WATER RESOURCES RESEARCH, VOL. 32, NO. 5, PAGES 1131–1161, MAY 1996

A reassessment of the groundwater inverse problem

Dennis McLaughlin

Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge

Lloyd R. Townley

CSIRO Division of Water Resources, Wembley, Western Australia

Abstract. This paper presents a functional formulation of the groundwater flow inverse problem that is sufficiently general to accommodate most commonly used inverse algorithms. Unknown hydrogeological properties are assumed to be spatial functions that can be represented in terms of a (possibly infinite) basis function expansion with random coefficients. The unknown parameter function is related to the measurements used for estimation by a "forward operator" which describes the measurement process. In the particular case considered here, the parameter of interest is the large-scale log hydraulic conductivity, the measurements are point values of log conductivity and piezometric head, and the forward operator is derived from an upscaled groundwater flow equation. The inverse algorithm seeks the "most probable" or maximum a posteriori estimate of the unknown parameter function. When the measurement errors and parameter function are Gaussian and independent, the maximum a posteriori estimate may be obtained by minimizing a least squares performance index which can be partitioned into goodness-of-fit and prior terms. When the parameter is a stationary random function the prior portion of the performance index is equivalent to a regularization term which imposes a smoothness constraint on the estimate. This constraint tends to make the problem well-posed by limiting the range of admissible solutions. The Gaussian maximum a posteriori problem may be solved with variational methods, using functional generalizations of Gauss-Newton or gradient-based search techniques. Several popular groundwater inverse algorithms are either special cases of, or variants on, the functional maximum a posteriori algorithm. These algorithms differ primarily with respect to the way they describe spatial variability and the type of search technique they use (linear versus nonlinear). The accuracy of estimates produced by both linear and nonlinear inverse algorithms may be measured in terms of a Bayesian extension of the Cramer-Rao lower bound on the estimation error covariance. This bound suggests how parameter identifiability can be improved by modifying the problem structure and adding new measurements.

1. Introduction

Inverse problems are likely to arise whenever mathematical models are used to explain or enhance observations. Examples may be found in fields as diverse as astronomy, medicine, meteorology, quantum mechanics, and hydrology. The generic inverse problem is concerned with the estimation of spatially variable model "parameters" which have physical significance but are difficult to measure. Under certain circumstances these parameters may be inferred from measurements of related "dependent variables." A typical example is the estimation of log hydraulic conductivity (a parameter) from scattered observations of piezometric head (a dependent variable). Over the years, hydrologists have become aware of the fact that such inverse problems are difficult to solve and may, in fact, be inherently "ill-posed" [*Yakowitz and Duckstein*, 1980; *Carrera and Neuman*, 1986a; *Dietrich and Newsam*, 1989, 1990]. At the same time, new developments in applied mathematics and other fields have provided conceptual insights and analytical tools which can help hydrologists deal with the inverse problems they encounter in practice.

Paper number 96WR00160.
0043-1397/96/96WR-00160$09.00

Groundwater inverse methods have been reviewed by *McLaughlin* [1975], *Yeh* [1986], *Kuiper* [1986], *Carrera* [1987], *Ginn and Cushman* [1990], and *Sun* [1994]. Other relevant reviews may be found in work by *Kubrusly* [1977], *Polis* [1982], *Lorenc* [1986], *Daley* [1991] and *Bennett* [1992]. In addition, there is an extensive inverse literature which spans many disciplines ranging from engineering applications to theoretical mathematics. It is not clear at first glance how all the concepts and methods described in this literature are related. As a start, it is convenient to characterize an inverse method by (1) the way it describes spatial variability (the "parameterization" it adopts), (2) the forward equation it uses to relate parameters to measurements, (3) the performance criterion it uses to define "good" parameter estimates, and (4) the solution technique it uses to find these estimates. Although all four factors are important, the inverse algorithms used by groundwater hydrologists differ most significantly in their approaches to parameterization. Parameterization deserves special consideration because it has a strong influence on the "well-posedness" of an inverse problem and on the physical plausibility of its solution.

Most groundwater inverse algorithms adopt either a blocked or a geostatistical (random field) description of spatial variability. The first approach divides the region of interest into a number of discrete blocks which are believed to correspond to distinct geologic units [see, e.g., *Carrera and Neuman*, 1986a, b, c; *Cooley*, 1977, 1979, 1982, 1983]. Each block is characterized by a set of spatially uniform hydrogeologic properties which are treated as parameters in an appropriate inverse problem. The geostatistical alternative views the properties of interest as stationary random fields which vary relatively smoothly over space [see, e.g., *Hoeksema and Kitanidis*, 1984; *Dagan*, 1985]. Although the two approaches are based on different parameterizations, they both treat hydrogeologic properties as spatial functions. This suggests that it should be possible to formulate a general inverse theory which encompasses both the blocked and geostatistical alternatives, as well as hybrid methods which lie between these extremes.

This paper shows how the methods of functional analysis may be used to develop a general groundwater inverse theory. The estimated parameters in this theory are scalar functions of location rather than vectors of spatially discretized variables. A functional approach to the inverse problem offers several advantages. It relates the blocked and geostatistical approaches and provides a framework for evaluating particular inverse algorithms proposed in the literature. It suggests how new inverse algorithms can be derived from alternative geologically motivated parameterizations. Finally, it brings the powerful tools of functional analysis to bear on the problems of ill-posedness and identifiability, and it suggests methods for improving the stability of inverse algorithms.

