CONF-9009254--1 --/

DE90 017737

CALIBRATION OF GROUNDWATER FLOW MODELS USING MONTE CARLO SIMULATIONS AND GEOSTATISTICS

R. J. HOEKSEMA
Calvin College, Grand Rapids, MI 49546
R. B. CLAPP
Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

ABSTRACT Groundwater flow models require specification of several input parameter fields which are inferred from limited data. In this paper the hydraulic conductivity and recharge rate are estimated for an unconfined aquifer under steady flow conditions. Usually point observations of conductivity and head are available for the estimation of the distributed conductivity field and the recharge rate. Use of numerical flow models require that these fields be prescribed as average values over finite elements. The geostatistical solution to this problem uses linear estimation to obtain the distributed conductivity field and the recharge rate. Monte Carlo simulations are used here to compute the covariances associated with the head data. Results obtained for both artificial and real data show that the head data is effective in improving the estimates that would result using conductivity data alone. The use of Monte Carlo simulations results in a method which can be used under a wider variety of modeling conditions than previous applications of the geostatistical approach.

INTRODUCTION

Groundwater flow models require specification of several physical input parameter fields. Such parameter fields might be the hydraulic conductivity or transmissivity, storage coefficient, recharge rate, etc. These parameter fields are all characterized by two things. First, they are all spatially distributed parameter fields. In other words, these fields vary throughout the aquifer. Second, our knowledge regarding the values of these fields is limited to randomly located, error-prone, point measurements. Typically, these measurements include variations of the parameter at a scale smaller than the scale of interest in the flow model. The calibration problem is one of providing estimates of these spatially distributed fields aggregated to finite regions for use in flow models. These fields must be inferred from the available measurements.

Along with model input parameters mentioned above, observations of output variables are also usually available. The primary output variable is the hydraulic head. Predicting the input parameters based on available observations of the output variables is generally

MASTER
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

referred to as the "inverse problem". A comprehensive review of work in this area is presented by Yeh (1986). Inverse solutions were developed in an attempt to find parameters which reproduce head measurements. Pure inverse methods attempt to solve the flow equation in reverse. In other words, the conductivities are found as a function of the head measurements. These methods generally produce ill-posed problems. Also, actual observations from the parameter fields (i.e. conductivities) are not used. Improvements on pure inverse methods generally involve using the conductivity measurements to stabilize or limit the inverse solution.

The geostatistical approach to the inverse problem follows the description given above (see Hoeksema & Kitanidis, 1984, 1985b, 1989). Sparse, point observations of head and conductivity are used to estimate the complete, spatially distributed conductivity field. The geostatistical approach finds the estimates as a linear function of all of the available data. The primary computational technique used is cokriging which requires covariances between all of the measurements and the quantities to be estimated.

The work presented in this paper is an extension of the geostatistical approach to the inverse problem as presented in Kitanidis & Vomvoris (1983) and Hoeksema & Kitanidis (1984, 1985b, 1989). The geostatistical approach can be presented as a five step procedure. 1.) The input parameter fields are treated as realizations of some random function or process. A model is proposed which specifies the spatial correlation structure of these unknown input parameters in terms of a few unknown mean and covariance function parameters. 2.) The differential equation of flow is used to get the mean and covariance associated with the heads as a function of the mean and covariance parameters given in step 1. 3.) The mean and covariance parameters introduced in step 1 are estimated using the available observations of both head and conductivity. 4.) The mean and covariance function parameters estimated in step 3 are validated. 5.) Finally, the input parameter fields (i.e. conductivity, recharge rate, etc.) are found using linear estimation techniques.

In the work of Hoeksema & Kitanidis (1984, 1985b, 1989) a linearized form of the flow equation was used to obtain the head covariances as a linear function of the transmissivity covariance function. This approach simplified the task of obtaining the head covariances. The resulting method was limited to the case of confined aquifers with prescribed head boundaries and no recharge. It also required that the variations in transmissivity were small. The primary goal here is to make the method more generally applicable by using Monte Carlo simulations to obtain the head covariances. This results in a more general method but it involves a loss in computational efficiency. The application presented in this paper is for the particular case of an unconfined aquifer with both prescribed head and zero flux boundary conditions. The recharge rate is an additional parameter.

This paper will first present the development of the five steps described above. It will highlight the differences between this application and those described in the Hoeksema and Kitanidis papers. Next, the results of implementation of this method in several test cases will be presented.

