematically, with zero means and unknown variance σ^2, mathematically,

$$E(e_i) = 0 \text{ and } V(e_i) = \sigma^2$$

In case of exact fit the number of observation point is the same as the number of cofficient of the approximated function. Thus for each observation, an equation similar to (2) may be obtained and one has the number of unknown coefficients of approximated function equal to the number of simultaneous equations, the solution of which will yield the values of coefficients of the approximated function. This method suffers from a serious drawback as the number of data points required is equal to the number of coefficients. Hence, the number of data points fix the degree of polynomial to be fitted. If data is avialable at large number of points, a higher degree polynomial should be adopted which lead to artificial undulations in surface violating the concept of smooth surface. Also,with larger number of data points, the number of simultaneous equations to be solved increases leading to more computer memory requirement.

4.3 Least Square Method

The method of least square makes the use of noisy functional values to generate smooth approximation to the functions, and then, these smooth approximations can be used to approximate the derivative of function more exactly than the exact approximation. If m is the number of term in approximating polynomial and n is the total number of data points then, for the least square method $m+1 < n$ and in this case smoothening of data is possible. If $m+1 > n$, we get an approximation the deviation of which from the true function is good but the derivatives will be much different from true function.

The most obvious frequency distribution of residue [e_i in eq.(3)], that one may assume is gaussian since intutively one can say that residue can take positive or negative value with equal probatility and larger residue will be occuring less frequently and vise-versa. If this assumption relating to the guassian distribution of residues holds good then, the minimization of the sum of squares of the residues yield most likelyhood estimate for the parameters. The influence of the finite size is much larger as compared to the number of parameters to be estimated. Thus, this is a necessary requirement for assuming least square solution to be identical to the most likelyhood solution.

For the third order surface polynomial, equation (3) may be extended as given below.

$$h_i = f(x_i, y_i) + e_i = \beta_1 + \beta_2 x_i + \beta_3 y_i + \beta_4 x_i^2 + \beta_5 x_i y_i + \beta_6 y_i^2 + \beta_7 x_i^3 + \beta_8 x_i^2 y_i + \beta_9 x_i y_i^2 + \beta_{10} y_i^3 + e_i \qquad \ldots(4)$$

For any value of (x_i, y_i), these are a distribution of h , whose mean is $f(x_i, y_i)$ and whose variance is σ^2. The best estimates of unknown quantities $\beta_1, \beta_2 \ldots\ldots\ldots \beta_{10}$ are obtained by minimization of sum of squares of deviations (residues) with respect to β_1, β_2, $\beta_3 \ldots\ldots \beta_{10}$ and are denoted by $b_1, b_2, b_3 \ldots\ldots\ldots b_{10}$. The sum of the square of residue is given by

$$SSR = \sum_{i=1}^{n} [h_i - f(x_i, y_i)]^2 \qquad \ldots(5)$$

Minimizing SSR, the partial derivatives of it with respect to $b_1, b_2 \ldots\ldots b_p$, where p is the total number of coefficient in the approximating surface polynomial, should be equated to zero, to yield the following set of simultaneous equation.

$$b_1 n + b_2\Sigma x_i + b_3\Sigma y_i + b_4\Sigma x_i^2 + b_5\Sigma x_i y_i + \ldots\ldots + b_{10}\Sigma y_i^3 = \Sigma h_i$$

$$b_1\Sigma x_i + b_2\Sigma x_i^2 + b_3\Sigma x_i y_i + b_4\Sigma x_i^3 + b_5\Sigma x_i^2 y_i + \ldots + b_{10}\Sigma x_i y_i^3 = \Sigma x_i h_i$$

$$b_1\Sigma y_i + b_2\Sigma x_i y_i + b_3\Sigma y_i^2 + b_4\Sigma x_i^2 y_i + b_5\Sigma x_i y_i^2 + \ldots + b_{10}\Sigma y_i^4 = \Sigma y_i h_i$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots(6)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$b_1\Sigma y_i^3 + b_2\Sigma x_i y_i^3 + b_3\Sigma y_i^4 + b_4\Sigma x_i^2 y_i^3 + b_5\Sigma x_i y_i^4 + \ldots + b_{10}\Sigma y_i^6 = \Sigma y_i^3 h_i$$

