WATER RESOURCES RESEARCH, VOL. 31, NO. 10, PAGES 2411–2419, OCTOBER 1995

Quasi-linear geostatistical theory for inversing

Peter K. Kitanidis
Civil Engineering, Stanford University, Stanford, California

Abstract. A quasi-linear theory is presented for the geostatistical solution to the inverse problem. The archetypal problem is to estimate the log transmissivity function from observations of head and log transmissivity at selected locations. The unknown is parameterized as a realization of a random field, and the estimation problem is solved in two phases: structural analysis, where the random field is characterized, followed by estimation of the log transmissivity conditional on all observations. The proposed method generalizes the linear approach of Kitanidis and Vomvoris (1983). The generalized method is superior to the linear method in cases of large contrast in formation properties but informative measurements, i.e., there are enough observations that the variance of estimation error of the log transmissivity is small. The methodology deals rigorously with unknown drift coefficients and yields estimates of covariance parameters that are unbiased and grid independent. The applicability of the methodology is demonstrated through an example that includes structural analysis, determination of best estimates, and conditional simulations.

Introduction

Mathematical models of groundwater flow and transport require many parameters before they can be used to make predictions. For example, the conductivity or transmissivity values of all elements are required before the solution to the flow equation can be computed in a finite-element formulation. In practice, these parameters are inferred from the data in a process known as "model calibration," "parameter identification," "history matching," or "inverse modeling." The literature on the subject was quite voluminous when it was reviewed by *Yeh* [1986], and the number of publications on this topic has probably doubled by now. A recent book on the subject is that of *Sun* [1994]. Despite the many methods that have been proposed, research on this topic continues to be as strong as ever. In fact, the practical significance of parameter estimation is now more apparent than ever as models improve in their description of physical processes and as more data are collected through hydrogeologic and geophysical methods. The development of accurate parameter estimation methods holds the promise of allowing the full utilization of measurements, which are collected at a high cost, in achieving the best possible site characterization.

It is impossible to review in the space available all the worthwhile methods that have been presented. However, note that many of the currently available methods are related to classical estimation methods, that are relations or generalizations of the least squares or maximum likelihood approach [e.g., *Yeh and Yoon*, 1981; *Cooley*, 1982, 1983; *Clifton and Neuman*, 1982; *Kitanidis and Vomvoris*, 1983; *Carrera and Neuman*, 1986]. Because of the inherent difficulties associated with the estimation of spatial functions (or "fields") from limited and imperfect data, the crux of the problem is how to parameterize the distributed parameter system [*Emsellem and DeMarsily*, 1971; *Neuman*, 1973; *Gavalas et al.*, 1976], that is, how to transform the question into a well-posed estimation problem. In my judgment, the principal conceptual differences among available methods are in the parameterization.

This work is along the lines of the geostatistical approach [e.g., *Kitanidis and Vomvoris*, 1983; *Dagan*, 1985; *Hoeksema and Kitanidis*, 1984, 1985, 1989; *Rubin and Dagan*, 1987a, b; *Wagner and Gorelick*, 1989; *Hoeksema and Clapp*, 1990]. A principal characteristic of the geostatistical approach is the random-field parameterization that is commonly used in stochastic groundwater mechanics [*Dagan*, 1989; *Gelhar*, 1993] and that has been popularized in geophysics by the seminal work of *Matheron* [1971]. The important practical advantage of this parameterization is that it eliminates the need for arbitrary and inflexible assumptions that would otherwise need to be made for the purpose of reducing the number of effective parameters to be estimated, such as the domain to be subdivided into a small number of zones or the parameter to be represented as a polynomial of the spatial coordinates.

This work generalizes the geostatistical approach as a method of solving inverse problems involving spatially distributed parameters and process (e.g., flow or transport) equations. A general formulation of the problem is first presented. Then a practical, computationally efficient, and statistically sound method is developed for the solution of some estimation problems that may involve nonlinear relations between data and unknown parameters but where sufficient data are available to allow reasonably accurate estimates. The method is illustrated on a classical problem involving the estimation of a large-contrast conductivity function in one-dimensional (1-D) flow from sparse head and conductivity observations.

Methodology

Formulation

Let $\mathbf{s}$ be a spatially variable parameter, such as the log conductivity, that is represented as a random process. The rationale for such a model is that it is a practical way to represent the structure of the unknown function without making overly strong or restrictive assumptions, as is the case when the unknown spatial function is represented as a sum of de-

Paper number 95WR01945.
0043-1397/95/95WR-01945$05.00

terministic functions with adjustable coefficients. *Matheron* [1971] and others have argued that deterministic trend functions are not appropriate for the description of erratic variability, especially small-scale variability or random-walk type of variability; such variability is more suitably modeled through covariance functions or variograms. Geological processes are complex, and it is difficult to foresee precisely the structure of variability of parameters such as hydraulic conductivity, porosity, etc. In most geostatistical applications the simplest mean function (to wit, a constant mean) is conservatively adopted, and the structure of the process is described through the variogram.

The first step, which in geostatistics is known as structural analysis, is to characterize the random field by functions of a few parameters that describe the statistics of the field. The usual approach is to represent the mean as a linear function of $\boldsymbol{\beta}$ and the covariance as a known function of $\boldsymbol{\theta}$, where $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are the statistical or "structural" parameters of the process. Since the analysis will be carried out numerically, consider that the spatial process has been discretized so that $\mathbf{s}$ is the $n \times 1$ vector of the discretized values and

$$E[\mathbf{s}] = \mathbf{X}\boldsymbol{\beta} \tag{1}$$

where $\mathbf{X}$ is a known $n \times p$ matrix, $\boldsymbol{\beta}$ are p unknown drift coefficients, and $E[\ ]$ denotes the expected value. Furthermore, $\mathbf{s}$ has a covariance matrix

$$E[(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T] = \mathbf{Q}(\boldsymbol{\theta}) \tag{2}$$

that is considered a known function of parameters $\boldsymbol{\theta}$. The exponent T stands for matrix transpose. If there are two or more spatially variable parameters that need to be estimated (such as log transmissivity and storage coefficient), this case is included here by considering $\mathbf{s}$ as the aggregation of all spatially variable parameters.

There may also be other unknown parameters $\mathbf{r}$, such as boundary conditions or the intensity of sources or sinks. The $\mathbf{r}$ parameters will be included here for the sake of generality. However, the emphasis of this work is on the estimation of the spatial process and its structural parameters. The $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $\mathbf{r}$ parameters are treated as unknown constants and are supposed to be few in number, certainly far fewer than the observations. The observations are related to the unknown spatial process and the other parameters through

$$\mathbf{z} = \mathbf{h}(\mathbf{s}, \mathbf{r}) + \mathbf{v} \tag{3}$$

where $\mathbf{z}$ is the $m \times 1$ vector of observations. The observation error $\mathbf{v}$ is random with normal distribution, zero mean, and covariance matrix $\mathbf{R}$ that is fixed or a known function of an expanded set of parameters $\boldsymbol{\theta}$. The standard deviations of the measurement errors, which are the square root of the diagonal elements of $\mathbf{R}$, define how closely the observations should be reproduced.