DETAILS OF THE GEOSTATISTICAL APPROACH

1. Specify the model for spatial variability of input parameters

In the first step of the geostatistical solution a model is proposed for the mean and covariance function of the input parameters. The input parameters will be the natural logarithm of the saturated hydraulic conductivity (hereafter referred to as simply lnK) and the recharge rate. The conductivity is assumed to follow a log-normal distribution. Two forms of lnK are used in this work. Point lnK refers to available point measurements and element lnK refers to values that would be used in a flow model. The lnK is considered to be spatially distributed but the recharge rate will be modeled as a constant over the aquifer. The variable used to represent lnK is Z and the variable used to represent the recharge rate is S. Point lnK values will be represented as Z_P while element lnK will be Z_E. The unknown parameters in the mean and covariance model will be estimated in step 3. The spatial variability model for point lnK is given by the following specification of mean and covariance:

$$E[Z_{Pi}] = Z_M = \text{constant} \tag{1}$$

$$Cov(Z_{Pi}, Z_{Pj}) = E[(Z_{Pi}-Z_M)(Z_{Pj}-Z_M)] = \theta_1\delta(\xi) + \theta_2 \exp(-\xi/\theta_3) \tag{2}$$

In (1) Z_M is the mean lnK which is assumed constant. In (2) ξ represents the separation distance between the two points i and j, and $\delta(\xi)$ is the Kronnecker delta function ($\delta(\xi)$ is zero if $\xi \neq 0$ and one if $\xi=0$). The mean parameter is Z_M and the covariance parameters are θ_1, θ_2, and θ_3 (in vector form, $\underline{\theta}$). The parameter θ_1 is the variance of unstructured lnK variability which is due to measurement error and variability of the lnK field on a small scale. The parameter θ_2 is the variance of structured lnK variability which is associated with separation of the measurements points (i.e. measurements of lnK near to each other tend to be more highly correlated than measurements far from each other). The correlation length associated with the structured lnK variability is θ_3. In geostatistical terminology, the model presented here is equivalent to an exponential variogram (sill = $2\theta_1 + 2\theta_2$) with a nugget (nugget = $2\theta_1$). Other covariance models can be used. The model presented in (2) has been found to be useful for a wide variety of problems (see Hoeksema & Kitanidis, 1985a) and is the only one used in this work.

The values of lnK used in a groundwater flow model are usually average values defined over a finite domain. Since estimates will eventually be computed for these area-averaged values we need to expand the covariance model to handle them as well. An element lnK is simply the average of a point lnK over the domain of interest. The models for mean and covariance of element lnK are then the following:

$$E[Z_{Ei}] = Z_M \tag{3}$$

$$Cov(Z_{Ei}, Z_{Ej}) = \frac{1}{D_i D_j}\int_{D_i}\int_{D_j} Cov(Z_{Pi}, Z_{Pj})\, dD\, dD \tag{4}$$

$$Cov(Z_{Pi}, Z_{Ej}) = \frac{1}{D_j}\int_{D_j} Cov(Z_{Pi}, Z_{Pj})\, dD \qquad (5)$$

The domains of integration, D_i and D_j, are the elements of the flow model. The covariance functions in (4) and (5) for one dimensional elements can be evaluated in a closed form. For a two-dimensional element Gauss quadrature is used for the numerical integration.

The recharge rate is assumed to follow a normal distribution. The mean and variance model for recharge rate is simple. Since the recharge rate is assumed in this application to take on a single value over the entire aquifer all that is needed is a mean and variance. It is assumed that the modeler has some prior estimate of the recharge rate taken as the mean, S_m. Also, the modeler has some measure of the uncertainty associated with the recharge rate estimate. This uncertainty is quantified in terms of a recharge rate standard deviation, σ_S.

2. Obtain the Mean and Covariance for Output Variables

The second step in the geostatistical solution is that of obtaining mean and covariance relationships for the output variables. In this step the head-related covariances are determined. These covariances will be determined as functions of the (as yet undetermined) lnK covariance parameters, $\underline{\theta}$.