The above set of simultaneous equations may be represented in a matrix notation as given below.

$$[E] \; [b] = [H] \qquad \ldots(7)$$

where

$$[E] = \begin{bmatrix} n & \Sigma x_i & \Sigma y_i & \Sigma x_i^2 & \ldots\ldots\ldots\ldots & \Sigma y_i^3 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i y_i & \Sigma x_i^3 & \ldots\ldots\ldots\ldots & x_i y_i^3 \\ \Sigma y_i & \Sigma x_i y_i & \Sigma y_i^2 & \Sigma x_i^2 y_i & \ldots\ldots\ldots\ldots & y_i^4 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \Sigma y_i^3 & \Sigma x_i y_i^3 & \Sigma y_i^4 & \Sigma x_i^2 y_i^3 & \ldots\ldots\ldots\ldots & y_i^6 \end{bmatrix}$$

$$[b] = \begin{bmatrix} b_1 \\ b_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ b_{10} \end{bmatrix}$$

$$[H] = \begin{bmatrix} \Sigma h_i \\ \Sigma x_i h_i \\ \Sigma y_i h_i \\ \Sigma x_i^2 h_i \\ \cdot \\ \cdot \\ \cdot \\ \Sigma x_i y_i^2 h_i \\ \Sigma y_i^3 h_i \end{bmatrix}$$

The solution of the above equation(7) will yield the value of b_i, b_2, b_3 .... b_{10}.

4.4 Statistical Analysis

The statistical problems involved in the regression analysis are as given below.

1. To obtain the best point and interval estimators of unknown regression parameter.
2. To determine the adequacy of assumed model.
3. To verify the set of relevant assumptions.

The choice of the appropriate model is not a statistical one but rather, it should be derived from the underlying physical situation.

4.4.1 Factors affecting statistical surface fitting.

The following factors can adversely affect the trend surface analysis. The effect of these factors can range from a slight distortion of trend to total invalidation of the results

1. Adequate control must be present, i.e., the number of data points must exceed the number of coefficients in the polynomial equation otherwise the results of regresion are invalid.

2. In order to perform statistical tests to check the validity of the model, degree of freedom must be sufficiently large.

3. The number and spacing of control points has a direct influence on the size of local deviations that can be detected in trend surface fitting.

To be more precise in explaining the points one and two above, it is obvious that the correlation will always increase with the addition of terms. When number of terms becomes equal to (n-1), the correlation equals 1.00, regardless of how widely the data points are scattered.

4.4.2 Analysis of variance

The significance of a regression may be tested by performing an analysis of variance. This is a process of separating the total variation of a set of observation into components associated with defined source of variation, viz, due to regression and due to deviation from regression. The terms expressing the variation of dependent variable, h_i are being given below.

a) Total sum of squares (SS_T) of h_i

$$SS_T = \sum_{i=1}^{n} h_i^2 - \left(\frac{(\sum_{i=1}^{n} h_i)^2}{n} \right)$$

$$= \sum_{i=1}^{n} (h_i - \bar{h})^2$$

b) Sum of square due to regression (SS_R) of h_i.

$$SS_R = \sum_{i=1}^{n} (\hat{h}_i - \bar{h})^2$$

c) Sum of square due to deviation from regression (SS_D).

$$SS_D = SS_T - SS_R$$

This is a measure of the failure of the least square line to fit the data point. This can also be defined as

$$SS_D = \sum_{i=1}^{n} (\hat{h}_i - h_i)^2$$

The goodness of fit to the points can be defined by

$$R^2 = \frac{SS_R}{SS_T}$$

$$\simeq 1.0 \text{ (For good estimation of data)}$$

$$R = \sqrt{SS_R/SS_T} = \text{Coefficient of correlation.}$$

5.0 ANALYSIS

A computer program was developed for fitting a surface through the observed water-table level at fixed points the space coordinate of which are known. The surface polynomial of varying degree were fitted through the observed points and the results were compared. The program consists of soloving equation (7) and formulation of the matrices involved. Analysis of variance,correlation and different statistical tests, i.e., F-test and t-test have also been incorporated in the program accordingly.