The spatially variable parameters of most interest in groundwater inverse applications include hydraulic conductivity, groundwater recharge and discharge, fluxes and piezometric heads on designated boundaries, and chemical rate coefficients. All of these parameters are uncertain, but some may be more important (have a greater effect on predictions) than others in any given application. Although a general discussion of the groundwater inverse problem would consider all potentially important parameters, our present purpose is better served by a more focused discussion. Therefore we concentrate on the problem of estimating large-scale variations in log hydraulic conductivity from point measurements of log conductivity and head. This is a reasonable choice for a review paper since hydraulic conductivity estimation has long been the primary focus of groundwater inverse research. In any case, the concepts we develop for the conductivity estimation problem extend naturally to problems where boundary conditions, chemical rate coefficients, and other spatially variable parameters must also be estimated from field data. Examples are provided in the work of *Cooley* [1977, 1979, 1982, 1983], *Carrera and Neuman* [1986a, b, c], *Rubin and Dagan* [1987a, b], and *Townley and Wilson* [1989], among others.

In the next section we discuss some important concepts and terms used in the rest of the paper. We then formulate and solve a functional (maximum a posteriori) version of the groundwater inverse problem. Next we show how a number of different groundwater inverse algorithms can be derived from or related to the maximum a posteriori solution. We conclude with a discussion of several important issues which are closely related to inverse estimation.

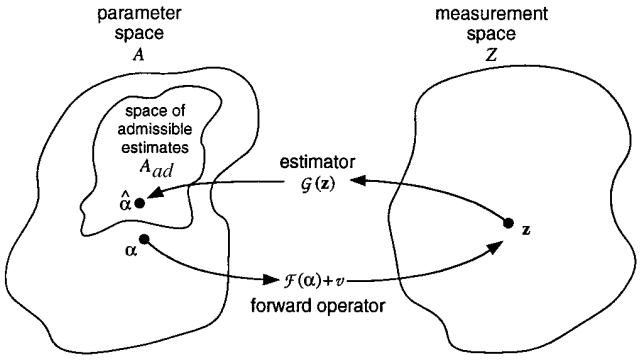

Figure 1. Relationship between parameter and measurement spaces for the groundwater inverse problem.

2. Background

2.1. Inversion and Well-Posedness

In groundwater inverse problems the estimated parameters are usually hydrogeologic properties and the measured dependent variables are quantities such as piezometric head, solute concentration, or temperature. We assume here that the dependent variables are functions of location and time but that the parameters are only functions of location. The M discrete measurements used for estimation can be collected in a vector $\mathbf{z}$ with scalar components $z_1, z_2, \cdots, z_M$. It is often natural to express the relationship between the measurement vector and a single scalar parameter function $\alpha(\mathbf{x})$ in the following "forward" form [*Tikhonov and Arsenin*, 1977; *Banks and Kunisch*, 1989; *Anger*, 1990]:

$$\mathbf{z} = \mathcal{F}(\alpha) + \mathbf{v} \tag{1}$$

where $\alpha(\mathbf{x})$ lies in a normed function space A, $\mathbf{z}$ lies in an M-dimensional Euclidean vector space Z, and the forward operator $\mathcal{F}$ is a functional that maps α to $\mathbf{z}$. The M-dimensional measurement error vector $\mathbf{v}$ also lies in Z. The more general multiparameter case can be handled by defining $\alpha(\mathbf{x})$ to be a vector of unknown functions. In the remainder of this paper italic fonts indicate scalars and scalar functions, bold lower case fonts represent vectors, bold upper case fonts indicate matrices or tensors, script fonts indicate scalar functionals, and bold script fonts indicate vectors of functionals. Subscripted fonts are generally scalar elements of vectors or matrices.

The basic goal of inverse estimation is to identify a generalized inverse operator $\mathcal{G}$ which maps the measurement vector to an estimate of α:

$$\hat{\alpha} = \mathcal{G}(\mathbf{z}) = \mathcal{G}[\mathcal{F}(\alpha) + \mathbf{v}] \tag{2}$$

The estimate $\hat{\alpha}(\mathbf{x})$ lies in a normed function space A_{ad} consisting of all admissible estimates of $\alpha(\mathbf{x})$. The estimate is selected to be "close" to the true parameter function $\alpha(\mathbf{x})$, where the definition of closeness remains to be defined. Figure 1 illustrates the definitions of the operators $\mathcal{F}$ and $\mathcal{G}$. If there is no measurement error ($\mathbf{v} = 0$) and the functional $\mathcal{F}$ is invertible on (A, Z) an obvious choice for $\mathcal{G}$ is $\mathcal{F}^{-1}$:

$$\hat{\alpha}(\mathbf{x}) = \mathcal{F}^{-1}(\mathbf{z}) = \mathcal{F}^{-1}[\mathcal{F}(\alpha)] = \alpha(\mathbf{x}) \tag{3}$$

In this special case the estimate is identical to the true parameter function and $A_{ad} = A$. If measurement noise is present or if the forward operator is not invertible, it is generally not possible to estimate $\alpha(\mathbf{x})$ perfectly.

It is important to know whether or not a given inverse problem will yield an acceptable solution in practice. This question is addressed by *Tikhonov and Arsenin* [1977], who extended *Hadamard*'s [1952] concept of "well-posedness" to the general inverse problem stated above. An inverse problem is well-posed if it satisfies the following three requirements [*Tikhonov and Arsenin*, 1977]: (1) For every measurement vector $\mathbf{z} \in Z$ there exists a parameter solution $\alpha \in A_{ad}$. (2) The parameter solution is unique. (3) The inverse problem is stable on the spaces (A_{ad}, Z); that is, small changes in the measurement produce small changes in the parameter solution, where "small" is defined with respect to the norms of A_{ad} and Z.

The first two of these requirements insure that a unique solution exists in the strict mathematical sense. The third requirement insures that the inverse solution is physically meaningful (not overly sensitive to measurement error).

The well-posedness of an inverse problem depends on the form of the inverse operator and on the definitions of the parameter and measurement spaces. An ill-posed problem can often be replaced by a well-posed problem if the definitions of $\mathcal{F}$, A_{ad}, and/or Z are modified [*Banks and Kunisch*, 1989]. Although there is no general theory for constructing well-posed inverse problems, groundwater hydrologists have developed many methods for dealing with ill-posedness. Some of the most successful include imposing stress on the groundwater system (e.g., pumping), using transient data, adding more or new kinds of measurements, and constraining the set of admissible parameter functions. It is important to note that ill-posedness does not imply that an inverse problem is meaningless. It merely indicates that the problem formulation must be modified or supplemented in some way before an acceptable solution can be obtained. The relevant issues can be illustrated with a few simple examples which introduce many of the concepts discussed in this paper.

