Comparison of Gaussian Conditional Mean and Kriging Estimation in the Geostatistical Solution of the Inverse Problem

ROBERT J. HOEKSEMA

Department of Engineering, Calvin College, Grand Rapids, Michigan

PETER K. KITANIDIS

St. Anthony Falls Hydraulic Laboratory, Department of Civil and Mineral Engineering
Minneapolis, Minnesota

Two separate applications of the geostatistical solution to the inverse problem in groundwater modeling are presented. Both applications estimate the transmissivity field for a two-dimensional model of a confined aquifer under steady flow conditions. The estimates are based on point observations of transmissivity and hydraulic head and also on a model of the aquifer which includes prescribed head boundaries, leakage, and steady state pumping. The model used to describe the spatial variability of the log-transmissivity describes large-scale fluctuations through a linear mean or drift intermediate and small-scale fluctuations through a two-parameter covariance function. The first application presented estimates the log-transmissivities using Gaussian conditional mean estimation. The second application uses an extended form of cokriging. The two methods are compared and their relative merits discussed. The extended cokriging application is applied to the Jordan Aquifer of Iowa. A comparison is also made between the conditional mean application and an analytical approach.

1. INTRODUCTION

The successful modeling of groundwater flows through the use of numerical techniques relies on a knowledge of the spatial variation of the aquifer system variables, inputs, and boundary conditions. The inverse problem in groundwater modeling is usually defined as estimation of the spatially varying transmissivity field with an emphasis on its subsequent use in groundwater models. The transmissivity field is estimated based on all available information, including (1) point observations of the piezometric head, (2) measurements of transmissivity (usually from pumping tests), (3) estimates of the areal extent and boundary features of the aquifer, (4) estimates of the rate of water removal from the aquifer, and (5) estimates of the leakage characteristics of confining layers. The most recent efforts in the solution of the inverse problem presuppose the piezometric head and transmissivity fields to be realizations of random functions. Regression procedures or statistical estimation can then be applied to solve the problem.

The scope of this work is that of estimating the transmissivity field for a two-dimensional model of a confined aquifer under steady flow conditions. The estimation procedure will be based on the five types of information described above. The emphasis here is on regional aquifer models. For this reason the measured values of transmissivity are assumed to represent aquifer behavior at a scale smaller than the dimension of a typical model discretization element.

The geostatistical approach to the inverse problem is described in detail in the works by Kitanidis and Vomvoris [1983] and Vomvoris [1982]. An application of the geostatistical solution is presented in the work by Hoeksema and Kitanidis [1984], and a similar technique is presented in the work by Dagan [1985]. The geostatistical solution can be viewed as a five-step procedure. First, a model for the statistical spatial variability of the transmissivity field is proposed. Second, the differential equation of groundwater flow is used to relate the spatial variability of the piezometric head to that of the transmissivity. Third, the observed values of head and transmissivity are used to estimate the unknown parameters in the transmissivity spatial variability model. Fourth, the validity of the fitted spatial variability model is tested. If the model passes the validity test then, finally, linear estimation procedures are applied to estimate the spatially distributed transmissivity field. The estimation variance is also computed to measure the accuracy of the predicted transmissivity field.

The work of Dagan [1985] describes an application of the geostatistical approach which uses an analytical technique (based on some simplifying assumptions) to relate the head and transmissivity variability. He also uses Gaussian conditional mean to estimate the transmissivity field. In this work, model-specific discretized equations of flow are used to relate the head and transmissivity variability. Both Gaussian conditional mean and a form of kriging will be presented as methods to estimate the transmissivity field.

The major contributions of this paper beyond those of Hoeksema and Kitanidis [1984] are several. First, the solution presented here provides for vertical flow (leakage or pumping) into or out of the aquifer. Second, comparisons are made here with the analytical technique of Dagan [1984] and also between the Gaussian conditional mean approach (classical linear estimation theory) and kriging. Finally, the kriging application presented here is new in that it uses linearized mean heads to avoid use of an estimated log-transmissivity mean in the kriging equations.

2. INVERSE PROBLEM SOLUTION USING GAUSSIAN CONDITIONAL MEAN

In this section an application of the geostatistical solution is developed which utilizes Gaussian conditional mean to estimate the transmissivity field. The development here follows the five-
step procedure described in the introduction. It should be emphasized at this point that the geostatistical solution is quite general, and the approach presented here is just one possible application.

### 2.1. Select a Statistical Spatial Variability Model for the Transmissivity Field

Throughout this work it is assumed that the transmissivity follows a lognormal probability distribution. Freeze [1975] and Hoeksema and Kitanidis [1985] provide evidence in support of this conclusion. Instead of dealing with the transmissivity, then, the natural logarithm will be used. The log-transmissivity will be frequently designated in the text as ln T. Its variable representation in equations is Y.

Since the log-transmissivity is assumed to be normally distributed, all that is needed is a model for its mean and covariance. The model consists of an ln T drift and stationary zero mean random terms. Here, the model used for the correlations is

\[
E(Y_i) = F_i = \theta_{d1} + \theta_{d2}x_i + \theta_{d3}y_i
\]

\[
\text{Cov}(Y_i, Y_j) = \theta_{i1} \delta_{ij} + \theta_{i2} \exp \left\{-D_{ij}/l_f\right\}
\]

where \(Y_i\) is the ln T value at point \(i\); \((x_i, y_i)\) are the cartesian coordinates of point \(i\); \(E[\cdot]\) represents mean or expected value; \(\delta_{ij}\) is Kronecker’s delta (\(\delta_{ij} = 1\) if \(i = j\) and \(\delta_{ij} = 0\) if \(i \neq j\)); \(D_{ij}\) is the scalar distance between points \(i\) and \(j\); and \(l_f\) is the correlation length. The drift parameters \(\theta_d\) and the covariance parameters \(\theta_c\) are considered to be unknowns which will be estimated from the data.

The first term in (2) represents small-scale ln T variability. This is variability due to measurement error and due to property variations at a very small scale. It is important to recognize this type of variability, since it has no effect on regional variations of hydraulic head. Therefore by estimating \(\theta_{i1}\), small-scale variability can be easily filtered out of the transmissivity field predictions. The second term in (2) represents the structured or large-scale ln T variability. In this model, \(l_f\) is the length over which the ln T values become uncorrelated, and will be assumed to be known. This assumption greatly simplifies the parameter estimation procedure, since it keeps the covariance function linear in the parameters. It can be shown that the resulting estimated ln T field is quite insensitive to the assumed value of \(l_f\) [see Hoeksema, 1984]. Hoeksema and Kitanidis [1985] provide an analysis of several real aquifers indicating typical values of \(l_f\).