The differential equation of flow in an unconfined aquifer with a horizontal base and recharge under steady conditions and simplified by the Dupuit-Forchheimer approximation (see Freeze & Cherry, 1979) is the following:

$$\frac{\partial}{\partial x}\left(Kh\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(Kh\frac{\partial h}{\partial y}\right) + S = 0 \qquad (6)$$

In (6) K is the saturated hydraulic conductivity, S is the recharge rate, and h is the hydraulic head using a horizontal base as the datum. Rewriting (6) in terms of the head-squared ($V = h^2$) and in terms of lnK, Z, results in the following:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial Z}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial Z}{\partial y}\frac{\partial V}{\partial y} + 2Se^{-Z} = 0 \qquad (7)$$

The equations of flow are simplified by the use of head-squared, V, instead of head, h. In fact for the case of a uniform conductivity field the flow equation is linearized by this substitution. For this reason V will be used as the primary output variable instead of h (h data are easily transformed to V). The goal, then, is to obtain the covariances for V measurements using the relationships given in (1) through (5) and the equation of flow (7).

The Monte Carlo (MC) simulation procedure is used to determine the relationship between the head related covariances, $Cov(V_i, V_j)$, $Cov(V_i, Z_{Bj})$, and $Cov(V_i, Z_{Pj})$ and the lnK covariance model parameters. Also the relationship between $E[V_i]$ and $\underline{\theta}$ will be sought. Since θ_1 is a measure of the uncorrelated random noise in the point lnK measurements, it has no influence on the head and therefore does not affect the above covariances. Also, the value of θ_3 will be simply assumed and not estimated. Therefore the required functional relationship is between these covariances and θ_2 only. It is impossible using MC simulation to establish an analytical expression for each of these covariance as a function of θ_2. Instead, a piecewise linear relationship will be developed by setting θ_1 to zero and finding the above covariances for several uniformly distributed values of θ_2.

For a specific value of θ_2 many unconditional simulations of lnK are generated. For the several examples presented later in this paper the number of simulations required to obtain stable covariances ranged between 5000 and 8000. A set of random element lnK values, $\underline{Z}_E^k$, are first generated which follow the model described in (3) and (4). This set of lnK's represents the values associated with the elements in the flow model. Next, a set of point lnK values is generated, $\underline{Z}_P^k$. This set follows the model specified in (1) and (2) and is conditioned on the element values generated above (and therefore are consistent with (5)). These point values are generated to represent possible point observations at the nodes of the model. The k superscript refers to the kth Monte Carlo simulation of these fields. Next, a random value of recharge rate is generated, S^K, based on the values of S_M and σ_S. The head simulation is obtained by using the element lnK simulation, $\underline{Z}_E^k$, and the simulated recharge, S^k, as input to a flow model. Finally, the appropriate covariances and means are computed by averaging the results from all of the simulations.

3. Estimate Mean and Covariance Model Parameters

The third step of the geostatistical solution involves estimating the lnK mean and covariance model parameters using the available measurements. The mean lnK, Z_M, is estimated using a simple average of the measured values. The mean lnK is only used in generation of lnK simulations (step 2) and not in either covariance parameter estimation or in estimation of the lnK field itself (step 5).

Since the head related covariances are determined as a function of the lnK covariance parameters it makes sense to use all of the available data (both lnK and V) to estimate these parameters. The approach followed in Hoeksema & Kitanidis (1985b) is to assume a reasonable value for the correlation length, θ_3 (see Hoeksema & Kitanidis, 1985a), and then use maximum likelihood estimation (MLE) to estimate θ_1 and θ_2. The MLE procedure used is described in detail in Kitanidis & Lane (1985). The MLE procedure used requires the measurement covariance as function of the parameters. Equation (2) gives the lnK measurement covariance as a function of $\underline{\theta}$ and the Monte Carlo simulations described in step 2 give the head-squared covariances as a piece-wise linear function of $\underline{\theta}$.

The MLE procedure used assumes that all of the data are jointly normally distributed. It is generally assumed that lnK is normally distributed. For small variations in lnK (i.e. $\theta_2 < 1.0$) the head-squared given by (7) is nearly linear in lnK (see Hoeksema & Kitanidis, 1984). Head-squared will then be considered to be normally distributed also. The normal distribution assumption is required only during the estimation of the parameters θ_1 and θ_2 (and not in the estimation of the lnK field by cokriging). This is a reasonable thing to do for two reasons. First, the lnK estimation procedure using cokriging (step 5) is quite insensitive to the actual values of these parameters therefore the error associated with θ_1 and θ_2 estimates should not adversely effect the final results. Second, MLE provides reasonable least squares estimates of the parameters even when the data are non-Gaussian (see Kitanidis, 1985).