Although the approach that will be presented is quite general, to fix ideas, consider as an example 1-D saturated flow without sources and sinks. We will estimate conductivity from "point" measurements of conductivity and head. In our formulation, $\mathbf{s}$ is the log conductivity, a spatial process with, for example, constant mean and exponential covariance function, and $\mathbf{r}$ is the head at the boundaries. Then, $\mathbf{X}$ is an $n \times 1$ vector of 1's, β is the process mean, and the covariance parameters $\boldsymbol{\theta}$ are the sill and length parameters of the exponential covariance function. If $\mathbf{s}$ and $\mathbf{r}$ were given, one could predict the "response" $\mathbf{h}(\mathbf{s}, \mathbf{r})$, which is the log conductivity at measurement locations and the head at measurement locations from a groundwater flow model (i.e., a method to solve the flow equation). Equation (3) simply states that the measurements are equal to the predicted responses plus random "observation error."

The vector $\mathbf{z}$ is a random vector, being a function of $\mathbf{s}$ and $\mathbf{v}$, which are random vectors. The probability distribution of $\mathbf{z}$ depends on the distribution of $\mathbf{s}$ and $\mathbf{v}$ and also on the function $\mathbf{h}$ and is generally hard to derive explicitly. Nevertheless, using the fact that $\mathbf{z}$ is jointly distributed with $\mathbf{s}$ (and, for notational simplicity, not always showing that every probability distribution function (pdf) is for some values of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\mathbf{r}$):

$$p(\mathbf{z}) = \int_s p(\mathbf{z}, \mathbf{s})\, d\mathbf{s} = \int_s p(\mathbf{z}|\mathbf{s})p(\mathbf{s})\, d\mathbf{s} \tag{4}$$

where

$$p(\mathbf{z}|\mathbf{s}) \propto |\mathbf{R}|^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))\right] \tag{5}$$

and $|\ |$ denotes matrix determinant. In terms of formulation, any probability distribution could be used for $p(\mathbf{s})$. This work will apply the theory of Gaussian random processes (subject, of course, to validation with data in any application):

$$p(\mathbf{s}) \propto |\mathbf{Q}|^{-1/2} \exp\left[-\tfrac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})\right] \tag{6}$$

Then

$$\begin{aligned} p(\mathbf{z}, \mathbf{s}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{r}) &\propto |\mathbf{R}|^{-1/2}|\mathbf{Q}|^{-1/2} \\ &\cdot \exp\left[-\tfrac{1}{2}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))\right] \\ &\cdot \exp\left[-\tfrac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})\right] \end{aligned} \tag{7}$$

$$\begin{aligned} p(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{r}) &\propto |\mathbf{R}|^{-1/2}|\mathbf{Q}|^{-1/2} \\ &\cdot \int_s \exp\left[-\tfrac{1}{2}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))\right] \\ &\cdot \exp\left[-\tfrac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})\right] d\mathbf{s} \end{aligned} \tag{8}$$

For consistency with other work in geostatistics, the unknown drift coefficients will be eliminated from the analysis. For the significance of this procedure, see *Hoeksema and Kitanidis* [1985] and *Kitanidis and Lane* [1985], who show that this elimination is needed to reduce bias in the estimation of the covariance parameters $\boldsymbol{\theta}$. Also, this way, one may use nonstationary models, such as intrinsic functions described through variograms or generalized covariance functions [*Matheron*, 1973; *Kitanidis*, 1983]. The elimination of $\boldsymbol{\beta}$ can be achieved by working with the restricted likelihood, obtained through averaging over all values of the drift coefficients $\boldsymbol{\beta}$:

$$\begin{aligned} p(\mathbf{z}|\boldsymbol{\theta}, \mathbf{r}) &= \int_s p(\mathbf{z}|\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{r})\, d\boldsymbol{\beta} \\ &\propto |\mathbf{R}|^{-1/2}|\mathbf{Q}|^{-1/2}|\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}|^{-1/2} \int_s I(\mathbf{s})\, d\mathbf{s} \end{aligned} \tag{9}$$

where the integrand is

$$I(\mathbf{s}) = \exp\left[-\tfrac{1}{2}\{(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r}))^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\mathbf{s}, \mathbf{r})) + \mathbf{s}^T\mathbf{G}\mathbf{s}\}\right] \tag{10}$$

$$\mathbf{G} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1} \tag{11}$$

The same averaging could be done with the other unknown parameters $\mathbf{r}$ but will not be carried out in this work.

One may then proceed to apply restricted maximum likelihood estimation, that is, find the values of the parameters $\boldsymbol{\theta}$ and $\mathbf{r}$ that maximize the expression $p(\mathbf{z}|\boldsymbol{\theta}, \mathbf{r})$ of (9). In practice, however, it is required to find an efficient way to handle the multiple integral, which cannot be computed analytically except in very special cases. Take into account that $\mathbf{s}$ can be a very large vector so that brute-force numerical integration is impractical. Furthermore, after estimating the unknown structural and other parameters, it is needed to estimate the pdf of $\mathbf{s}$ conditional on $\mathbf{z}$, $p(\mathbf{s}|\mathbf{z})$. Both problems are intractable in their exact form. Although up to this point the analysis has been exact, it is now necessary to make some practical yet accurate approximations.

Quasi-Linear Approximation

In many cases, the integrand $I(\mathbf{s})$ is a peaked function, with most of the value of the integral contributed from the area near the peak, which will be referred to as "local area." This happens when $\|\mathbf{Q}\|$ is small and/or $\|\mathbf{R}\|$ is small and/or there are many observations. Then $\mathbf{h}$ can be approximated by its low-order Taylor series near the neighborhood of the peak, provided that $\mathbf{h}$ is a smooth function of $\mathbf{s}$. This type of method for the asymptotic calculation of integrals is actually quite old, its original invention being credited to Laplace [see *Nayfeh*, 1981]. In the quasi-linear theory, we will linearize locally the $\mathbf{h}$ function.