### 2.2. Relate Head Variability to ln T Variability

The differential equation relating the piezometric head \(\phi\) to the log-transmissivity \(Y\) under the conditions listed in the introduction is as follows:

\[
\frac{\partial \bar{Y}}{\partial x} \frac{\partial \phi}{\partial x} + \frac{\partial ^2 \bar{Y}}{\partial x^2} \frac{\partial ^2 \phi}{\partial y^2} + \frac{\partial \bar{Y}}{\partial y} \frac{\partial ^2 \phi}{\partial y^2} = -\left(Q_i + \frac{K}{M} (\phi_i - \bar{\phi})\right) e^{-T}
\]

where \(Q_i\) is the specified vertical inflow (plus sign) or outflow (minus sign) rate; \(K\) is the hydraulic conductivity of the confining layer; \(M\) is the thickness of the confining layer; and \(\phi_i\) is the piezometric head on the other side of the confining layer. The head perturbation approach is applied here in which the log-transmissivity \(Y\) is replaced by its mean, \(F\), plus a zero mean perturbation \(f\) \((Y = F + f, \text{where } F = E[Y])\). Also, the head \(\phi\) is replaced by its expected value \(H\) plus a zero mean perturbation \(h\) \((\phi = H + h, \text{where } H = E[\phi])\). As long as the perturbation \(f\) is small [see Gutjahr and Gelhar, 1981; Dagan, 1982; and Mizell et al., 1982], products of perturbations can be ignored, and \(e^{-T}\) can be approximated by \((1 - f)\). These simplifications yield the following differential equation:

\[
0 \frac{\partial^2 \bar{H}}{\partial x^2} + \theta_{d2} \frac{\partial \bar{H}}{\partial x} + \theta_{d3} \frac{\partial \bar{H}}{\partial y} + \frac{\partial ^2 \bar{H}}{\partial y^2} + \frac{\partial ^2 \bar{h}}{\partial y^2} = -\left(Q_i + \frac{K}{M} (\phi_i - H)\right) e^{-T}
\]

Taking expected values of both sides results in the equation which gives the expected head \(H\):

\[
\theta_{d2} \frac{\partial ^2 \bar{H}}{\partial x^2} + \theta_{d3} \frac{\partial ^2 \bar{H}}{\partial y^2} = -\left(Q_i + \frac{K}{M} (\phi_i - H)\right) e^{-T}
\]

Note that given the drift parameters \(\theta_d\), the expected head field can be computed from (5) utilizing a numerical groundwater flow model. Subtracting (5) from (4) yields the stochastic differential equation relating the perturbations \(f\) and \(k\):

\[
\theta_{d2} \frac{\partial ^2 \bar{h}}{\partial x^2} + \theta_{d3} \frac{\partial ^2 \bar{h}}{\partial y^2} + \frac{\partial \bar{h}}{\partial x} + \frac{\partial \bar{h}}{\partial y} = \frac{-\theta_{d2} - \theta_{d3} - \theta_{d1} + \frac{K}{M}}{2} e^{-T} k
\]

Following a finite difference or finite element approach, (6) can be written in discrete form as

\[
\Delta h = Bf
\]

or

\[
h = A^{-1} Bf
\]

where \(h\) is the vector of head perturbations, and \(f\) the vector of ln T perturbations associated with either the finite difference "blocks" or with the finite element nodes. It should be noted that the matrices \(A\) and \(B\) are both functions of the drift parameters \(\theta_d\) (remember that \(f\) is a function of \(\theta_d\)).

The goal here is to obtain the spatial variability model for the head. While (5) provides the mean or expected head \(H\), (8) can be used to obtain the head covariance matrix \(\Omega_{\bar{h} \bar{h}}\) and the head-in T cross-covariance matrix \(\Omega_{\bar{h} \bar{Y}}\). These are as follows:

\[
\Omega_{\bar{h} \bar{h}} = E[hh^T] = (A^{-1} B) E[ff^T] (A^{-1} B)^T
\]

\[
\Omega_{\bar{h} \bar{Y}} = E[hf^T] = (A^{-1} B) E[ff^T]
\]

In (9) and (10) \(E[ff^T]\) is the ln T covariance matrix whose values are computed from (2). Note that \(\Omega_{\bar{h} \bar{h}}\) and \(\Omega_{\bar{h} \bar{Y}}\) are therefore linear in the ln T covariance parameters \(\theta_c\).

A few details have been omitted in the development of (9) and (10) for the cause of simplicity. First, these equations give the covariance matrices for the heads and ln T associated with the aquifer discretizations scheme; needed is the covariance matrices for the measurements. Secondly, the covariance matrices, \(E[ff^T]\) should take into account the integrated effect of the finite area associated with each element in the \(f\) vector. This requires averaging of (2) over finite areas. (The result here is that \(\Omega_{\bar{h} \bar{h}}\) and \(\Omega_{\bar{h} \bar{Y}}\) do not depend on \(\theta_{i1}\).) The interested reader should consult Hoeksema and Kitanidis [1984] for a more complete description of these details.
2.3. Estimate the Unknown Statistical Parameters

The measured values of head and log-transmissivity are next used to estimate the unknown parameters of the model described in section 2.2. (Note that the complete vector of both drift parameters \( \theta_d \) and covariance parameters \( \theta_c \) will be designated simply as \( \theta \)). Maximum likelihood parameter estimation is used here to simultaneously estimate all five unknown parameters. The Gauss-Newton procedure is applied to determine the maximum likelihood estimates \( \hat{\theta} \) [see Kitanidis and Vomvoris, 1983; Kitanidis and Lane, 1985]. Required for application of the Gauss-Newton method is the complete head and in \( T \) measurement vector \( z \), the measurement expectations \( E(z) \); the derivative of the measurement expectations with respect to each parameter, \( \partial E(z)/\partial \theta_i \); the measurement covariance matrix (obtained via equations (2), (9), and (10)) as a function of the parameters, \( Q_{xx}(\theta) \); and the measurement covariance matrix derivatives with respect to each parameter, \( \partial Q_{xx}(\theta)/\partial \theta_i \). These terms can be computed given the expected head derivatives with respect to the drift parameters. A method for computing these derivatives is given in the appendix. Hoeskema [1984] provides more details related to obtaining the required vectors and matrices.