Consider one-dimensional steady state saturated flow in the region defined by $x \in [0, L]$. If the piezometric head is specified at each boundary, the relevant flow equation is

$$\frac{\partial}{\partial x}\left[K(x)\,\frac{\partial h(x)}{\partial x}\right] = 0 \qquad x \in [0, L] \tag{4}$$

with boundary conditions

$$\begin{aligned} h(x) &= h_0 \qquad x = 0 \\ h(x) &= h_L \qquad x = L \end{aligned} \tag{5}$$

where $K(x)$ is the hydraulic conductivity and $h(x)$ is the piezometric head at location x. We assume that the unknown parameter $\alpha(x) = \ln K(x)$ lies in the space $A = A_{ad} = L_2[0, L]$ of functions which are square integrable on $[0, L]$. Also, we assume for the moment that noise-free head measurements are available and differentiable everywhere so that the measurement vector $\mathbf{z}$ can be replaced by a measurement function $z(x)$ which lies in the space $Z = C_1[0, L]$ of functions which are differentiable on $[0, L]$. The solution to (4) at any x can be written as

$$\begin{aligned} z(x) = h(x) &= \mathcal{F}(\alpha)(x) \\ &= h_0 + \frac{h_L - h_0}{\displaystyle\int_0^L e^{-\alpha(\xi)}\,d\xi}\int_0^x e^{-\alpha(\xi)}\,d\xi \qquad x \in [0, L] \end{aligned} \tag{6}$$

The $\mathcal{F}(\alpha)$ obtained for this problem is not invertible since $h(x)$ remains unchanged everywhere in $[0, L]$ if $\alpha(x)$ is shifted by an arbitrary constant. The problem is ill-posed because a given $z(x)$ does not yield a unique $\alpha(x)$. Also, note that the forward operator is nonlinear in α since $\mathcal{F}(\alpha_1 + \alpha_2)$ does not equal $\mathcal{F}(\alpha_1) + \mathcal{F}(\alpha_2)$ for all admissible α_1 and α_2 (this comment also applies if $\mathcal{F}$ is written as a functional operating on K rather than $\ln K$).

Now suppose that the downstream head boundary condition is replaced by the following specified flux condition:

$$K(x)\,\frac{\partial h(x)}{\partial x} = Q \qquad x = L \tag{7}$$

where Q is a known flux. The resulting solution is

$$z(x) = h(x) = \mathcal{F}(\alpha)(x) = h_0 + Q\int_0^x e^{-\alpha(\xi)}\,d\xi \qquad x \in [0, L] \tag{8}$$

The forward operator is now invertible on (A, Z) so long as Q is nonzero. A unique parameter estimate can be obtained from

$$\hat{\alpha}(x) = \mathcal{F}^{-1}(z) = \ln\left[\frac{1}{Q}\,\frac{\partial z(x)}{\partial x}\right] \tag{9}$$

The ill-posed problem has been made well-posed by forcing the system with a specified flux (i.e., by modifying the forward operator to make it invertible).

If head measurements are available only at M discrete points (as specified in the beginning of this section), the measurement space Z is an M-dimensional Euclidean space, and $\mathcal{F}$ is no longer invertible because an infinite number of different $\alpha(x)$ functions can reproduce any given set of M head measurements. In this case the problem can be made well-posed by changing the definition of either the admissible parameter space A_{ad} or the measurement space Z. We briefly consider both possibilities in the following paragraphs.

Suppose that two noise-free head measurements z_1 and z_2 are available at locations $x_1 = L/4$ and $x_2 = 3L/4$. The problem can then be made well-posed if $\alpha(x)$ is constrained to have the following form:

$$\alpha(x) = a_1\phi_1(x) + a_2\phi_2(x) \qquad x \in [0, L] \tag{10}$$

where

$$\begin{aligned} \phi_i(x) &= 1 \qquad \text{if } x \in [(i-1)L/2,\ iL/2) \qquad i = 1, 2 \\ \phi_i(x) &= 0 \qquad \text{otherwise} \qquad i = 1, 2 \end{aligned} \tag{11}$$

This is equivalent to defining A_{ad} to be the function space spanned by the two linearly independent basis functions ϕ_1 and ϕ_2. If (10) is substituted for $\alpha(\mathbf{x})$ in (8) and $z(x)$ is evaluated at x_1 and x_2, the result is a system of two equations in the two unknowns a_1 and a_2. The unique solution is

$$\hat{a}_1 = \ln\left[\frac{QL}{4(z_1 - h_0)}\right] \tag{12}$$

$$\hat{a}_2 = \ln\left[\frac{QL}{4(z_2 - 2z_1 + h_0)}\right] \tag{13}$$

The inverse problem has been made well-posed by modifying the definition of A_{ad} so that $\mathcal{F}(\alpha)$ is invertible on (A_{ad}, Z). Note that $\phi_i(\mathbf{x})$ need not be continuous since the inverse

solution for the constrained point measurement problem does not require differentiation of $z(\mathbf{x})$.

An alternative is to construct a "measured" head function by fitting a differentiable function (e.g., a second-order polynomial) to the upstream head boundary condition and the two measured heads at x_1 and x_2. This has the effect of "embedding" the original two-dimensional Euclidean measurement space in a function space spanned by the basis functions $\{1, x, x^2\}$. The fitted measurement function may be written as

$$z(x) = h(x) = h_0 + b_1 x + b_2 x^2 \tag{14}$$

where

$$b_1 = \frac{2}{3L}[9z_1 - z_2 - 8h_0] \tag{15}$$

$$b_2 = -\frac{8}{3L^2}[3z_1 - z_2 - 2h_0] \tag{16}$$

This head measurement function may be inserted into (9) to give a unique parameter estimate. In this case the problem has been made well-posed by modifying the definition of Z so that the forward operator is invertible on (A_{ad}, Z).