To illustrate the principle involved in this approximation, consider the hypothetical case that $h(s) = s^2$ and

$$I(s) = 5\exp\left[-\frac{1}{0.02}\{(1.000 - s^2)^2 + (1.205 - s^2)^2\}\right] \tag{12}$$

The peak of the function is at $s_0 = 1.05$. About this value, one may linearize the measurement equation:

$$h(s) = s^2 \simeq s_0^2 + 2s_0(s - s_0) = -s_0^2 + 2s_0 s \tag{13}$$

Figure 1 depicts the actual $h(s)$ and its linearized approximation as well as the actual $I(s)$ and the one obtained if the linearized approximation is used. The important point is that the linearization, even though not a globally dependable approximation, is accurate in the local domain and quite satisfactory for purposes of approximating the definite integral (because the integrand vanishes anyway where the linearization is not valid, as is demonstrated in Figure 1b). Note that this approach is different from the linear theory [*Kitanidis and Vomvoris*, 1983; *Dagan*, 1985; etc.], where the linearization is effectively performed about the prior mean.

Thus the essential requirement for the quasi-linear approximation to work is that the local area must be small enough and the measurement function $\mathbf{h}$ must be sufficiently smooth for the measurement function to be accurately represented through its linearized form in the local area. The algorithm to carry out the quasi-linear approximation will be described in detail in the next section. Let it be emphasized here that the mathematical analysis is rigorous and the approach is conceptually straightforward: the objective is to optimize the maximum likelihood (equation (9)) with the integration being performed through a linearization about the point where the integrand has its peak (Laplace's method).

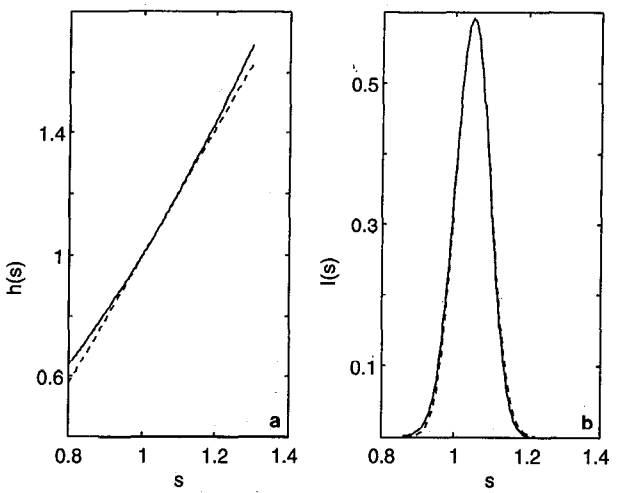

Figure 1. (a) Plot of $h(s)$ (solid line) and its linearized approximation (dashed line). (b) Plot of $I(s)$ (solid line) and its linearized approximation (dashed line).

Algorithm

The process is iterative because it involves nonlinear optimization. At the $(k + 1)$th iteration, start with the most recent guess of the parameters $\hat{\boldsymbol{\theta}}_k$ and $\hat{\mathbf{r}}_k$ as well as the spatial process, $\hat{\mathbf{s}}_k$. At the start, $k = 0$, one may use some rough estimates of the structural parameters from an experimental variogram analysis, similar to the linear geostatistical method [*Kitanidis and Vomvoris*, 1983]. Here, however, a preliminary estimate of the spatial process $\mathbf{s}$ is also required; one may use the mean $\mathbf{X}\boldsymbol{\beta}$ with some rough estimate of $\boldsymbol{\beta}$.

Step 1: Find the Peak of the Integrand

Considering for the moment that $\hat{\boldsymbol{\theta}}_k$ and $\hat{\mathbf{r}}_k$ are fixed, minimize with respect to $\tilde{\mathbf{s}}$:

$$\Psi = -\ln\left(I(\tilde{\mathbf{s}})\right)$$

$$= \tfrac{1}{2}[(\mathbf{z} - \mathbf{h}(\tilde{\mathbf{s}}, \mathbf{r}))^T\mathbf{R}^{-1}(\mathbf{z} - \mathbf{h}(\tilde{\mathbf{s}}, \mathbf{r})) + \tilde{\mathbf{s}}^T\mathbf{G}\tilde{\mathbf{s}}] \tag{14}$$

The appropriate iterative procedure is the Gauss-Newton method, obtained by linearizing $\mathbf{h}$ about the most recent estimate and solving the optimization problem, starting with $\tilde{\mathbf{s}}_0 = \hat{\mathbf{s}}_k$. There are different ways to implement numerically the Gauss-Newton method. It is advantageous to use a formulation that avoids the inversion of matrices of the size of $\mathbf{s}$ and other operations of the same order. The reason is that $\mathbf{s}$ could be very large if the discretization is fine. The recommended procedure is as follows:

Find the derivative of $\mathbf{h}$ about $\mathbf{s}$ at $\tilde{\mathbf{s}}_l$:

$$\mathbf{H}_l = (\partial\mathbf{h}/\partial\mathbf{s})|_{s=\tilde{\mathbf{s}}_l} \tag{15}$$

Define

$$\boldsymbol{\Sigma}_l = \mathbf{H}_l \mathbf{Q} \mathbf{H}_l^T + \mathbf{R} \tag{16}$$

$$\mathbf{z}_{0l} = \mathbf{z} - \mathbf{h}(\tilde{\mathbf{s}}_l) + \mathbf{H}_l \tilde{\mathbf{s}}_l \tag{17}$$

Solve the system of equations

$$\begin{bmatrix} \boldsymbol{\Sigma} & \vdots & \mathbf{H}_l\mathbf{X} \\ \cdots & \vdots & \cdots \\ (\mathbf{H}_l\mathbf{X})^T & \vdots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^T \\ \cdots \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_l\mathbf{Q} \\ \cdots \\ \mathbf{X}^T \end{bmatrix} \tag{18}$$

where $\boldsymbol{\Lambda}$ is an $m \times n$ matrix of coefficients and $\mathbf{M}$ is a $p \times n$ matrix of multipliers. Then

$$\tilde{\mathbf{s}}_{l+1} = \boldsymbol{\Lambda} \mathbf{z}_{0l} \tag{19}$$

Step 2: Estimate Structural Parameters

The iterations in step 1 converge to the value $\tilde{\mathbf{s}}$ where the integrand has its peak. For this value, linearize the $\mathbf{h}$ function with respect to $\mathbf{s}$ about the value $\tilde{\mathbf{s}}$, i.e.,

$$h(\mathbf{s}, \mathbf{r}) = h(\tilde{\mathbf{s}}) + \mathbf{H}(\mathbf{s} - \tilde{\mathbf{s}}) \tag{20}$$

where

$$\mathbf{H} = (\partial \mathbf{h}/\partial \mathbf{s})|_{s=\tilde{s}} \tag{21}$$

Compute

$$\boldsymbol{\Sigma} = \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R} \tag{22}$$

$$\mathbf{z}_0 = \mathbf{z} - \mathbf{h}(\tilde{\mathbf{s}}) + \mathbf{H}\tilde{\mathbf{s}} \tag{23}$$

$$\mathbf{V} = -\mathbf{X}\mathbf{M} + \mathbf{Q} - \mathbf{Q}\mathbf{H}^T\boldsymbol{\Lambda}^T \tag{24}$$

Here and in step 3, $\mathbf{z}_0$ and $\mathbf{H}$ are treated as constant.