2.4. Check Validity of Fitted Spatial Variability Model

Model validation can be accomplished by tests performed on a set of uncorrelated residuals obtained from the complete measurement vector \( z \), the fitted measurement expectation vector \( E(z) \), and the fitted measurement covariance matrix \( Q_{xx} \). This set of residuals is computed as

\[
s = V^{-1}(z - E(z))
\]

where \( VV^T = Q_{xx} \). Under the model assumptions \( s \) should be a vector of independent normal standard normal deviates.

2.5. Estimate the \( \ln T \) Field

The last step involves estimating the \( \ln T \) field. If the assumption that the \( \ln T \) is normally distributed is valid, then the approach taken here uses the Gaussian condition mean to estimate the \( \ln T \) field and the conditional variance to estimate the quality of the estimated field.

If the log-transmissivity estimate averaged over a finite area or block located at point \( 0 \) is designated as \( \hat{Y}_0 \), then the conditional mean gives \( \hat{Y}_0 \) as

\[
\hat{Y}_0 = E[Y_0] + Q_{Y0}Q_{xx}^{-1}(z - E(z))
\]

where \( E[Y_0] \) is available given estimates of the drift parameters, and \( Q_{Y0} \) represents the cross covariance (row) matrix between the block-averaged log-transmissivity being estimated and each measurement. The matrix \( Q_{Y0} \) is obtained using (2) and (10). The conditional variance is given by

\[
\text{Var}(\hat{Y}_0) = \text{Var}(Y_0) - Q_{Y0}Q_{xx}^{-1}Q_{Y0}^T
\]

The unconditional variance \( \text{Var}(Y_0) \) is obtained via averaging of (2) over the block.

2.6. Comments on the Gaussian Conditional Mean Approach

The results of extensive testing reported in the work by Hoeskema [1984] indicate that the approach described in sections 2.1 through 2.6 provides excellent estimates of the \( \ln T \) field. Some of these results will be presented later. The primary difficulty associated with using the Gaussian conditional mean approach is that the computed variance (equation (13)) tends to underestimate the true squared \( \ln T \) prediction error. This is related to two problems. First, the simultaneous estimation of both mean and covariance parameters leads to biased estimates of the covariance parameters. This bias in \( \hat{\theta} \) subsequently effects the conditional variance given by (13). The bias in \( \hat{\theta} \) is due to not accounting for the sampling variability in the estimates of \( \theta_p \). The bias gets worse as the correlation length increases. Also, if the correlation length is significant relative to the aquifer dimension, the addition of more data will not eliminate this bias. The second cause for the underestimation of the conditional variance is that equations (12) and (13) assume that the unconditional mean \( E[Y_0] \) is known with certainty.

3. Inverse Problem Solution Using Extended Corriging

The major criticism against the use of the Gaussian conditional mean is that it attempts to simultaneously estimate the unknown covariance and drift coefficients and then use these values in the linear estimation stage. This procedure determines estimation variances which are too small compared to the actual mean square errors. If the measurement mean vector \( E[z] \) is a linear function of the drift parameters, then the mean can be filtered out of the data and parameter estimation can be applied without bias. Furthermore, universal kriging or cokriging can then be applied to estimate the \( \ln T \) field without the need to know the drift parameters. Since the expected head \( H \) is a nonlinear function of the drift parameters, this standard procedure cannot be applied as is. An extension of this procedure is thus developed, apparently for the first time. The analysis, named extended cokriging, will be based on linearization of the expected head field. This procedure also allows the introduction of additional unknown parameters. In this work a procedure for estimation of log-transmissivities in the presence of an unknown leakage coefficient will be developed.

The \( \ln T \) spatial variability model presented in section 2.1 and the relationship between the head and \( \ln T \) variability discussed in section 2.2 remain unchanged.

3.1. Estimate the Unknown Statistical Parameters

The goal of this application is to reduce the dependence of the estimated covariance parameters and the estimated log-transmissivity field on the values of the drift parameters \( \theta_p \). This is done through an iterative procedure which involves two steps. The first step uses maximum likelihood estimation to improve the covariance parameter estimate \( \hat{\theta} \), based on the data and the current value of \( \theta_p \). The second step uses weighted least squares to obtain new approximate values for the drift parameters.

The bias in the parameter estimation procedure of section 2 is due to the simultaneous estimation of measurement mean and covariance related parameters. The procedure described in this section will reduce the bias by using a transformation of the data to filter out the head and \( \ln T \) mean. Filtering out the \( \ln T \) mean \( F_L \) is straightforward, since it is linear in the unknown drift parameters \( \theta_p \). However, to filter out the mean head \( H_0 \), the relationship for \( H_0 \) must first be linearized about the current approximate value of \( \theta_p \).

If \( \hat{H} \) represents the value of the mean or expected head using the current approximate value of the drift parameters \( \theta_p \), then the linearized expected head is given as

\[
H_0 \approx \hat{H}_0 + \sum_{j=1}^{n} \frac{\partial \hat{H}_0}{\partial \theta_p} (\theta_p - \theta_p^0)
\]
where $ND$ is the number of unknown drift parameters. The error in using (14) is on the order of $(\hat{\theta}_D - \theta_D)^2$. Next a new quantity $H'_i$ is introduced where $H'_i$ is just the expected head at point $i$, $H_i$ minus a quantity which is a function of the approximate drift parameters:

$$H'_i = H_i - \left( \hat{H}_i - \sum_{j=1}^{ND} \frac{\partial H_j}{\partial \theta_j} \theta_j \right)$$

(15)

The point is that this new expected head $H'_i$ can now be expressed as being linear in the actual drift parameters:

$$H'_i = \sum_{j=1}^{ND} \frac{\partial H_i}{\partial \theta_j} \theta_j$$

(16)

If the head measurements, $\phi_i$, are similarly transformed to $\phi'_i$ where

$$\phi'_i = \phi_i - \left( \hat{H}_i - \sum_{j=1}^{ND} \frac{\partial H_j}{\partial \theta_j} \theta_j \right)$$