Note that problem modifications such as those discussed above are not intended to produce perfect estimates of the true parameter function. They are simply methods for obtaining a unique inverse solution which is consistent with the original problem statement. In practical applications, constraints on the parameter and measurement spaces should be chosen to reflect physical understanding of the hydrogeologic system under investigation. This is the best way to insure that solutions to modified inverse problems are geologically reasonable as well as unique.

So far our discussion has focused on problems with noise-free measurements and on estimators which are derived from the inverse operator $\mathcal{F}^{-1}$. This so-called direct approach to inversion has been applied to groundwater problems by *Nelson* [1960, 1961], *Kleinecke* [1971], *Neuman* [1973], and *Scott* [1992]. Most inverse research suggests that $\mathcal{F}^{-1}$ is not generally the best choice for $\mathcal{G}$ when noise is present [*Tikhonov and Arsenin*, 1977; *Tarantola*, 1987; *Banks and Kunisch*, 1989]. Other estimators may yield estimates which are, in some sense, "closer" to the true parameter value. Generally speaking, the best estimators are those which minimize the effect of the noise term in (2) while capturing the dominant features of the parameter function. One possibility is the estimator obtained by minimizing the following "regularized" least squares performance index [*Wahba*, 1990; *Fitzpatrick*, 1991]:

$$\mathcal{J}(\alpha) = [\mathbf{z} - \mathcal{F}(\alpha)]^T \mathbf{W}_v [\mathbf{z} - \mathcal{F}(\alpha)] + \int_D \int_D [\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})] W_\alpha(\mathbf{x}, \boldsymbol{\xi}) \cdot [\alpha(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\, d\mathbf{x}\, d\boldsymbol{\xi} \tag{17}$$

where $\bar{\alpha}(\mathbf{x})$ is a "prior" estimate of the unknown parameter function, $\mathbf{W}_v$ is a positive definite weighting matrix, and $W_\alpha(\mathbf{x}, \boldsymbol{\xi})$ is a positive-definite weighting function. In this case the estimate is constrained to lie in a function space A_{ad} with a squared norm given by [*Fitzpatrick*, 1991]:

$$\|\alpha\|^2_{A_{ad}} = \int_D \int_D \alpha(\mathbf{x}) W_\alpha(\mathbf{x}, \boldsymbol{\xi}) \alpha(\boldsymbol{\xi})\, d\mathbf{x}\, d\boldsymbol{\xi} \tag{18}$$

The first (measurement fit) term in (17) penalizes deviations of model predictions from measurements, while the second (regularization) term penalizes deviations of estimates from $\bar{\alpha}(\mathbf{x})$, which can be viewed as a reasonable "first guess" for $\alpha(\mathbf{x})$. The regularization term helps make the problem well-posed by keeping the estimate "close" to the prior, where "closeness" is measured in terms of the A_{ad} norm [*Tikhonov and Arsenin*, 1977; *Banks and Kunisch*, 1989; *Chavent*, 1991]. Once the weighting factors in (17) are selected, variational methods may be used to find the parameter function that minimizes $\mathcal{J}(\alpha)$. Equation (17) is the basis for the functional inverse algorithm described in the next section.

Inverse problems based on minimization of (17) are generally better posed then those based on direct inversion techniques. In fact, (17) will always yield a unique and stable minimum $\hat{\alpha}(\mathbf{x}) = \bar{\alpha}(\mathbf{x})$ when $\mathbf{W}_v = 0$. Unfortunately, this is a degenerate solution which does not depend on the measurements at all! In more realistic situations where $\mathbf{W}_v$ is nonzero the least squares problem may have a unique global minimum, but this minimum may be difficult to find (see work by *Chavent* [1991] for a detailed discussion of this point). In such cases, ill-posedness manifests itself in the form of performance indices which have many local minima or which are relatively insensitive to changes in $\alpha(\mathbf{x})$. So, although least squares inverse algorithms are able to deal with measurement error, they still must be applied with caution.

2.2. Upscaling, Effective Properties, and the Measurement Process

The simple examples discussed above do not consider how the hydraulic conductivity function might vary over space at real field sites. This is an important issue which we must address in order to define more precisely just what it is we wish to estimate. It is now widely recognized that hydraulic conductivity can vary by orders of magnitude over scales ranging from a few centimeters to hundreds of meters or more. This is confirmed by field data obtained from air and water permeameters, flowmeter measurements, piezometer tests, and larger-scale pumping tests (see work by *Gelhar* [1993, pp. 284–297] for a brief review with references). We clearly cannot expect to estimate centimeter-scale conductivity variations from point head and conductivity measurements spaced tens of meters apart. In fact, it is not obvious why we would even want to do so. Most groundwater flow investigations are concerned with large-scale variations in the piezometric head (e.g., drawdown in the vicinity of a pumping field) or with aggregate large-scale fluxes (e.g., total discharge to a river). In such cases, small-scale fluctuations about regional trends are important only to the extent that they influence point measurements.