Carrying out the integration and simplifying the obtained expressions,

$$p(\mathbf{z}|\boldsymbol{\theta}, \mathbf{r}) \propto |\boldsymbol{\Sigma}|^{-1/2} |\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}|^{-1/2} \exp\left[-\tfrac{1}{2}\mathbf{z}_0^T \cdot (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1})\mathbf{z}_0\right] \tag{25}$$

The problem of maximum likelihood estimation is equivalent to minimizing

$$L = \tfrac{1}{2}\ln|\boldsymbol{\Sigma}| + \tfrac{1}{2}\ln|\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}| + \tfrac{1}{2}\mathbf{z}_0^T(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1})\mathbf{z}_0 \tag{26}$$

$$\frac{\partial L}{\partial \theta_i} = \tfrac{1}{2} Tr\left[\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\right] - \tfrac{1}{2} Tr\left[(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}\right]$$

$$+ \tfrac{1}{2}\mathbf{z}_0^T\Bigg(-\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1} \cdot \mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}\Bigg)^{-1}$$

$$\cdot \mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}$$

$$+ \boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Sigma}^{-1}\Bigg)\mathbf{z}_0$$

Let

$$\boldsymbol{\Xi} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1}\mathbf{H}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{H}^T\boldsymbol{\Sigma}^{-1} \tag{27}$$

Then equations defining the gradient of L, $\mathbf{g}$, take the simple form

$$g_i = \frac{\partial L}{\partial \theta_i} = \tfrac{1}{2} Tr\left[\boldsymbol{\Xi}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\right] - \tfrac{1}{2}\mathbf{z}_0^T\left(\boldsymbol{\Xi}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Xi}\right)\mathbf{z}_0 \tag{28}$$

The Fisher information matrix $\mathbf{F}$ has ijth element

$$F_{ij} = E\left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right] = \tfrac{1}{2} Tr\left[\boldsymbol{\Xi}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i}\boldsymbol{\Xi}\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j}\right] \tag{29}$$

One can proceed to apply Gauss-Newton iterations:

$$\boldsymbol{\theta}_{l+1} = \boldsymbol{\theta}_l - \mathbf{F}^{-1}\mathbf{g} \tag{30}$$

For most cases, this scheme converges quickly. For strongly nonlinear cases it may be necessary to add a line search. For ill-conditioned $\mathbf{F}$ matrices it may be necessary to apply the Marquardt modification. (However, it is advisable to introduce few and identifiable $\boldsymbol{\theta}$ parameters so that $\mathbf{F}$ is a well-conditioned matrix.)

When the iterative procedure has converged, updated estimates of the structural parameters are obtained. Note also that mean square estimation error of the structural parameters is given approximately through $\mathbf{F}^{-1}$, through the well-known Cramer-Rao inequality [*Schweppe*, 1973; *Rao*, 1973].

Step 3: Estimate the Spatial Function

Iterations consisting of successive applications of steps 1 and 2 continue until the estimates of the structural parameters converge. Then $\mathbf{s}$ conditional on the $\mathbf{z}$ is approximately Gaussian, with mean given by $\tilde{\mathbf{s}}$ and covariance matrix given by $\mathbf{V}$. The justification for this statement follows. From

$$p(\mathbf{z}, \mathbf{s}) = p(\mathbf{z}|\mathbf{s})p(\mathbf{s}) \tag{31}$$

where $p(\mathbf{s})$ is from (6), and $p(\mathbf{z}|\mathbf{s})$ using the linearization is

$$p(\mathbf{z}|\mathbf{s}) \propto |\mathbf{R}|^{-1/2} \exp\left[-(\mathbf{z}_0 - \mathbf{H}\mathbf{s})^T\mathbf{R}^{-1}(\mathbf{z}_0 - \mathbf{H}\mathbf{s})\right] \tag{32}$$

where $\mathbf{z}_0 = \mathbf{z} - \mathbf{h}(\tilde{\mathbf{s}}) + \mathbf{H}\tilde{\mathbf{s}}$. Consequently, the joint pdf $p(\mathbf{z}, \mathbf{s})$ is (approximately) Gaussian, and as a consequence, the conditional pdf

$$p(\mathbf{s}|\mathbf{z}) = \frac{p(\mathbf{z}, \mathbf{s})}{p(\mathbf{z})} \tag{33}$$

is also Gaussian. One can then carry out all the calculations to show that $p(\mathbf{s}|\mathbf{z})$ has mean that is the same with the final estimate of $\tilde{\mathbf{s}}$ and a covariance matrix that is the same as $\mathbf{V}$.

Thus without additional calculations, the conditional mean and covariance of the spatial parameter are available.

Step 4: Conditional Simulations

One can generate realizations of $\mathbf{s}$ that are conditional on all observations. The practical meaning of conditional simulations is that they form a set (or "ensemble") of functions that, given available information, are equally likely answers to the question, "What is the actual parameter field $\mathbf{s}$?" A comprehensive review of conditional simulation methods in the context of the inverse problem is given by *Gutjahr et al.* [1994].

The first step toward generating a conditional realization is to generate an unconditional one. Efficient methods to accomplish this task are usually methods based on spectral decomposition. For illustration purposes, we mention the Choleski decomposition method, which is the most straightforward al-

though usually not the most efficient approach: decompose $\mathbf{Q} = \mathbf{C}\mathbf{C}^T$ and then compute

$$\mathbf{s}_{ui} = \mathbf{X}\boldsymbol{\beta} + \mathbf{C}\mathbf{u}_i \tag{34}$$

where $\mathbf{u}_i$ is a vector of independent, identically distributed normal variates. In this approach, there is no need to look for the "correct" estimates of $\boldsymbol{\beta}$ because the conditional realization finally produced is unaffected by the values of $\boldsymbol{\beta}$; thus one may set $\boldsymbol{\beta} = 0$ in (34). Also generate a realization $\mathbf{v}_i$ which is random with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}$. The realizations $\mathbf{s}_{ui}$ and $\mathbf{v}_i$ are generated independently.