(17)

then $E[\phi'_i] = H'_i$ and $(\phi'_i - \hat{H}_i) = (\phi_i - H_i)$. The result is a new quantity $\phi'_i$ whose mean is (approximately) linear in the $\ln T$ drift parameters and whose variance, covariance, and $\ln T$ cross covariance are identical to that of the actual head $\phi_i$. The values of $\partial H_j/\partial \theta_j$ are evaluated using the methods described in the appendix. The new measurement vector is $x'$, where $x' = [\phi'_1, \phi'_2, \ldots, \phi'_n; y_1, y_2, \ldots, y_m]$ and its covariance matrix $Q_{d}(\theta_D)$ is the same as that described in section 2 for the original measurements, $x$. The measurement mean vector $E[x]$ is now linear in the drift parameters (in the vicinity of their current estimates) and is given as

$$E[x'] = X \theta_D$$

(18)

where $X$ is the matrix of measurement expectation derivatives. (Note that $X_D$ represents the unknown drift parameter vector, while $X_D$ is a current approximation to $\theta_D$ so that $(\theta_D - X_D)$ is “small”).

To filter out the unknown mean from the measurements the transformation $z_{mod} = Ux'$ is used, where $U$ is the transformation matrix given in the work by Kitanidis [1983] as

$$U = T_{\theta_D}(I - X(X^TX)^{-1}X^T)$$

(19)

In (19) $T_{\theta_D}$ represents a transformation which simply eliminates $ND$ rows from the projection matrix $(I - X(X^TX)^{-1}X)$. This is necessary, since the projection matrix has rank no larger than $n + m - ND$. Transformed in this fashion the modified measurement vector has a zero mean and covariance given by

$$Q_{mod}(\theta_D) = U Q_{d}(\theta_D) U^T$$

(20)

The covariance parameter estimates can now be improved using the Gauss-Newton method of maximum likelihood estimation. Note that since $z_{mod}$ has a zero mean and only the covariance parameters are being improved, the only derivatives needed are $\partial Q_{mod}(\theta_D)/\partial \theta_1$ and $\partial Q_{mod}(\theta_D)/\partial \theta_2$.

The goal of this procedure is to reduce the dependence of the covariance parameter estimation procedure on the current approximation to the drift parameters. Unlike in the linear case, where the modified measurement vector $z_{mod}$ does not depend on the drift coefficient estimates, in the nonlinear case the dependence of $z_{mod}$ is only reduced to second-order terms (i.e., $z_{mod}'$ is a function of $(\theta_D - \hat{\theta}_D)^2$).

After the covariance parameter estimates have been improved, new approximations to the drift parameters are obtained using weighted least squares (WLS). The WLS estimate is given as $\hat{\theta}_D = (X^TX)^{-1}X^TWx'$, where $W$ is the matrix of weights. The weights are obtained as $W = Q_{d}^{-1}(\theta_D)$. The new drift parameter estimates are then used in the next iteration to obtain the new $Q_{d}(\theta_D)$. The iterative procedure terminates when the negative log-likelihood (in the procedure to estimate $\theta_D$) reaches a stationary value.

3.2. Estimate the $\ln T$ Field

The log-transmissivity field will be estimated in this analysis through extended cokriging. The $\ln T$ estimate is assumed to be a linear combination of the $\ln T$ measurements $y_i$ and the transformed heads $\phi'_i$. Therefore

$$\hat{y}_i = \sum_{j=1}^{n} \mu_j \phi'_j + \sum_{j=1}^{m} \lambda_j Y_j$$

(21)

where $\mu_j$ and $\lambda_j$ are the cokriging coefficients to be determined. First it is required that the estimate $\hat{y}_i$ be unbiased. To do this the quantity $E[\hat{y}_i - y_i]$ is set to zero:

$$E[\hat{y}_i - y_i] = \theta_D \sum_{j=1}^{n} \mu_j \frac{\partial H_j}{\partial \theta_1} + \theta_D \sum_{j=1}^{n} \mu_j \frac{\partial H_j}{\partial \theta_2}$$

$$+ \theta_D \sum_{j=1}^{m} \lambda_j \frac{\partial H_j}{\partial \theta_D}$$

$$+ \theta_D \sum_{j=1}^{m} \lambda_j \frac{\partial H_j}{\partial \theta_D}$$

$$+ \lambda_j \sum_{j=1}^{m} \lambda_j y_j - \theta_D \sum_{j=1}^{m} \lambda_j y_j - \theta_D y_0$$

(22)

For this expression to be zero regardless of the value of the drift parameters, the following relations must hold:

$$\sum_{j=1}^{n} \mu_j \frac{\partial H_j}{\partial \theta_1} + \lambda_j \sum_{j=1}^{m} \lambda_j y_j = 1$$

(23)

$$\sum_{j=1}^{n} \mu_j \frac{\partial H_j}{\partial \theta_2} + \lambda_j \sum_{j=1}^{m} \lambda_j y_j = x_0$$

(24)

$$\sum_{j=1}^{m} \lambda_j \frac{\partial H_j}{\partial \theta_D} + \lambda_j \sum_{j=1}^{m} \lambda_j y_j = y_0$$

(25)

Next, the estimation variance must be minimized. The estimation variance (assuming that equations (23) through (25) hold) is given as

$$E[(\hat{y}_i - y_i)^2]$$

$$= \sum_{i=1}^{n} \sum_{k=1}^{n} \mu_i \mu_k \text{Cov}(\phi_i, \phi_k) + \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_j \lambda_k \text{Cov}(y_i, y_j)$$

$$+ 2 \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_i \lambda_j \text{Cov}(\phi_i, y_j) - 2 \sum_{i=1}^{n} \mu_i \text{Cov}(\phi_i, y_0)$$

$$- 2 \sum_{j=1}^{m} \lambda_j \text{Cov}(y_j, y_0) + \text{Var}(y_0)$$

(26)

Minimization of the estimation variance subject to the constraints of (23) through (25) results in the following set of equations to be solved simultaneously along with (23), (24), and (25) for the coefficients $\mu$ and $\lambda$:

$$\sum_{i=1}^{n} \mu_i Q_{mod}(k, i) + \sum_{j=1}^{m} \lambda_j Q_{mod}(k, j) + \mu_i \frac{\partial H_k}{\partial \theta_1}$$

$$+ \lambda_j \frac{\partial H_k}{\partial \theta_2}$$

$$+ \lambda_j \frac{\partial H_k}{\partial \theta_D}$$

$$= \text{Cov}(\phi_k, y_0), k = 1, \ldots, n$$

(27)
\[
\sum_{i=1}^{n} x_i \mu_i Q_{xj}(i, j) + \sum_{j=1}^{m} \lambda_j Q_{rr}(i, j) + v_1 + v_2 x_i + v_3 y_j
\]

\[= \text{Cov} (Y_i, Y_j), \quad i, j = 1, \ldots, m \quad (28)\]

In the above equations \(v_1\), \(v_2\), and \(v_3\) are the Lagrange multipliers.