These ideas can be made more precise if we suppose that the values of log conductivity and head at a given point can be expressed as the sum of a "large-scale" component or trend and a "small-scale" component or fluctuation, as shown in Figure 2:

$$\alpha_p(\mathbf{x}) = \alpha_l(\mathbf{x}) + \alpha_s(\mathbf{x}) \tag{19}$$

$$h_p(\mathbf{x}, t) = h_l(\mathbf{x}, t) + h_s(\mathbf{x}, t) \tag{20}$$

Here the p, l, and s subscripts refer to point, large-scale, and small-scale values, respectively. The large-scale log conductivity and head are qualitatively defined to be the components of the point variables that we can expect to estimate from available measurements. The corresponding small-scale values are

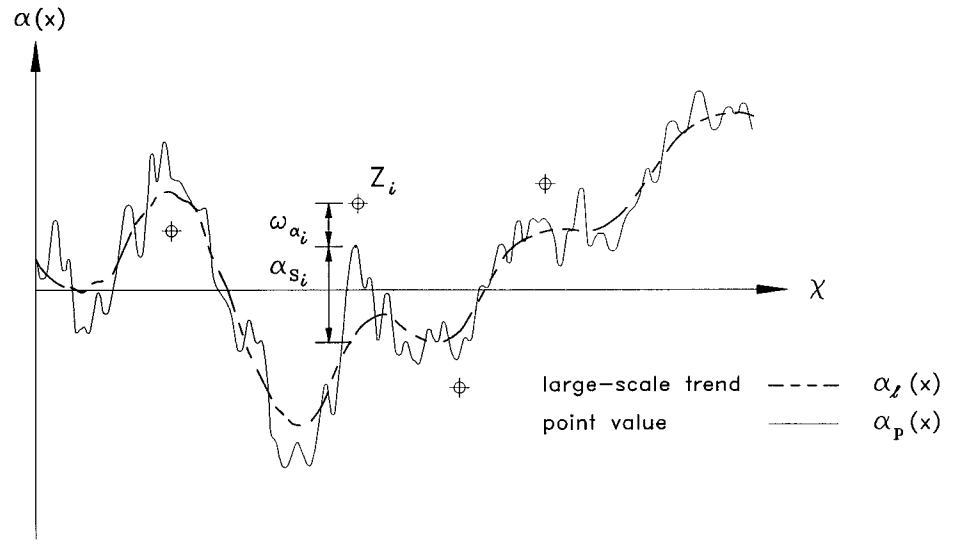

Figure 2. Large- and small-scale components of a hypothetical log hydraulic conductivity profile.

unidentifiable deviations from the large-scale trends. Clearly, the distinction between the two scales is application-dependent. Although it is possible to imagine more than two scales of variability in some situations, the simple decomposition proposed above is sufficient for present purposes.

It is convenient to adopt a stochastic (Bayesian) view of spatial variability and assume that each component of the head or log conductivity is a distinct random field characterized by its own statistical properties [*Rajaram and McLaughlin*, 1991]. This assumption facilitates our discussion of the various inverse methods used in groundwater hydrology. If the mean of the small-scale component is chosen to be 0, then the large-scale component is just the (nonstationary) mean of the point value:

$$\alpha_l(\mathbf{x}) = E_{\alpha_s}[\alpha_p(\mathbf{x})] \tag{21}$$

$$h_l(\mathbf{x}, t) = E_{h_s}[h_p(\mathbf{x}, t)] \tag{22}$$

where it is understood that the expectations E_{α_s} and E_{h_s} are taken over the probability densities of the small-scale fields. We assume that the large-scale component of the log conductivity has a constant mean $E_{\alpha_l}[\alpha_l] = E_{\alpha_l}\{E_{\alpha_s}[\alpha_p(\mathbf{x})]\}$, where E_{α_l} is taken over the probability density of α_l. Fluctuations of α_l and α_s about their respective means are characterized by the large- and small-scale covariances $C_{\alpha_l}(\mathbf{x}, \boldsymbol{\xi})$ and $C_{\alpha_s}(\mathbf{x}, \boldsymbol{\xi})$. Note that α_l and α_s are generally nonstationary random functions with covariances that depend on two spatial coordinates $\mathbf{x}$ and $\boldsymbol{\xi}$. The prior statistics of these random functions can be estimated from field data or inferred in other ways, depending on one's perspective. Inference of prior log conductivity statistics is an important practical issue which we consider further in section 5.1. For now, we will assume that these statistics are known.

A number of investigators have examined the problem of deriving an equation for the large-scale head when the log conductivity statistics are given. If certain assumptions are made, this upscaled equation can have the same general form as the familiar point groundwater flow equation (see (35)). The hydraulic conductivity appearing in the upscaled equation is, however, an "effective" conductivity tensor which typically depends on both the large- and small-scale properties of the log conductivity [*Gelhar and Axness*, 1983; *Rubin and Gomez-Hernandez*, 1989; *Durlovsky*, 1991; *Indelman and Dagan*, 1993; *Beckie et al*., 1994]. An example is the expression derived by *Gelhar and Axness* [1983] for the case of three-dimensional flow in a statistically isotropic medium:

$$[\mathbf{K}_{eff}(\mathbf{x})]_{ij} = e^{\alpha_l(\mathbf{x})}\left[1 + \frac{\sigma^2_{\alpha_s}}{6}\right]\delta_{ij} \qquad i, j = 1, 2, 3 \tag{23}$$

where δ_{ij} is equal to 1 if $i = j$, and 0 if $i \neq j$. This effective conductivity depends on the variance of the small-scale log conductivity $\sigma^2_{\alpha_s}$ as well as the value of the large-scale log conductivity $\alpha_l(\mathbf{x})$ at $\mathbf{x}$. Anisotropic extensions of (23) also depend on the directional correlation scales of $\alpha_s(\mathbf{x})$ [*Gelhar*, 1993, pp. 111–113]. Since we assume here that the large-scale log conductivity is a random field, the effective conductivity is also a random field, with statistical properties that can be derived from those of α_l.

Other effective conductivity formulas can be derived from other upscaling theories. In applications where there is no explicit use of an upscaling theory, the effective conductivity is often assumed to be an isotropic tensor with all diagonal elements equal to the geometric mean, so $[\mathbf{K}_{eff}(\mathbf{x})]_{ij} = \exp[\alpha_l(\mathbf{x})]\delta_{ij}$. Since the large-scale log conductivity α_l appears in most commonly used effective conductivity expressions it is the parameter that we emphasize in our discussion of the inverse problem. In some applications it may be useful to also estimate small-scale statistics (such as $\sigma^2_{\alpha_s}$) in the inverse procedure.