Next, a conditional simulation $\mathbf{s}_{ci}$ is produced from $\mathbf{s}_{ui}$ by minimizing with respect to $\mathbf{s}_{ci}$:

$$(\mathbf{s}_{ci} - \mathbf{s}_{ui})^T\mathbf{G}(\mathbf{s}_{ci} - \mathbf{s}_{ui}) + (\mathbf{z} + \mathbf{v}_i - \mathbf{h}(\mathbf{s}_{ci}, \mathbf{r}))^T \cdot \mathbf{R}^{-1}(\mathbf{z} + \mathbf{v}_i - \mathbf{h}(\mathbf{s}_{ci}, \mathbf{r})) \tag{35}$$

The method can be interpreted intuitively as making the smallest possible modification to $\mathbf{s}_{ui}$ that will allow reproduction of the data within their measurement error (as specified by $\mathbf{R}$). The addition of the term $\mathbf{v}_i$ ensures that the uncertainty introduced by measurement error is accounted for in the generation of the conditional simulations.

It can be demonstrated through linearized error analysis that the $\mathbf{s}_{ci}$ ensemble has the mean and covariance matrix computed earlier, under the condition that the conditional mean and the conditional simulation are close enough that the same linearization applies to both of them. Thus this method can be rigorously shown to be consistent with the quasi-linear estimation method when the conditional variance is small, i.e., when the difference $\mathbf{s}_{ci} - \hat{\mathbf{s}}$ is small. Otherwise, this method of conditional simulations is an alternative nonlinear estimation method, different in principle from the quasi-linear approach described earlier. Also, instead of the standard approach of generating unconditional simulations $\mathbf{s}_{ui}$ according to the Gaussian distribution, one might in an ad hoc fashion generate $\mathbf{s}_{ui}$ realizations according to some other distribution and proceed to condition them on the data by minimizing (35).

After $\mathbf{s}_{ci}$ has been computed, such as the log conductivity, realizations of the dependent variable, such as head, are generated by solving the flow equation. This approach is straightforward and has the following features:

1. All measurements are reproduced within a margin specified by their measurement covariance matrix $\mathbf{R}$. For example, if the error is negligible ($\mathbf{R} \simeq \mathbf{0}$), the measurements should be reproduced exactly.
2. The joint conditional simulations of log conductivity and head satisfy the flow equation, not the linearized form of this equation.
3. The mean and covariance of the ensemble of the conditional realizations tend to be the same, with the conditional mean and covariance of $\mathbf{s}$ computed in the quasi-linear approach as the conditional variance becomes smaller.

Comparison With Other Methods

The approach proposed here is a generalization of the linear geostatistical approach [e.g., *Kitanidis and Vomvoris*, 1983] for the estimation of log conductivity from point measurements of log conductivity and head. The linear approach is obtained as a special case from the quasi-linear if we skip step 1, linearize at step 2 about the drift (unconditional mean) and estimate the structural parameters through optimization. Then estimates of the spatial process are determined from step 3, and conditional simulations are obtained from step 4. It is obvious that the linear approach is computationally advantageous because it avoids the nonlinear optimization problem of step 1 and also completes the computations in one round (steps 1 and 2). However, as our previous discussion illuminates, the linear approach is in principle less accurate than the quasi-linear approach when the variance of the log-conductivity is large because the linearization is done about the prior estimate, which may be quite different from the actual values. The linear approach was intended for small-variance cases.

Table 1. Conductivity Measurements

x	K
0.025	0.020
0.075	0.035
0.125	0.038
0.275	1.923
0.725	0.23
0.875	0.262

The method of *Carrera and Neuman* [1986] uses a geostatistical parameterization and handles large-variance cases, as does the method proposed herein. However, it differs in the estimation of parameters. The approach of Carrera and Neuman is maximum likelihood estimation of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\mathbf{r}$ simultaneously with maximum a posteriori probability (MAP) estimation of all parameters $\mathbf{s}$, that is, maximization of (7) with respect to $(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{r}, \mathbf{s})$. This method will be referred to as MAP. Unlike MAP, the method proposed herein maximizes (9) with respect to $(\boldsymbol{\theta}, \mathbf{r})$ and then computes the pdf of $\mathbf{s}$ conditional on $\mathbf{z}$. This difference is not minor for the following two reasons:

1. While the method proposed herein yields unbiased estimates of the covariance parameters $\boldsymbol{\theta}$, MAP cannot make this claim. *Kitanidis* [1994] has proved that the method used here yields unbiased estimates of $\boldsymbol{\theta}$, whereas MAP estimates of $\boldsymbol{\theta}$ are biased with the bias becoming larger as the size of $\mathbf{s}$ increases. Unless the domain is subdivided into a number of zones much smaller than the number of observations, the $\boldsymbol{\theta}$ estimates may yield suboptimal estimates of the $\mathbf{s}$ parameters and, perhaps more importantly, may underestimate the mean square error of estimation of $\mathbf{s}$.
2. The MAP estimation method cannot handle nonstationary intrinsic functions, such as those with a linear or power variogram. Indeed, while the MAP method is equipped to deal with geostatistical models described through covariance functions and drift functions, the method proposed herein also addresses the estimation of parameters of variograms or generalized covariance functions.

Illustrative Example

Consider steady 1-D flow without sources or sinks:

$$\frac{d}{dx}\left(K(x)\frac{d\phi}{dx}\right) = 0 \tag{36}$$

where ϕ is hydraulic head and K is hydraulic conductivity. Tables 1 and 2 contain measurements of K and ϕ at selected locations. All quantities have been made dimensionless. The domain of interest has coordinates from 0 to 1.

It is given as boundary or auxiliary conditions that $\phi(0) = 1$ and the discharge is $q = 0.12$. Notice that the downgradient

Table 2. Head Measurements

x	ϕ
0.25	0.302
0.30	0.299
0.55	0.293
0.60	0.292
0.70	0.254
0.75	0.228
0.80	0.198
0.85	0.181

boundary head $\phi(1)$ is not given. The objective is to estimate the log conductivity $Y = \ln K$ and the head ϕ and to evaluate the uncertainty associated with the estimation over a grid that covers the domain of interest with spacing $\Delta x = 1/20$. The data were simulated by solving the flow equation and the observations involve no other measurement error other than that caused by the finite machine accuracy (10^{-16}) and the computations.

Preliminary examination of the conductivity data shows variability in conductivity of 2 orders of magnitude over a relatively short distance, indicating a relatively large-contrast case. Initially, the log conductivity was represented as a realization of a stationary function with exponential covariance function

$$C_Y(x_i, x_j) = v \exp\left(-\frac{|x_i - x_j|}{l}\right) \tag{37}$$

where v and l are structural parameters to be estimated. The measurement errors were taken to be independent and identically distributed with variance σ_R^2; that is, the measurement error covariance matrix was taken equal to

$$\mathbf{R} = \sigma_R^2 \mathbf{I} \tag{38}$$

where $\mathbf{I}$ is the identity matrix.