The method as described here should produce excellent results under rather general conditions. The required conditions are that the covariance of the drift parameters is small and the expected head field is not too nonlinear in the drift coefficients. The neglected terms in (14) are of the order \(\partial^2 H/\partial \theta_{ap}^2 (\theta_{ap} - \bar{\theta}_p)^2\), where \(H\) is the expected head, \(\theta_{ap}\) represents a drift parameter, and \(\theta_{ap}\) is its most recent approximation. If the expected head field is not too nonlinear then \(\partial^2 H/\partial \theta_{ap}^2\) is small; if the covariance of the drift parameter is small then \(\theta_{ap}\) obtained through WLS should provide a good approximation to \(\theta_p\). These two conditions combined make (14) valid.

Cokriging is an improvement over Gaussian conditional mean estimation in the case of unknown drift coefficients. The conditional mean approach applied in section 2 assumes that the expected \(ln T\) and head fields are known with certainty. This results in an underestimation of the estimation variance. Extended cokriging, instead, provides an estimation variance which is more nearly correct. The cokriging variance tends to be, if anything, somewhat conservative (high) simply because it effectively assumes that the measurement means are completely unknown a priori, while in reality some prior information may be available about them. Both methods were found in our applications to provide practically equally good estimates of the log-transmissivity field.

3.3. Additional Unknown Parameters

The method just presented estimates the \(ln T\) field in a fashion which attempts to minimize the dependence on the unknown \(ln T\) drift parameters. This formulation readily allows the addition of other unknown parameters. In this case an unknown leakage coefficient will be considered.

In this application the conductivity of (3), \(K\), is replaced by \(K_0 K_{\theta ap}\) where \(\theta_{ap}\) is an unknown leakage coefficient multiplier and \(K_0\) varies spatially. This unknown parameter does not affect the \(ln T\) mean but it does influence the mean head; consequently, it may be treated in the same way as the unknown drift parameters. The derivative of the expected head with respect to this parameter is obtained using the technique described in the appendix. Values of \(\partial H/\partial \theta_{ap}\) can then be obtained for all the head measurement point locations. These values appear as an extra column in the \(X\) matrix of (18).

Fig. 1. Aquifer model used to generate data and test inverse solution. Shaded area represents area of leakage from confining layer.

Fig. 2. Contour map of generated block averaged log-transmissivities. Also shown are the 16 generated point log-transmissivities used in the inverse solution.
The parameter estimation procedure is exactly as described earlier; the only difference is the additional column in the $X$ matrix and the increase in $ND$ from 3 to 4. The addition of the unknown leakage coefficient effectively removes one measurement from the parameter estimation procedure (i.e., $z_{mod}$ has length $n + m - ND$). Additional unknown parameters would have the same effect.

The cokriging equations change slightly with the addition of the unknown leakage coefficient. Equation (22) now includes the term

$$
\theta_{p4} \sum_{i=1}^{n} \mu_i \frac{\partial \hat{H}_i}{\partial \theta_{p4}}
$$

which results in an additional constraint equation

$$
\sum_{i=1}^{n} \mu_i \frac{\partial \hat{H}_i}{\partial \theta_{p4}} = 0
$$

(29)

The cokriging equation (27) is changed as follows:

$$
\sum_{i=1}^{n} \mu_i Q_{10}(k, i) + \sum_{j=1}^{m} \lambda_j Q_{10}(k, j)
$$

$$
+ v_1 \frac{\partial \hat{H}_1}{\partial \theta_{p1}} + v_2 \frac{\partial \hat{H}_2}{\partial \theta_{p2}} + v_3 \frac{\partial \hat{H}_3}{\partial \theta_{p3}} + v_4 \frac{\partial \hat{H}_4}{\partial \theta_{p4}}
$$

$$
= \text{Cov}(\phi_r, Y_{oh}, k = 1, \ldots, n)
$$

(30)

The relationship for the cokriging variance is unchanged.

**TABLE 1. Results of Runs Comparing Gaussian Conditional Mean Estimation to Cokriging**

<table>
<thead>
<tr>
<th>Run</th>
<th>$\theta_{p1}$</th>
<th>$\theta_{p2}$</th>
<th>$\theta_{p3}$</th>
<th>$\theta_{p4}$</th>
<th>$\text{MSE}(\epsilon_{a})$</th>
<th>$\text{MSE}(\epsilon_{er})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.271</td>
<td>0.213</td>
<td>5.80</td>
<td>0.0128</td>
<td>0.0077</td>
<td>0.0578</td>
</tr>
<tr>
<td>B</td>
<td>0.274</td>
<td>0.261</td>
<td>5.90</td>
<td>0.0137</td>
<td>0.0052</td>
<td>0.0412</td>
</tr>
<tr>
<td>C</td>
<td>0.256</td>
<td>0.247</td>
<td>5.84</td>
<td>0.0142</td>
<td>0.0057</td>
<td>0.0526</td>
</tr>
<tr>
<td>*</td>
<td>0.300</td>
<td>0.300</td>
<td>7.00</td>
<td>0.0080</td>
<td>0.0000</td>
<td>$\cdots$</td>
</tr>
</tbody>
</table>

Run A uses Gaussian conditional mean estimation. Run B uses cokriging. Run C uses cokriging along with an estimated leakage coefficient.

*Data is generated using these values.

4. SIMULATION RESULTS

The predictive capabilities of the geostatistical solution to the inverse problem can best be demonstrated by application to a hypothetical aquifer. The obvious advantage is that the true transmissivity and piezometric head fields are known and can be compared to the predicted results.