The head and log conductivity measurements used to estimate the large-scale log conductivity are typically defined over a smaller scale (i.e., have a smaller "support") than the characteristic scales of $\alpha_l(\mathbf{x})$ or $h_l(\mathbf{x})$. The most extreme examples are measurements derived from soil samples or piezometers that average over scales of a few centimeters. For all practical purposes, such measurements are point observations. Equations (19) and (20) can be used to develop the following expressions for point measurements of log conductivity and head at $(\mathbf{x}_i, t_i)$:

$$z_{\alpha i} = \alpha_{pi} + \omega_{\alpha i} = \alpha_{li} + \alpha_{si} + \omega_{\alpha i} \tag{24}$$

$$z_{hi} = h_{pi} + \omega_{hi} = h_{li} + h_{si} + \omega_{hi} \tag{25}$$

where $\alpha_{pi} = \alpha_p(\mathbf{x}_i)$, $h_{pi} = h_p(\mathbf{x}_i, t_i)$, etc. The additive noise terms $\omega_{\alpha i}$ and ω_{hi} are included to account for the effect of instrumentation and recording errors. From the perspective of inverse estimation the small-scale fluctuations α_{si} and h_{si} can also be regarded as noise since they obscure the large-scale trends α_{li} and h_{li}, which are of primary interest. For this reason it is convenient to combine the small-scale fluctuation and instrumentation-recording error terms in each measurement equation. If we drop the l subscripts on large-scale variables to simplify notation, the log conductivity and head point measurement equations become

$$z_{\alpha i} = \alpha_i + v_{\alpha i} \tag{26}$$

$$z_{hi} = h_i + v_{hi} \tag{27}$$

where $\alpha_i = \alpha_{li}$, $h_i = h_{li}$, and the composite noise terms are $v_{\alpha i} = \alpha_{si} + \omega_{\alpha i}$ and $v_{hi} = h_{si} + \omega_{hi}$. The properties of $v_{\alpha i}$ and v_{hi} depend on the nature of small-scale variability in the vicinity of the sampling point and on the technology used to collect the measurements. It should be noted that the analysis outlined here for point measurements can be generalized to provide explicitly for temporal and spatial averaging. This is accomplished by integrating (24) or (25) over appropriate time and space intervals [*Bencala and Seinfeld*, 1979; *McLaughlin*, 1995].

In this section we have identified some of the basic issues which must be considered when developing a practical groundwater inverse procedure. In particular, we have seen that the success of inverse estimation depends strongly on the way the problem is posed. It is important to have realistic expectations which recognize the limitations of the data available for estimation. With this in mind, we can now formulate a general approach to the problem.

3. A Functional Approach to the Groundwater Inverse Problem

Our functional approach to the inverse problem requires us to specify (1) a stochastic parameterization which describes how the large-scale log conductivity varies over space, (2) a forward equation which relates the unknown log conductivity to the measurements used for estimation, (3) an estimation performance index, and (4) a solution algorithm for computing the optimum estimate. Each of these topics is discussed in one of the following subsections. We begin with a Bayesian description of spatial variability which can be applied to both blocked and geostatistical formulations of the inverse problem. We then derive a generic forward equation from the classic groundwater flow equation and an analysis of the measurement process. This is followed by a discussion of the maximum a posteriori (or "most probable") approach to estimation. When the unknown log conductivity and measurement error are jointly Gaussian the maximum a posteriori estimate minimizes a generalized least squares performance index similar to (17). We conclude with a discussion of variational methods for solving this functional minimization problem. The general problem formulation developed in this section provides a convenient framework for our discussion (in section 4) of several widely used groundwater inverse algorithms.

3.1. Parameterization

In section 2 we identified the large-scale log conductivity $\alpha(\mathbf{x})$ as the parameter to be estimated in our inverse procedure. We now consider in more detail how this unknown parameter function might be characterized. Since $\alpha(\mathbf{x})$ lies in a function space A its properties will be determined by the way we structure this space. In particular, if $\alpha(\mathbf{x})$ is random, this must be reflected in the definition of the norm of A [see *Kuelbs et al*., 1972; *Kuo*, 1975; *Fitzpatrick*, 1991]. If A satisfies suitable mathematical requirements, any random function in A can be expressed in a discrete form, as an expansion in a set of linearly independent "basis functions":

$$\alpha(\mathbf{x}) = \sum_{i=1}^{N} a_i \phi_i(\mathbf{x}) \tag{28}$$

where $\phi_i(\mathbf{x})$ is the ith basis function, a_i is the corresponding basis function coefficient, and N is the (possibly infinite) dimension of A. When N is finite we may assemble the basis functions and their coefficients in the N-dimensional vectors $\boldsymbol{\phi}(\mathbf{x})$ and $\mathbf{a}$ and then write the expansion as a dot product:

$$\alpha(\mathbf{x}) = \boldsymbol{\phi}^T(\mathbf{x})\mathbf{a} \tag{29}$$

Truncated finite-dimensional basis expansions are typically introduced when inverse algorithms are discretized for numerical implementation. The choice of a basis is not unique, although some alternatives may be more useful than others in any given application. Figure 3 shows several methods for expressing a large-scale log conductivity function as a weighted sum of simple basis functions. The cell and node-based approaches break the function into many small parts, each characterized by a different unknown coefficient. The regionalized method divides the function into a few constant blocks which are meant to correspond to distinct geological units. The polynomial/spline approach expresses the function as a smooth function with a relatively small number of unknown coefficients. The influence function approach expresses the function as a sum of elementary functions (e.g., covariances) which decay from measurement values towards the prior mean. All of these methods have been used in groundwater inverse algorithms.