A sample problem was taken with ten points in space having known coordinates and water table levels which are given is table 1.

TABLE 1

OBSERVED WATER TABLE LEVELS

Sl.No.	Coordinates (in m) x	y	Observed water-table level (in m)
1.	18298.73	22349.78	208.25
2.	16970.59	21755.05	207.08
3.	15716.32	19955.19	205.21
4.	17708.45	20669.96	207.67
5.	18781.80	20432.56	207.36
6.	20559.29	20330.78	209.38
7.	15863.89	18272.67	205.26
8.	17976.69	17685.74	206.18
9.	18050.57	13984.28	203.94
10.	19922.03	20307.36	208.98

For the above problem surface was fitted for varying degree of polynomial When the higher order products of x and y are considered as independent variables, it was observed from the correlation matrix that the higher order products are highly correlated among themselves. But for least square method the independent variable choosen should be truely independent and should not be correlated among themselves. Hence, ultimately it was found that only x and y can be taken as independent variable and not the higher order product of x and y. The comparison between the above two results were shown in table 2. Thus, taking only two independent variable, i.e, x and y, the results obtained are as given below.

TABLE 2

COMPARISON BETWEEN OBSERVED AND PREDICTED WATER TABLE LEVELS

Sl.No.	Observed Value (m)	Computed value (m)	
1.	208.25	208.38	
2.	207.08	207.11	R = 0.98
3.	205.21	205.33	F =98.23
4.	206.67	207.01	
5.	207.36	207.84	Standard
6.	209.38	209.12	Error
7.	205.26	204.65	of
8.	206.18	205.96	Estimate
9.	203.94	204.28	= 0.366
10.	208.98	208.63	

In order to have better results some other variables should be taken as independent upon which watertable level is likely to depend, even taking x and y as independent variable, is not very much justified because in this way watertable level at any point is being assumed to depend only on space coordinates and all the other important variables are being neglected.

6.0 CONCLUSIONS

The main conclusions drawn from the analysis of surface fitting are summerised below.

1. The higher order product of space coordinates should not be taken as independent variables but only space coordinates taken individually as independent variables may yield equally good results.

2. In order to obtain better result, the other variables upon which watertable level is likely to depend, should be taken as independent variables in the solving problem of surface fitting.

REFERENCES

1. Battacharya B.K.,(1969), Bicubic spline interpolation as a method of treatment of potential field data, Geophyiscs, Vol. 34(3), pp. 402-423.

2. Brige, R.T. and J.W. Winberg, (1947),'Least square fitting of data by means of polynomials,'Rev. Mod. Phy., Vol. 19, pp 298-360.

3. Brittles, A.B. and Morel, E.H, (1979),'calculation of aquifer parameters from sparse data,'water Resources Research, Vol. 15(4), pp 832-844.

4. Chang, S. and W.W-G.Yeh, (1976), 'A proposed algorithm for the solution of large scale inverse problem in groundwater,'Water Reso. Research, Vol.12(5), pp 365-375.

5. Chatterjee, S.,(1977),'Regression analysis by example, 'Wiely New Yord, N.Y., 228.

6. Chayes, F., (1970)' On deciding wether trend surfaces of progressively highly order are meaningful,'Bulletin of Geological Society of America, Vol. 81, pp 1273-1278.

7. Cooley, R.L., (1982), 'Incorporation of prior information on parameters into nonliniear regression groundwater flow model,'1' theory, Water Resources Research, Vol. 18(4), pp 965-974.