A detailed description of the procedure used to generate the random transmissivity and head fields is given in the work by Hoeksema and Kitanidis [1984] and Hoeksema [1984]; only a summary is presented here. The log-transmissivity data is generated in two steps. First, the field of block-averaged log-transmissivities are generated using the spatial, variability model of (1) and (2). Note that since the values represent the average $\ln T$ over a finite area, small-scale variability is not present (i.e., $\theta_{r1}$ is set to zero), and the decay coefficient (exp $(-D_{ij}/L_1)$) must be averaged over the appropriate areas. The block-averaged $\ln T$ field is the one which should be used in a groundwater flow model of the aquifer and is therefore the one which the solution of the inverse problem is compared with. The random piezometric head field is obtained by using

**Fig. 5. Contour map comparison of the predicted head field from run B (solid curves) and the true head field (dashed curves). Heads are given in meters.**
the block-averaged \( T \) field in the groundwater flow model. In the second step, random log-transmissivities are generated which represent point observations of the true field. These are generated using (1) and (2) (with a nonzero \( \theta_{11} \)) but are conditional on the block-averaged values already generated. Note that the variability of the point measurements is larger than that of the block-averaged values. Thus the testing procedure takes a subset of the point (highly variable) in \( T \) measurements and a subset of point piezometric head values (it is assumed that there is no measurement error associated with the heads) and attempts to predict the block-averaged \( T \) field. The predicted \( T \) field can then be used in the groundwater flow model to obtain a predicted head field for comparison with the true (generated) head field.

The discretized model of the hypothetical aquifer is shown in Figure 1. Note that a finite difference method discretization is used. Figure 1 also shows the variation in piezometric head along the constant head boundary. This model of the aquifer includes leakage from a confining layer represented by the shaded region. The values used in the \( T \) and head field generation procedure are

\[
\begin{align*}
\theta_{11} &= 0.3 & \theta_{12} &= 0.3 & l_f &= 50 \text{ km} \\
\theta_{21} &= 7.00 & \theta_{22} &= 0.008/\text{km} & \theta_{12} &= 0
\end{align*}
\]

The contour map for the generated block-averaged \( T \) field is shown in Figure 2. Also shown in Figure 2 are the 16-point observations which will be used to predict the given \( T \) field. Figure 3 shows the contour map for the true head field along with the 16 observed values.

Following the solution of the inverse problem, three statistics are computed to measure the quality of the predicted fields. The first is the mean squared error of the \( T \) estimation error, \( \text{MSE}(\epsilon_T) \). This is the average squared difference of the true block-averaged \( T \) minus the predicted \( T \) computed at the 48 interior model blocks. In a similar fashion the mean squared error of the head prediction error \( \text{MSE}(\epsilon_h) \) is computed. The average is taken over all interior model blocks. The third statistic computed is the mean estimation variance ratio \( m(\text{evr}) \). Here, the \( T \) prediction error is squared and divided by the computed estimation variance. The average value is then computed for the 48 interior model blocks. Under the assumption that the estimation variance is properly computed, \( m(\text{evr}) \) should have a value of about 1. If \( m(\text{evr}) \) is greater than 1, then the estimation variance is generally too small or underestimated.

Three runs are made using the 16 head and 16 \( T \) measurements shown in Figures 2 and 3. The results are given in Table 1. The descriptions of the three runs are as follows:

Run A: the correct leakage parameters are assumed. Gaussian conditional mean estimation is employed.

Run B: cokriging estimation is used assuming the correct leakage parameters.

Run C: the leakage coefficient is estimated as described in section 3.3. Cokriging estimation is used.

The results shown in Table 1 are typical of those observed in many other runs. As is clearly shown, the covariance parameters and the estimation variance are all underestimated in the first run which uses Gaussian conditional mean estimation. This underestimation was present in virtually every run made using several different inflow conditions, \( T \) field realizations, and measurement combinations. The attempt to reduce this bias by filtering out the mean is somewhat successful in run B. Again, very typical of many other runs, \( \theta_{11} \) is increased only slightly toward the generating value of 0.3 and \( \theta_{12} \) improved to a greater extent but still some error is present. The estimation variance is no longer underestimated. The results from many other runs give values of \( m(\text{evr}) \) generally greater than 1 if Gaussian conditional mean estimation is used. However, if cokriging estimation is used, these are reduced (as is shown here) to values distributed about 1. The effect on the estimated \( T \) and head fields of these two methods are given by the statistics \( \text{MSE}(\epsilon_T) \) and \( \text{MSE}(\epsilon_h) \). In this case \( \text{MSE}(\epsilon_h) \) improves in run B. Note also that in run C, where the leakage coefficient is estimated, the fit quality is still good despite the loss of one measurement in the parameter estimation procedure.

To observe the predicted results from the inverse solution, the results of run B are shown in Figures 4 and 5. Figure 4 gives the \( T \) contour map predicted using cokriging. Note that the predicted result is smoother than the true field, since it represents the average of all possible realizations which have the given spatial variability structure and the given measured values. Figure 5 shows the predicted head field (resulting from

![Fig. 6. Contour map of the log-transmissivity field predicted based on measured values of \( T \) alone.](image)

![Fig. 7. Contour map comparison of the predicted head field from the run using only \( T \) measurements (solid curves) and the true head field (dashed curves). Heads are given in meters.](image)
the application of the predicted transmissivity field in a groundwater flow model of the aquifer) as the solid curve and the true head field as the dashed curve. Note how closely the predicted field matches the true field.

A fourth run is also made to show the effectiveness of the head data in improving the prediction of ln T and head fields. Figures 6 and 7 show the predicted ln T and head fields which result from the use of kriging estimation based on the 16-ln T measurements alone. Note how the detail in the ln T field is lost and the predicted head field does not match nearly as well (MSE of $e_\theta$) increased by a factor of 15). Typically, these predictions based on ln T measurements alone are far worse than those which use head measurements.

5. Application of the Geostatistical Solution to the Jordan Aquifer of Iowa

In a previous paper [Hoeksema and Kitanidis, 1984] the geostatistical solution was applied to the Jordan Aquifer of Iowa. A description of this aquifer and the data screening process is also described there. The cokriging estimation procedure described here (in section 3) is also used to predict the transmissivity distribution in the Jordan Aquifer.