4. Validation of the Estimated Covariance Model Parameters

The procedure used in this work to validate the results of covariance model parameter estimation is the same as that used in Hoeksema & Kitanidis, (1984, 1985b). Model validation can be accomplished by tests performed on a set of uncorrelated residuals obtained from the parameter estimation procedure. If the assumptions regarding the normality of the data are correct and the estimated covariance model parameters are indeed maximum likelihood then this set of residuals should have a zero mean and unit variance. Tests can be performed to check the above assumptions.

5. Prediction of the Input Parameters Using Linear Estimation

Linear estimation procedures are used to determine the best estimates of the input parameter fields. In our case the input parameters are the element lnK field and the recharge rate. The lnK field is estimated using cokriging and in this case the recharge rate is found using a simple least squares procedure. This estimation is the heart of the calibration process. Cokriging seeks to find the best, unbiased, minimum variance estimate of the unknown quantities as a linear function of the available measurements. The measurements in this problem are the point observations of lnK and head-squared. The form of the cokriging equations used in this work appear in Hoeksema, et. al. (1989). The cokriging equations require the covariance values established in the first three steps.

To estimate the recharge rate it is first assumed that cokriging finds the best estimate of the lnK field. Then the linear relationship between head-squared and recharge rate is used to fit the best recharge rate to the head-squared data using a simple least squares technique. For a given hydraulic conductivity field the recharge rate and the head-squared are linearly related. If V_{si} is the head-squared at measurement point i associated with the cokriged lnK field and the mean recharge rate, S_M, and V_i is the same except for a value of recharge rate equal to S, then

$$V_i = V_{si} + \alpha_i (S - S_M) \qquad (8)$$

(A measurement point is a location where measurements of head and therefore V are available). The α_i can be easily computed by selecting an arbitrary $(S-S_M)$ and using the flow model to obtain the resulting V_i at each point. The estimated recharge rate, $\hat{S}$, is then the value of S in (8) which minimizes the difference between V_i and the measured (true) values designated as V_{ti}.

$$\hat{S}=S_M+\frac{\sum_{i=1}^{n}\alpha_i(V_{ti}-V_{si})}{\sum_{i=1}^{n}\alpha_i^2} \qquad (9)$$

RESULTS OF MODEL TESTING

So far this paper has presented an application of geostatistics to the problem of groundwater model calibration using MC simulations for computing the head-squared covariances. The testing of this approach is done via the use of two computer programs. The first program is designed to generate artificial data sets for testing the method. The second program performs the calibration for both artificially generated data and for real problems.

One-dimensional example

The first test case presented is a one-dimensional aquifer. The aquifer properties were generated with a mean lnK, Z_M, of -11.0, and lnK covariance model parameters $\underline{\theta}$ of 0.1, 1.5, and 100 m. The recharge rate mean and standard deviation, S_M and σ_S, were set to $0.5(10)^{-7}$ m s^{-1} and $0.2(10)^{-7}$ m s^{-1}. The total model length was 200 m with 21 nodes, zero flux left end boundary, and prescribed head-squared (25 m^2) right end boundary. The data generation program used an actual recharge rate of $0.393(10)^{-7}$ m s^{-1}.

The solid lines in Figs 1 and 2 show the generated fields. In each figure the lower line shows the generated element lnK field and the upper line shows the generated head-squared field. Two different calibration runs were made. RUN1DA used 7 point lnK measurements and 7 head-squared measurements uniformly distributed