Application of the method developed in this work yielded the following parameter estimates

$$v = 12.7 \qquad l = 1.0 \qquad \sigma_R = 5 \times 10^{-6} \tag{39}$$

However, the estimates of v and l were highly positively correlated with very large errors of estimation, as is indicated by the covariance matrix of estimation errors, which is approximated by the inverse of matrix $\mathbf{F}$ (equation (29)). In simple terms, a good fit was also obtained by estimates $v = 6.8$, $l = 0.5$. It is evident that the data are insufficient and the domain is too small to estimate both the variance and the scale parameters. However, estimation of their ratio is more feasible. For simplicity, it is preferable to adopt the linear-variogram model:

$$\gamma_Y(x_i, x_j) = \theta|x_i - x_j| \tag{40}$$

This model was adopted, and the estimate $\sigma_R = 5 \times 10^{-6}$ was maintained. Thus the problem of structural parameter estimation is reduced to estimating a single parameter, the slope of the linear variogram θ. The estimate of θ and its standard error SE (from (29)) is

$$\theta = 12.36 \qquad \text{se}(\theta) = 5.4 \tag{41}$$

Note that even with one structural parameter to be estimated, there is significant uncertainty about its exact value.

Figure 2 compares the actual log conductivity with that estimated. Note that the estimated log conductivity reproduces the observations and follows the actual log conductivity quite closely wherever there are observations. Furthermore, the approximate 95% confidence interval is shown (given by $\hat{\mathbf{s}} \pm 2$ diag $(\mathbf{V})$). The actual log conductivity happens to be within the confidence interval. The best estimate, when used as input into the flow equation, reproduces the head observations, as is demonstrated in Figure 3. Both log K and head observations are reproduced within 10^{-5}. From Figure 3 one can see that the head is estimated quite accurately from the available observations, because it is such a smooth (or "predictable") function, except where extrapolation of the head is attempted, between $x = 0.85$ and $x = 1$. Note that the last head observation is at $x = 0.85$.

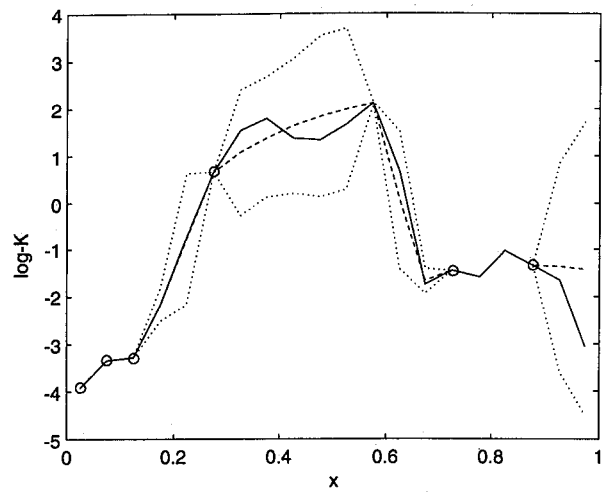

Figure 2. Log conductivity: actual (solid line), estimated (dashed line), approximate 95% confidence interval (dotted lines), and observations (open circles).

Three typical conditional realizations are shown in Figure 4 and the corresponding head functions are shown in Figure 5. The joint realizations of log K and head are consistent with the flow equation (36), reproduce both the log K and the head measurements appropriately (i.e., within their observation er-

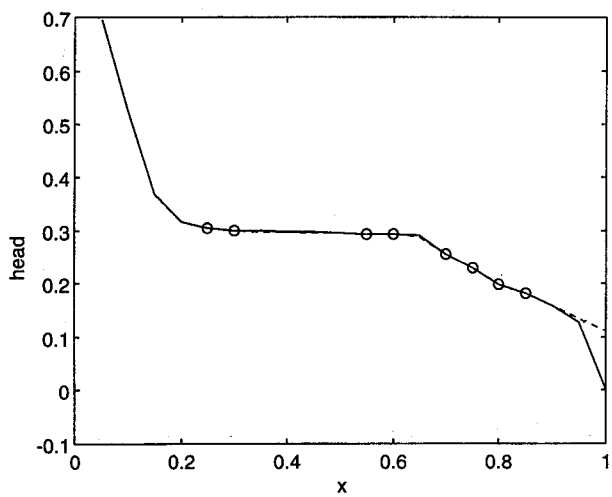

Figure 3. Head: actual (solid line), calculated using best estimates of conductivity (dashed line), and observations (open circles).

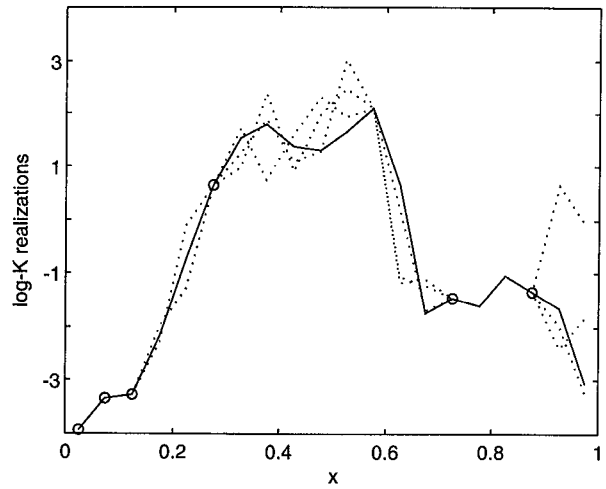

Figure 4. Actual log conductivity (solid line), three realizations (dotted lines), and observations (open circles).

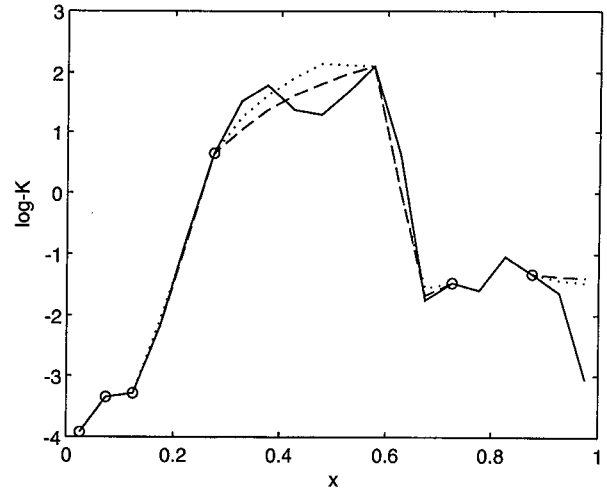

Figure 6. Log conductivity: Actual (solid line), best estimate (dashed line), mean of 400 conditional simulations (dotted line), and observations (open circles).

ror, which is about 10^{-5} in this case), and are consistent with the geostatistical structure that was identified previously.