8. Daniel, C.F.S. Wood and J.W. Gorman, (1971), 'Fitting equation to data,'Wiley Interscience, New York, N.Y., 342 pp.

9. Davis, J.C.,(1973), 'Statistics and data analysis in geology,'John Wiley, New York, N.Y.

10. Edwards, K. (1972),'Estimating area rainfall by fitting surfaces to irregularly spaced data, proceedings of the Geilo Symposium, Worls Meteo. Org. Publ. no. 327.

11. Emsellew, Y. and G.de Marsilly, (1971), ' An automatic solution for the inverse problem, 'Water Resources Research, Vol.7(5), pp 1264-1283.

12. Fritch, J.M., (1971), 'Objectve analysis of a 2D field by the cubic spline technique, 'Monthaly Weather Review, Vol. 99.

13. Guest, P.G. (1961),'Numerical method of curve fitting, ' Combridge Univ. press, New York.

14. Hardy, R.L.,(1971), 'Multi-quadric equations of topography and other irregular surfaces, Jour. Geophysical Research, Vol. 76, No. 8.

15. Kashyap, D. and Chandra, S., (1982), 'A nonliniear optimization method for aquifer parameter estimation,'Jour. Hydrology, Vol. 57, pp 163-173.

16. Koch, G.S. Jr. and R.F. Link, (1971), 'Statistical analysis of geological data', Vol. 2., John Wiley and sons, Inc, New York, 438 pp.

17. Mathematical model study for induced recharge in Ganga Basin.(1980), Hydrology report - 4. School of Hydrology, University of Roorkee, India.

18. Mohan Rao, K.M.,(1983), 'Interpolation technique for groundwater modelling studies,'A.M.E. dissertation in Hydrology, University of Roorkee, India.

19. Nelson, R.W., (1968),'In-place determination of permeability distribution for heterogeneous porous media through analysis of energy dissipation.,'Society of Petroleum Engineers.Journal, pp 33-42.

20. Neuman, S.P., (1983),'Calibration of distributed parameters of groundwater flow models viewed as multiple objective decision process under uncertaintly.,'Water Resources Research, Vol. 9(4), pp 1006-1021.

21. Pinder, G.F., and W.G. Gray, (1977),'Finite element simulation in surface and subsurface hydrology,'295 pp., Academic, New York.

22. Ralston, A., (1965),'First course in numerical analysis, Mc Graw-Hill Inc., 578 pp.

23. Rice, J.R., (1964), 'The application of functions, Vol. I, 'Addison-Wesley publishing company, Inc.Reading, Mass.

24. Sadeghipour, J. and W.W.G. Yeh, (1984),'Parameter identification of groundwater aquifer models,'A generalised least square approach,'Water Resources Research, Vol. 20(7), pp 971-979.

25. Sagar, B., C.C. Kaisiel and L. Duckstein, (1975), 'A direct method for identification of the parameter of dynamic non-homogeneous aquifer,'Water Resources Research, Vol. 2(4), pp 563-570.

26. Sammel, E.A., (1964), 'Evaluation of numerical analysis method for determining variation in transmissibilities, 'Int. Assoc. Sci. Hydrol., Gentbrugge, Belgium.

27. Show, E.M., and P.P. Lynn, (1972),'Areal evaluation using two surface fitting techniques, 'Hydrological science Bulletin, Vol. 17(4).

28. Slater, G.E., and E.J. Duner, (1971),'Adjustment of reservoir simulation model to match field performance, 'Soc. Petrol. Eng. J., Vol. 11(3), pp 295-305.

29. Stallmon, R.W., (1963)'Calculation of resistance and error in an electric analogy of steady flow through nonhomogeneous aquifers Geol. Survey (U.S.) Water supply paper 20, pp 1544-1546.

30. Yeh, W.W-G, and G.W. Tauxe, (1971), 'Optimal identification of aquifer diffusivity using quasilinearization, 'Water Resources Research, Vol. 7(4), pp 955-962.

31. Watson, G.S., (1971), 'Trend surface analysis,'Jour. Inter. Assoc. Mathematical Geology, Vol. 3, No. 3, pp. 215-226.

...-...