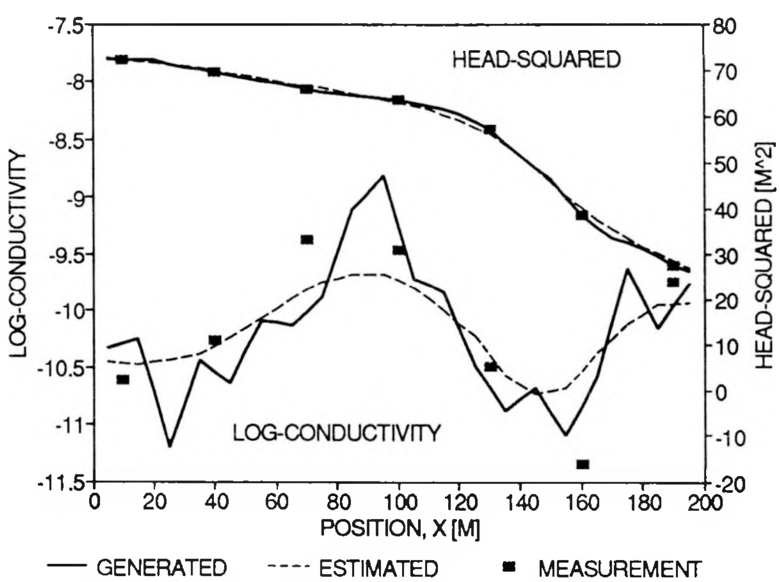

FIG. 1 One-dimensional example - RUN1DA

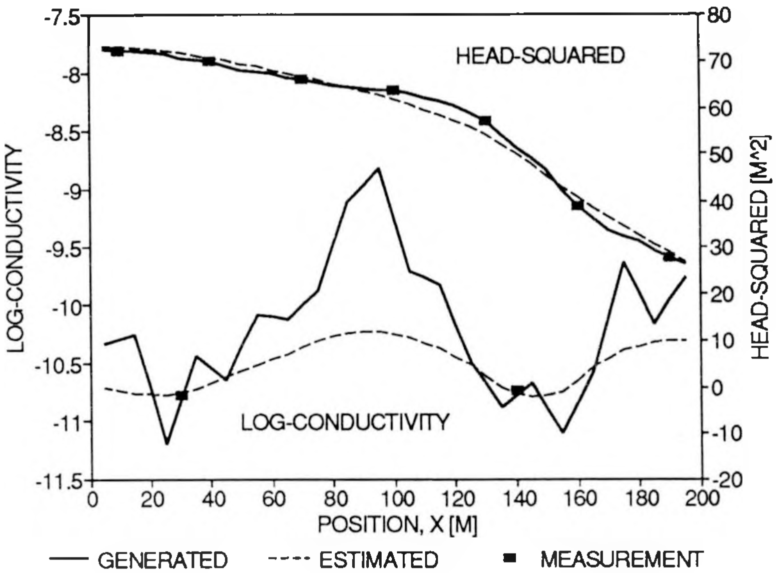

FIG. 2 One-dimensional example - RUN1DB.

over the model length (Fig. 1). The second run, called RUN1DB, is intended to show the influence of head-squared data on the estimate of lnK when only minimal lnK data are available. RUN1DB uses the same 7 head-squared measurements as RUN1DA but it used only 2 point lnK measurements (see Fig. 2). Note that the lnK measurements are error prone point values which tend to deviate from the element values.

The estimated element lnK from the calibration is shown as a dashed line in Figs 1 (RUN1DA) and 2 (RUN1DB). Also shown is the estimated head-squared which is found by using the estimated lnK and recharge rate as input to a flow model. Fig. 1 shows that 7 measurements of head and conductivity are sufficient to reproduce the essential features of the true fields. The effect of a non-zero θ_1 term is seen in that the estimated element lnK does not pass directly through the point lnK measurements. Fig. 2 shows the effect of the head data in the lnK estimation. Even though only two point measurements of lnK were used the primary shape of the true lnK is reproduced.

The recharge rate is also estimated. The actual value used in the generation of the data was $0.393(10)^{-7}$ m s^{-1}. The predicted recharge rate is $0.450(10)^{-7}$ m s^{-1} for RUN1DA and $0.369(10)^{-7}$ m s^{-1} for RUN1DB. Since the RUN1DB conductivity estimate is generally lower than the RUN1DA estimate the resulting recharge rate must be estimated lower to maintain reasonable head-squared values.

Several statistics can be used to measure the performance of the calibration. For generated data the true fields are known, so the average squared errors associated with the element lnK estimates and with the nodal head-squared estimates can be computed. For comparison, similar statistics are computed for the case of an element lnK field equal to the mean of the point lnK measurements. To obtain a head-squared estimate from this mean lnK model a recharge rate is used which minimizes the difference between the estimated head-squared field and the available (7 in this case) head-squared measurements. For RUN1DA the average squared error of lnK estimation is reduced from 0.354 for the mean model to 0.132 for the geostatistical calibration. The reduction for the average squared error of head-squared estimation is reduced from 6.58 m^2 to 0.669 m^2. These error reductions demonstrate the effectiveness of the calibration.