Once a set of basis functions is selected, uncertainties in $\alpha(\mathbf{x})$ can only come from uncertainties in the a_i coefficients, which must be random variables if $\alpha(\mathbf{x})$ is a random function. The statistical properties of the parameter function and its basis function coefficients are closely related. This is most clearly illustrated for the case of Gaussian functions and coefficients, which are completely characterized by their respective means and covariances. The mean and covariance of $\alpha(\mathbf{x})$ may be expressed in terms of the basis functions and the prior a_i statistics:

$$\bar{\alpha}(\mathbf{x}) = \sum_{i=1}^{N} \bar{\alpha}_i \phi_i(\mathbf{x}) = \boldsymbol{\phi}^T(\mathbf{x})\bar{\mathbf{a}} \tag{30}$$

$$\begin{aligned} C_\alpha(\mathbf{x}, \boldsymbol{\xi}) &= \overline{[\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})][\alpha(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \overline{[a_i - \bar{a}_i][a_j - \bar{a}_j]}\phi_i(\mathbf{x})\phi_j(\boldsymbol{\xi}) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} [\mathbf{C}_a]_{ij}\phi_i(\mathbf{x})\phi_j(\boldsymbol{\xi}) \\ &= \boldsymbol{\phi}^T(\mathbf{x})\mathbf{C}_a\boldsymbol{\phi}(\boldsymbol{\xi}) \end{aligned} \tag{31}$$

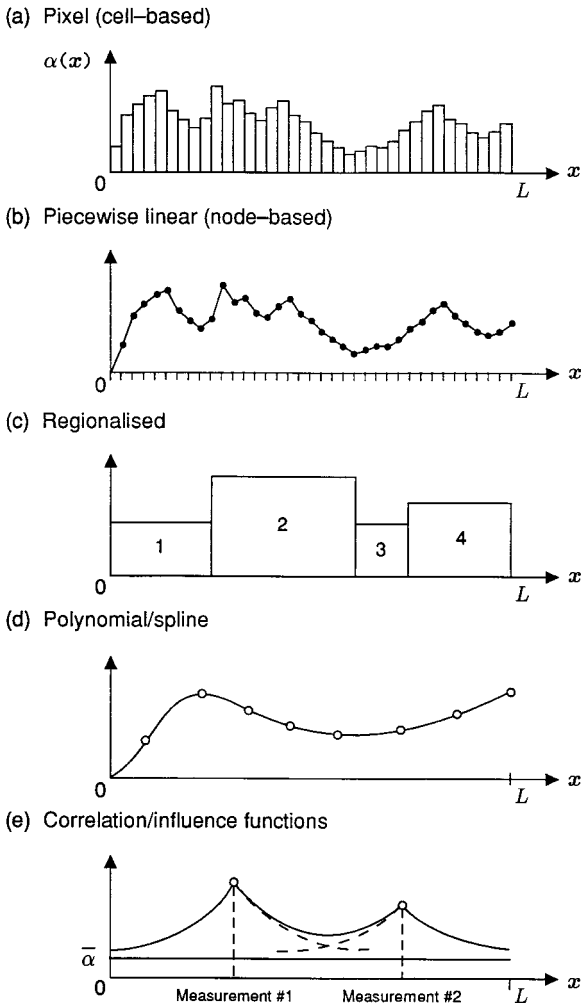

Figure 3. Some alternative methods for describing spatial variability with finite-dimensional basis function expansions.

where $\bar{\alpha}_i$ is the mean of a_i and $[\mathbf{C}_a]_{ij}$ is the covariance between a_i and a_j. The concise vector expressions which appear in the final equalities of these equations apply only when N is finite.

Equations (30) and (31) indicate how a functional parameterization (the mean and covariance of $\alpha(\mathbf{x})$) can be derived from a given discrete parameterization (a set of basis functions and the means and covariances of the corresponding a_i's). We can also derive a discrete parameterization from a given functional parameterization when $\alpha(\mathbf{x})$ is mean-squared continuous. The basis functions are obtained by solving the following Karhunen-Loeve integral equation for $\phi_i(\mathbf{x})$ [*Papoulis*, 1984, pp. 303–305; *Braud et al.*, 1993]:

$$\int_D C_\alpha(\mathbf{x}, \boldsymbol{\xi})\phi_i(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \lambda_i\phi_i(\mathbf{x}) \qquad (32)$$

$$i = 1, \cdots, N$$

(no summation over i). This equation defines an eigenvalue problem where λ_i is the eigenvalue associated with the ith eigenfunction $\phi_i(\mathbf{x})$. The eigenfunctions (sometimes called "empirical orthogonal functions") form an orthonormal basis for A. The coefficients (a_i's) associated with the Karhunen-Loeve basis functions are uncorrelated random variables with means and covariances given by

$$\bar{a}_i = \int_D \phi_i(\mathbf{x})\bar{\alpha}(\mathbf{x})\, d\mathbf{x} \qquad (33)$$

$$[\mathbf{C}_a]_{ij} = \lambda_i\delta_{ij} \qquad (34)$$

(no summation over i). The dimension of the parameter space (the number of basis functions) is generally infinite if $\alpha(\mathbf{x})$ is a stationary random field. However, since the magnitude of the eigenvalues decreases with i, a reasonable approximation of $\alpha(\mathbf{x})$ can often be obtained with a finite N. It is worth noting that discrete approximations to the Karhunen-Loeve basis functions can be constructed from the eigenvectors of $\mathbf{C}_a$ (see section 4.2.1).

The above paragraphs suggest that functional and discrete descriptions of random parameters such as the large-scale log conductivity are interchangeable in the Gaussian case, provided that the number of basis functions can be infinite. This enables us to derive blocked and geostatistical groundwater inverse algorithms from a common mathematical framework which can be expressed in either a functional or a discrete form. Since the functional form is usually more convenient in geostatistical applications, while the discrete form is usually more convenient in blocked applications, we will move freely between the two alternatives. Nevertheless, it is important to remember that the functional and discrete parameterizations described here are based on the same Bayesian estimation philosophy and lead to inverse algorithms which are, in many cases, formally equivalent. This will become more apparent as our discussion develops.