The data set used here is the same 29 head and 56 log-
transmissivity measurements used in the final run of the previous paper (Hoeksema and Kitanidis [1984], Figure 14). Again the head measurement error variance $\sigma^2_\theta$ will be estimated. In this discussion $\sigma^2_\theta$ is treated as an additional covariance parameter, $\theta_d$. The leakage parameters used here are the same as in the previous final model with the addition of the unknown multiplier, $\theta_{d_4}$, which will be estimated. The integral scale is set to 28 miles (45 km) as given in the work by Hoeksema and Kitanidis [1985]. The results of model fittings are as follows:

$$\begin{align*}
\hat{\theta}_1 &= 0.178 \\
\hat{\theta}_2 &= 0.369 \\
\hat{\theta}_3 &= 521.7 \text{ ft}^2 \ (48.47 \text{ m}^2)
\end{align*}$$

$$\text{Cov} (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = \\
\begin{bmatrix}
0.0132 & -0.0174 & 0.0379 \\
-0.0174 & -1.729 & 0.240(10^{-3}) \\
0.0379 & -1.729 & -0.0059
\end{bmatrix}$$

$$\begin{align*}
\theta_{d_1} &= 7.65 \\
\theta_{d_2} &= 0.00275/\text{mi} \ (0.00171 \ \text{km}) \\
\theta_{d_3} &= -0.00309/\text{mi} \ (0.00192 \ \text{km})
\end{align*}$$

The resulting predicted transmissivity field is quite similar to that of the previous paper. Contour maps for the predicted transmissivity field and its upper and lower 95% confidence interval are given as Figures 8–10. The predicted transmissivity field of Figure 8 is applied to a flow modeling program to predict to hydraulic head distribution. The contour map for the heads is shown as Figure 11.

6. Comparison With an Analytic Approach

Dagan [1985] presents a solution to the inverse problem which is similar to the one presented here. In his paper he first presents the inverse problem solution in terms of conditional probabilities. Later, specific application is made using Gaussian conditional mean and variance estimation. The primary difference between his solution and the one presented here (section 2) is that the head and head-ln T covariances are obtained using an analytical approach instead of the numerical approach (i.e., through discretized equations of flow, equation (7)). The assumptions which are used to provide the analytical solution are as follows: (1) there is no vertical inflow into the aquifer; (2) the mean in $T$ is constant (i.e., no drift as presented here); (3) the solution is sought in the neighborhood of the point of interest which is sufficiently far from the
boundary (i.e., the domain is assumed unbounded); (4) the average head gradient is constant.

Head and log-transmissivity fields are again generated for a hypothetical aquifer to compare the results of this analytical solution and the numerical solution presented in section 2. The aquifer model is square with 300 km sides. The correlation length is set to 30 km, and the intermediate scale variance \( \theta_2 \) is set to 0.4. According to the approach presented by Dagan [1984], the small-scale variability (\( \theta_1 \)) is set to zero. The mean head field has a slope of \(-0.5 \text{ m/km}\) in the x-coordinate direction. Fifteen measurements of head and \( \ln T \) are used to predict the \( \ln T \) field (the closest measurement to the boundary is 40 km). Figure 12 shows the contour map for the generated \( \ln T \) field along with the 15 observed values (note the lack of small-scale variability). The primary comparison which is made between the two solutions is in their ability to reduce the prediction MSE’s through the use of head data. (In all runs the true values of the \( \ln T \) mean, variance, and correlation length are used to predict the \( \ln T \) field).

The first run made use of the 15 \( \ln T \) measurements only. The resulting prediction MSE’s are

\[
\text{MSE}(e_h) = 0.128 \quad \text{MSE}(e_g) = 2.63 \text{ m}^2
\]

Figure 13 shows the predicted \( \ln T \) contours. The next run applies the geostatistical solution presented here in section 2 using the 15-\( \ln T \) measurements along with 15 head measurements. The resulting prediction MSE’s are

\[
\text{MSE}(e_h) = 0.112 \quad \text{MSE}(e_g) = 1.24 \text{ m}^2
\]

Figure 14 shows the predicted \( \ln T \) field contours. Note that the improvement here is not nearly as dramatic as presented earlier. This is primarily due to the lack of small-scale \( \ln T \) variability in the data which allows the first run to provide a very good solution. Next the analytical solution presented by Dagan [1985] is applied using the same data as the previous run. The results are as follows:

\[
\text{MSE}(e_h) = 0.129 \quad \text{MSE}(e_g) = 1.63 \text{ m}^2
\]

Figure 15 shows the predicted \( \ln T \) field contours. (Note that the measurements used to predict the log-transmissivity at any block are those within a 90-km radius of the block centerpoint).

The analytical solution, in this example, clearly provides a better estimate of the \( \ln T \) field than that using no head data. Even though MSE\((e_h)\) is not reduced, the reduction in
MSE($\varepsilon_\rho$) indicates a better solution. The numerical solution does a better job than the analytical solution in this example. This is primarily due to the influence of the boundaries in the model. (The reader is reminded that the analytical solution assumes an unbounded domain). The boundaries are seen to have a greater effect on the head-lnT cross covariances than on the head covariances.

Small-scale lnT variability can also be easily included in the analytical approach presented by Dagan [1984]. This is done by adding the variance of the small-scale lnT fluctuations, $\theta_{1s}$, to the lnT variance terms in the measurement covariance matrix. Again the results obtained using the numerical approach and the analytical approach can be compared.

The head and lnT fields are now generated under the same conditions as before except $\theta_{1s}$ is set to 0.4 instead of zero. The contour map for the block average log-transmissivity field is given as Figure 16. Also shown are the 15-lnT measurements used in the inverse solution.

The first run, again, uses only the 15-lnT measurements. The predicted lnT field is shown in Figure 17. Note that the presence of small-scale variability in the lnT measurements greatly reduces the ability of the procedure to obtain a detailed lnT prediction without the aid of head data. The prediction MSE's are

$$\text{MSE}(\varepsilon_\rho) = 0.254 \quad \text{MSE}(\varepsilon_\phi) = 14.0 \text{ m}^2$$

The next run applies the geostatistical solution presented here in section 2 using the 15-lnT measurements along with 15 head measurements. The predicted lnT field is shown in Figure 18. Note how much additional detail is obtained from the head data. The prediction MSE's are as follows:

$$\text{MSE}(\varepsilon_\rho) = 0.109 \quad \text{MSE}(\varepsilon_\phi) = 0.767 \text{ m}^2$$

Note again the dramatic improvement in these measures of fit quality.