Two-dimensional example

A two-dimensional run was also made to test the method. The model is rectangular with sides of 200 and 120 m in length. The boundaries were zero flux along 3 sides and prescribed head along the fourth. The mean lnK was set to -11.0 with covariance parameters ($\underline{\theta}$) of 0.25, 0.9, and 100 m. The recharge rate was generated by the program to have a value of $0.303(10)^{-7}$ m s^{-1}.

The calibration run used 7 head and 7 lnK measurements. Like the one-dimensional case an average squared error of estimation can be computed for both element lnK and nodal head-squared. In the move from the mean model to the geostatistical calibration, the average squared error of estimation was reduced from 0.514 to 0.155 for lnK and from 15.0 m^2 to 11.6 m^2 for head-squared. The recharge rate was estimated as $0.352(10)^{-7}$.

Two-dimensional case study

A preliminary calibration was performed using data from a hydrogeologic study of an old landfill near Holland Michigan, USA (Prein & Newhof, 1989). The site, now used as a park, is approximately 20 ha and is bounded on 2 sides by a small creek. The creek forms the prescribed head boundary for the flow model. Eight head measurement and 5 conductivity measurements were available from monitoring wells. The head measurements varied from 6.57 to 7.15 m above the base of the aquifer. The conductivity measurements varied from $3.04(10)^{-4}$ to $5.46(10)^{-4}$ m s^{-1} (lnK from -8.10 to -7.51).

The only measure of performance available for this calibration is the error between the actual head-squared measurements and the head-squared estimates at measurement points. The mean model, as described above, gives an average squared error of head-squared estimation at measurement points of 0.582 m^2. The geostatistical calibration results in an average squared error of 0.0585 m^2 with an estimated recharge rate of $0.383(10)^{-7}$ m s^{-1}. This reduction in error is quite significant considering the relatively flat head field.

CONCLUSIONS

This report has described the application of the geostatistical approach to the problem of calibrating a groundwater flow model. The method uses Monte Carlo simulation to obtain head related covariances and is applied to the particular case of flow in an unconfined aquifer with recharge. As developed in this study the method predicts both the conductivity field and the recharge rate.

The primary goal of this work was to show the feasibility of using MC simulations with geostatistics to perform model calibration. The results of tests done on artificial aquifer models show that the method works quite well when compared to using just the mean lnK value. The quality of the results are based on both matching the predicted lnK field with the true field but also by comparing the predicted heads to the true heads. The tests done using only two lnK measurements show that much of the basic lnK variation can be recovered from the head data.

ACKNOWLEDGEMENTS This research was sponsored by the Office Environmental Restoration, U.S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

REFERENCES

Freeze, R. A. & J. A. Cherry (1979) Groundwater, Prentice-Hall, Englewood Cliffs, N.J.

Hoeksema, R. J. & P. K. Kitanidis (1984) An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. Wat. Resour. Res. 20 (7), 1003-1020.

Hoeksema, R. J. & P. K. Kitanidis (1985) Analysis of the spatial structure of properties of selected aquifers. Wat. Resour. Res. 21 (4), 563-572.

Hoeksema, R. J. & P. K. Kitanidis (1985) Comparison of Gaussian conditional mean and kriging estimation in the geostatistical solution of the inverse problem. Wat. Resour. Res. 21 (6), 825-836.

Hoeksema, R. J. & P. K. Kitanidis (1989) Prediction of transmissivities, heads, and seepage velocities using mathematical modeling and geostatistics. Adv. Water Resour. 12 (2), 90-102.

Hoeksema, R. J., R. B. Clapp, A. L. Thomas, A. E. Hunley, N. D. Farrow & K. C. Dearstone (1989) Cokriging model for estimation of water table elevation. Wat. Resour. Res. 25 (3), 429-438.

Kitanidis, P. K. (1985) Minimum-variance unbiased quadratic estimation of covariances of regionalized variables. J. Int. Assoc. Math. Geol. 17 (2), 195-208.

Kitanidis, P. K. & R. W. Lane (1985) Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method. J. Hydrol. 79, 53-71.

Kitanidis, P. K. & E. G. Vomvoris (1983) A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Wat. Resour. Res. 19 (3), 677-690.

Prein & Newhof, Professional Consultants (1989) Holland Charter Township Quincy Park Hyrodgeological Study, Grand Rapids, Michigan.

Yeh, W. W-G. (1986) Review of parameter identification procedures in groundwater hydrology: The inverse problem. Wat. Resour. Res. 22 (2), 95-108.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."