Additionally, 400 conditional realizations were generated, and their statistics were compared with the conditional mean and covariance as computed through the algorithm. Figure 6 compares the conditional mean with the average of the 400 realizations. The agreement is excellent everywhere except where the conditional variance is large (refer to Figure 2). The difference is due mostly to the nonlinearity and only slightly to sampling error (i.e., the part that could be removed if a much larger number of realizations had been used). Perhaps the most interesting feature in Figure 6 is that it shows where the assumption of small conditional variance is less valid. That is, the difference occurs where the conditional variance is not as small as needed for the quasi-linear approximation to be accurate. Figure 7 compares the variance of log K at each block with the variance over the 400 realizations. By repeating the experiment with another set of 400 realizations, it was found that much of the scatter is due to sampling variability. Nevertheless, the computed variance may be slightly smaller than the variance over the set of the 400 realizations. When all 400 realizations are superposed on the same graph (Figure 8), their scatter has some striking similarities as well as some interesting differences compared with the confidence interval shown in Figure 2. A noticeable difference is that at some points where the conditional variance is relatively large, the distribution of the conditional simulations is asymmetric. Finally, Figure 9 shows the median and the nonparametrically determined 95% confidence intervals computed from the 400 realizations. The results are quite similar to those of Figure 2.

Discussion and Conclusions

A principal advantage of the geostatistical approach to the inverse problem is that it uses a parameterization that is flexible and suitable to account for erratic parameter variability without discretization of the domain into zones. The parameters that are fitted are the few parameters that describe the geostatistical structure. Identification of the structural parameters, which is a well-posed estimation problem, is the core of this approach. Once the structure has been resolved, the prob-

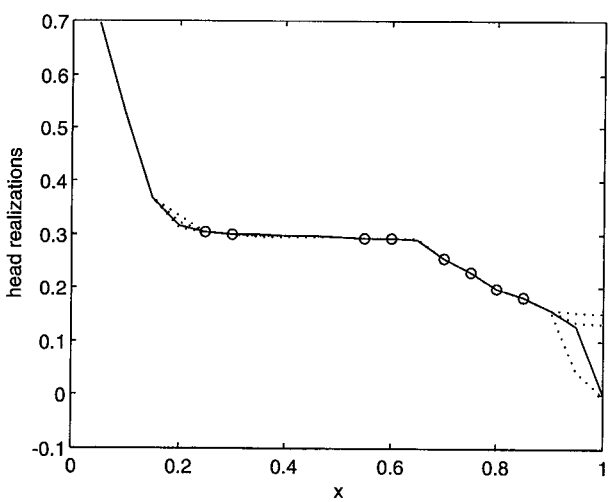

Figure 5. Actual head (solid line), three realizations (dotted lines), and measurements (open circles).

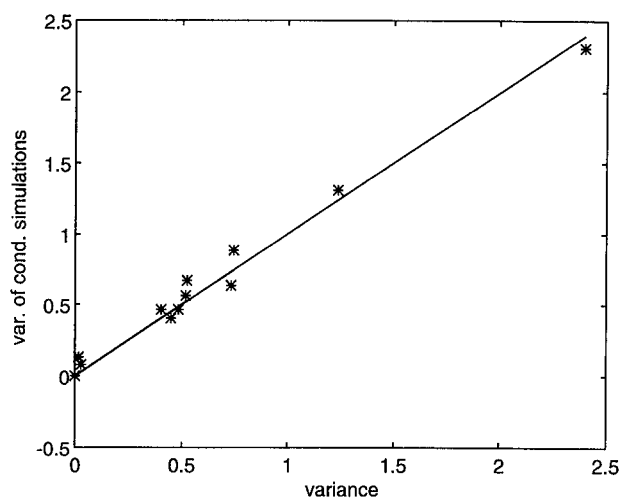

Figure 7. Comparison of estimation variance with the variance over the set of the 400 conditional realizations (asterisks). Points on the straight line would indicate perfect agreement.

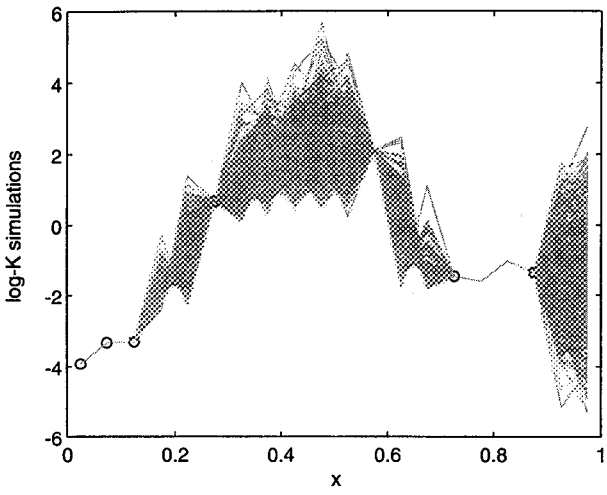

Figure 8. Plot of 400 conditional realizations.

ability distribution of the hydrogeologic parameters conditional on all observations can be derived. The hydrogeologic parameter estimates are stable and not strongly affected by the grid size. During both stages, all observations and the flow equations are taken into account. However, the linear method [*Kitanidis and Vomvoris*, 1983] assumed that the variability of the log conductivity or log transmissivity is low, which is typically interpreted to mean that the variance of the log conductivity is less than 1, because it linearized the flow equation about the prior estimate of the parameters.

Despite this limitation there is evidence that the linearized approach may perform well even in cases where the log conductivity variance is larger than 1. For example, *Zimmerman and Gallegos* [1993] simulated flow in a hypothetical formation and generated data (measured transmissivity and head). The data were subsequently used to estimate the transmissivity and the results were evaluated through calculation of travel times. In that study, linear methods were actually found to outperform nonlinear ones. Furthermore, although the consensus is that the linear method is appropriate only when the variance of log conductivity is less than 1, in my experience the method may work reliably for larger variances provided that the variation happens to be gradual or in the direction perpendicular to the streamlines. I believe that the linear theory performs the least satisfactorily when there is an abrupt change of conductivities along a streamline and when the head measurements have very small observation errors. I also believe that good fitting of the structural parameters to the data may be at least as important as accounting for nonlinearity.

To account for cases where the variance of the parameter field is large, a more general theory was presented in this work. Furthermore, a quasi-linear approximation was developed that is exact if the conditional variance of the parameter field is small. In practice, this approximation should work well if sufficient observations are available no matter how high the degree of variability. This is a nonlinear estimation method that unlike the linear method is guaranteed to reproduce all measurements (with a tolerance specified by their standard observation error). The probabilistic error analysis, however, assumes that the flow equation can be linearized about the best estimates, in the same sense as in popular nonlinear least squares methods [*Draper and Smith*, 1981; *Shah et al.*, 1978; *Cooley*, 1982; *Carrera and Neuman*, 1986; etc.], although this method differs from the conventional least squares approach in some important ways that are summarized in the next paragraph.