Finally, the analytical solution is used to predict the lnT field based on the same 15 lnT and 15 head measurements. The resulting lnT contour map is shown as Figure 19. The prediction MSE's are

$$\text{MSE}(\varepsilon_\rho) = 0.148 \quad \text{MSE}(\varepsilon_\phi) = 7.06 \text{ m}^2$$

Again the analytical approach provides a better solution than that based on lnT measurements alone, but it can not match
Fig. 19. Contour map of predicted block-averaged log-transmissivities ($\theta_{1} = 0.4$ and $\theta_{2} = 0.4$). Prediction is based on 15 ln $T$ and 15 head measurements. The analytical solution is used.

the numerical solution which fully takes into account the effect of the aquifer boundaries.

The primary advantages of an analytical approach are that it is free of discretization error and it can provide a quick (and inexpensive) solution. The numerical approach has the advantage that it can incorporate more details into the model. Since the emphasis in the inverse problem is that of calibration of a groundwater flow model, a numerical (model specific) solution may indeed be desirable.

7. SUMMARY AND CONCLUSIONS

The goal of this paper is to develop and compare two separate applications of the geostatistical solution to the inverse problem in groundwater modeling. Both applications attempt to estimate the area averaged values of transmissivity throughout an aquifer based on the aquifer model and point measurements of piezometric head and transmissivity. Both applications utilize the same log-transmissivity (ln $T$) spatial variability model (equations (1) and (2)), and both applications use discretized equations of flow to obtain a model for the spatial variability of the hydraulic head in terms of the unknown parameters of the ln $T$ variability model. In the first application all five unknown parameters (in equations (1) and (2)) are simultaneously estimated using maximum likelihood estimation. Then, the transmissivity field is predicted using the Gaussian conditional mean and variance. In the second application the expected or mean head is linearized about the drift parameters. This allows filtering of the measurement means from the data, thereby reducing the bias in the parameter estimation step and allowing cokriging or extended cokriging to be used to estimate the ln $T$ field. The second model also demonstrates the inclusion of additional unknown parameters into the model.

The two applications of the geostatistical solution have been applied to hypothetical aquifer models to test their behavior and predictive capabilities. The results of a typical test is given here in section 4. Based on many runs made with several different transmissivity field realizations, several different vertical inflow conditions and several different measurement groups a number of conclusions are obtained.

The first conclusion is that the geostatistical solution (both applications) does provide a good estimate of the transmissivity or log-transmissivity field even in the presence of leakage into (or out of) the aquifer. This is best seen by comparing the predictions based on the geostatistical solution to predictions obtained without the use of head data (see Figures 2 through 7). Typically, the use of head data reduces the mean squared error of ln $T$ estimation by about 2 times and the mean squared error of estimation of the resulting head field by about 35 times (the values of 2 and 35 are median values).

The approach using Gaussian conditional mean estimation nearly always underestimates the covariance parameters $\theta_{i}$ and the ln $T$ estimation variance (equation (13)). This is due to the simultaneous estimation of mean and covariance parameters and also due to the uncertainty in the estimate of the ln $T$ drift ($\partial H_{01}/\partial x$ in equation (12)). The approach using extended cokriging was developed primarily to combat this problem and is seen to effectively eliminate the bias in the ln $T$ estimation variance.

To develop the extended cokriging approach the expected or mean head field was made linear in the drift parameters in the vicinity near the approximate values of these parameters. The error associated with this linearization is small. Some simulation runs indicate that the mean squared error of ln $T$ estimation increases on the average by about 5% in the move from the conditional mean approach to extended cokriging. However, it is not clear at this point whether this increase is statistically significant.

The extended cokriging approach also very effectively takes into account other unknown model parameters. This is demonstrated with an unknown leakage coefficient. The effect of each additional unknown included in the model is to reduce by one the number of independent measurements available in the parameter estimation step. The results shown in Table 1 indicate that no major loss of predictive capability occurs even when the leakage coefficient is left as an unknown.

This paper also demonstrates the results of using cokriging estimation to predict the transmissivity field in the Jordan (sandstone) Aquifer of Iowa. The predicted transmissivity contour map is shown as Figure 8. The advantage of a statistical solution to the inverse problem is that it also provides an estimate of the quality of the predicted field. This is shown in the two contour maps for the upper and lower 95% confidence intervals (Figures 9 and 10).

Finally, a comparison is made between the numerical approach developed here (the one using Gaussian condition mean estimation) and an analytical approach suggested by Dagan [1985]. The analytical approach is seen to provide a quick, inexpensive, and free of discretization error solution to the inverse problem. However, analytical solutions can be obtained only for simple region geometries, boundary conditions, and inputs. Numerical procedures may involve some discretization error but can handle much more complex cases.

APPENDIX

The goal of this appendix is to describe the method used to obtain derivatives of the mean head $H$ with respect to the drift parameters $\theta_{D}$. The differential equation describing the spatial variation of the mean head is (5) in the text. The derivative $\partial H/\partial H_{01}$ is designated as $H_{D1}$. To obtain the differential equation for $H_{D1}$ the following substitutions are made in (5). The parameter $\theta_{D1}$ is replaced by $\theta_{D1} + \theta_{D}$, and $H$ is replaced by $H + H_{D1}$. The results obtained for $H_{D1}$, $H_{D2}$, $H_{D3}$, and $H_{D4}$ ($H_{D4}$ is the mean head derivative with respect to the leakage coefficient multiplier) are as follows:

$$
\theta_{D2} \frac{\partial H_{D1}}{\partial x} + \frac{\partial^{2} H_{D1}}{\partial x^{2}} + \theta_{D3} \frac{\partial H_{D1}}{\partial y} + \frac{\partial^{2} H_{D1}}{\partial y^{2}} - \frac{K}{M} e^{-r} H_{D1}
$$


R. J. Hoeksema, Department of Engineering, Calvin College, Grand Rapids, MI 49506.

P. K. Kitanidis, St. Anthony Falls Hydraulic Laboratory, Department of Civil and Mineral Engineering, Mississippi River at Third Avenue Southeast, Minneapolis, MN 55414.

References


(Received September 4, 1984; revised January 24, 1985; accepted February 7, 1985.)