Like the linear theory, the more general methodology described here avoids the estimation of the drift coefficients and can deal with nonstationary functions described through variograms and generalized covariance functions. Also, the method is not maximum a posteriori probability estimation of both field and structural parameters, as is essentially the case for nonlinear least squares methods. The nonlinear method presented here is maximum likelihood estimation of the structural parameters only (after the random field has been averaged out) followed by estimation of the random field conditional on the observations. This is a significant difference from a statistical estimation perspective because the two methods yield different estimates of structural parameters.

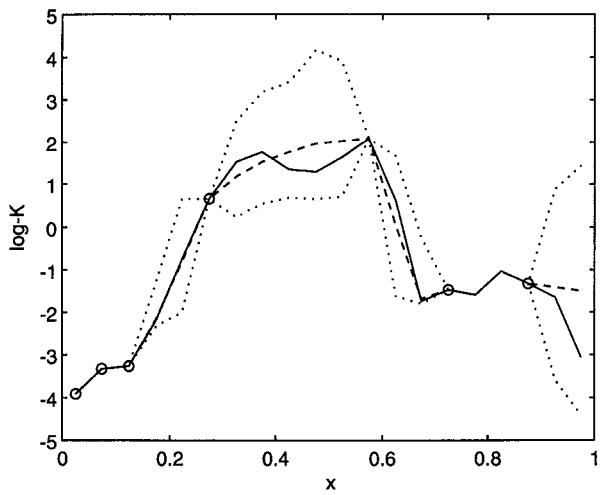

Figure 9. Actual log conductivity (solid line), median of 400 realizations (dashed line), nonparameteric 95% confidence intervals from 400 realizations (dotted lines), and observations (open circles).

Acknowledgments. Funding for this publication was provided by the office of Research and Development, U.S. Environmental Protection Agency, under agreement R-815738-01 through the Western Region Hazardous Substance Research Center. The content of this publication does not necessarily represent the views of the agency. Thanks to the reviewers for useful comments.

References

Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum likelihood method incorporating prior information, *Water Resour. Res.*, *22*(2), 199–210, 1986.

Clifton, P. M., and S. P. Neuman, Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona, *Water Resour. Res.*, *18*(4), 1215–1234, 1982.

Cooley, R. L., Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 1, Theory, *Water Resour. Res.*, *18*(4), 965–967, 1982.

Cooley, R. L., Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 2, Applications, *Water Resour. Res.*, *19*(3), 662–676, 1983.

Dagan, G., Stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem, *Water Resour. Res.*, *21*(1), 65–72, 1985.

Dagan, G., *Flow and Transport in Porous Media*, 465 pp., Springer-Verlag, New York, 1989.

Draper, N., and H. Smith, *Applied Regression Analysis*, 2nd ed., 709 pp., Wiley-Interscience, 1981.

Emsellem, Y., and G. DeMarsily, An automatic solution for the inverse problem, *Water Resour. Res.*, *7*(5), 1264–1283, 1971.

Gavalas, G. R., P. C. Shah, and J. H. Seinfeld, Reservoir history matching by Bayesian estimation, *Soc. Pet. Eng. J.*, *16*, 337–350, 1976.

Gelhar, L. W., *Stochastic Subsurface Hydrology*, 390 pp., Prentice-Hall, Englewood Cliffs, N. J., 1993.

Gutjahr, A., B. Bullard, S. Hatch, and L. Hughson, Joint conditional simulations and the spectral approach for flow modeling, *Stoch. Hydrol. Hydraul.*, *8*(1), 79–108, 1994.

Hoeksema, R. J., and R. B. Clapp, Calibration of groundwater flow models using Monte Carlo simulations and geostatistics, in *Model-CARE 90: Calibration and Reliability in Groundwater Modelling*, *IAHS Publ.*, *195*, 33–42, 1990.

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

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

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

Kitanidis, P. K., Statistical estimation of polynomial generalized covariance functions and hydrologic applications, *Water Resour. Res.*, *19*(4), 909–921, 1983.

Kitanidis, P. K., On the geostatistical approach to the inverse problem, interim report, Civil Eng., Stanford Univ., Stanford, Calif., 1994.

Kitanidis, P. K., and R. W. Lane, Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *J. Hydrol.*, *79*, 53–79, 1985.

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

Luenberger, D. G., *Introduction to Linear and Nonlinear Programming*, 356 pp., Addison-Wesley, Reading, Mass., 1973.

Matheron, G., *The Theory of Regionalized Variables and Its Applications*, 212 pp., Ecole de Mines, Fontainebleau, France, 1971.

Matheron, G., The intrinsic random functions and their applications, *Adv. Appl. Prob.*, *5*, 439–468, 1973.

Nayfeh, A. H., *Introduction to Perturbation Techniques*, 519 pp., John Wiley, New York, 1981.

Neuman, S. P., Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty, *Water Resour. Res.*, *9*(4), 1006–1021, 1973.

Rao, C. R., *Linear Statistical Inference and Its Applications*, 625 pp., John Wiley, New York, 1973.

Rubin, Y., and G. Dagan, Stochastic identification of transmissivity and effective recharge in steady groundwater flow, 1, Theory, *Water Resour. Res.*, *23*(7), 1185–1192, 1987a.

Rubin, Y., and G. Dagan, Stochastic identification of transmissivity and effective recharge in steady groundwater flow, 2, Case study, *Water Resour. Res.*, *23*(7), 1193–1200, 1987b.

Schweppe, F. C., *Uncertain Dynamic Systems*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

Shah, P. C., G. R. Gavalas, and J. H. Seinfeld, Error analysis in history matching: The optimum level of parametrization, *Soc. Pet. Eng. J.*, *18*, 219–228, 1978.

Sun, N.-Z., *Inverse Problems in Groundwater Modeling*, 337 pp., Kluwer Acad., Norwell, Mass., 1994.

Wagner, B. J., and S. M. Gorelick, Reliable aquifer remediation in the presence of spatially variable hydraulic conductivity: From data to design, *Water Resour. Res.*, *25*(10), 2211–2225, 1989.

Yeh, W. W-G., Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*(1), 95–108, 1986.

Yeh, W. W-G., and Y. S. Yoon, Aquifer parameter identifiability with optimum dimension in parameterization, *Water Resour. Res.*, *17*(3), 664–672, 1981.

Zimmerman, D. A., and D. P. Callegos, A comparison of geostatistically based inverse techniques for use in performance assessment analysis at the WIPP site: Results from test case No 1, in IV Annual International Conference, pp. 1426–1436, Am. Nucl. Soc., Lagrange Park, Ill., 1993.

P. K. Kitanidis, Civil Engineering, Stanford University, Stanford, CA 94305-4020.

(Received January 3, 1995; revised June 5, 1995;
accepted June 22, 1995.