3.2. The Forward Equation

The forward equation relates the unknown parameter $\alpha(\mathbf{x})$ to the various measurements used in the estimation algorithm. In order to derive a forward equation we must first consider particular kinds of measurements. The relationship between $\alpha(\mathbf{x})$ and the large-scale piezometric head at any given time and location is governed by the following upscaled groundwater flow equation [*Bear*, 1979]:

$$S\frac{\partial h}{\partial t} - \nabla \cdot \mathbf{K}_{eff}(\alpha)\nabla h = Q \qquad \mathbf{x} \in D,\ 0 < t \le T \qquad (35)$$

with the following initial and boundary conditions:

$$h(\mathbf{x}, t) = h_0(\mathbf{x}) \qquad \mathbf{x} \in D,\ t = 0$$

$$h(\mathbf{x}, t) = h_b(\mathbf{x}, t) \qquad \mathbf{x} \in \partial D_d,\ 0 \le t \le T$$

$$-\mathbf{K}_{eff}(\alpha)\nabla h(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}) = q_b(\mathbf{x}, t) \qquad \mathbf{x} \in \partial D_n,\ 0 \le t \le T$$

where h is the large-scale head and $\mathbf{K}_{eff}(\alpha)$ is the effective hydraulic conductivity tensor. The large-scale log conductivity α enters the flow equation through the effective conductivity tensor, which may be derived from an appropriate upscaling theory (see the discussion accompanying (23)).

In (35) the variables Q and S are the effective recharge rate and storage coefficient, respectively. The spatial domain D may be one-, two-, or three-dimensional. A head boundary condition is specified on the portion of the domain boundary denoted by ∂D_d while a flux condition is specified on the portion denoted by ∂D_n, with $\mathbf{n}(\mathbf{x})$ indicating the unit outward pointing vector normal to ∂D_n at $\mathbf{x}$. The functions $h_0(\mathbf{x})$, $h_b(\mathbf{x}, t)$, and

$q_b(\mathbf{x}, t)$ are the large-scale initial head, boundary head, and boundary outflow, respectively. In practice the boundary may include a free surface, but we assume here that the location of the boundary is known and time-invariant. We also assume that Q, S, h_0, h_b, and q_b are all known. These simplifications facilitate our discussion but are not crucial (see section 5.1 for further discussion of this point).

An explicit solution to (35) may be written, at least formally, in terms of the flow equation Green's function $G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha)$ evaluated at some specified location and time $(\mathbf{x}, t)$ [*Courant and Hilbert*, 1953; *Greenberg*, 1971]:

$$\begin{aligned} h(\mathbf{x}, t) &= \mathcal{F}_h(\alpha)(\mathbf{x}, t) \\ &= \int_0^T \int_D G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha) Q(\boldsymbol{\xi}, \tau)\, d\boldsymbol{\xi}\, d\tau \\ &\quad + \int_D S(\boldsymbol{\xi}) G(\mathbf{x}, \boldsymbol{\xi}, t, 0|\alpha) h_0(\boldsymbol{\xi})\, d\boldsymbol{\xi} \\ &\quad + \int_0^T \int_{\partial D_d} \mathbf{K}_{eff}(\alpha) \nabla_{\xi} G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha) \\ &\quad \cdot \mathbf{n}(\boldsymbol{\xi}) h_b(\boldsymbol{\xi}, \tau)\, d\boldsymbol{\xi}\, d\tau \\ &\quad + \int_0^T \int_{\partial D_n} G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha) q_b(\boldsymbol{\xi}, \tau)\, d\boldsymbol{\xi}\, d\tau \end{aligned} \tag{36}$$

where $\mathcal{F}_h(\alpha)(\mathbf{x}, t)$ is a nonlinear functional (the head prediction operator) that maps the function $\alpha(\mathbf{x})$ to the scalar $h(\mathbf{x}, t)$. Equation (36) shows that this functional may be decomposed into four integral terms, each associated with one of the source terms or auxiliary conditions introduced in (35). Note that $G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha)$ depends on the entire $\alpha(\mathbf{x})$ function and that the spatial gradient is taken with respect to the $\boldsymbol{\xi}$ variable. Since it is usually not possible to obtain a closed form expression for the Green's function, $h(\mathbf{x}, t)$ is typically obtained by solving the flow equation numerically rather than by actually evaluating the integrals in (36). Nevertheless, it is helpful to introduce an explicit integral expression for the prediction operator. This enables us to apply the techniques of functional analysis and variational calculus to the inverse problem.

Equation (36) may be used to derive a forward operator which relates $\alpha(\mathbf{x})$ to a vector $\mathbf{z}_h$ of discrete point head measurements taken at the M_h measurement points $(\mathbf{x}_1, t_1), \cdots, (\mathbf{x}_{M_h}, t_{M_h})$. The predicted large-scale heads at the measurement points may be assembled in an M_h-dimensional vector functional $\mathcal{F}_h(\alpha)$ (the head measurement operator) with components given by

$$h_i = h(\mathbf{x}_i, t_i) = \mathcal{F}_{hi}(\alpha) = \mathcal{F}_h(\alpha)(\mathbf{x}_i, t_i) \tag{37}$$

It follows from (25) that the ith component of the point head measurement equation may be written as

$$z_{hi} = h_i + v_{hi} = \mathcal{F}_{hi}(\alpha) + v_{hi} \tag{38}$$

where v_{hi} is the composite measurement noise associated with measurement i. This noise term accounts for the effects of small-scale fluctuations of the actual point head from the large-scale trend and for instrumentation/recording errors (see the discussion following (24)).

An analogous forward operator can be identified for point measurements of log hydraulic conductivity. To do this we define a linear functional $\mathcal{F}_\alpha(\alpha)$ (the log conductivity prediction operator) which samples the value of the large-scale log conductivity at a particular location $\mathbf{x}$:

$$\alpha(\mathbf{x}) = \mathcal{F}_\alpha(\alpha)(\mathbf{x}) = \int_D \delta(\mathbf{x} - \boldsymbol{\xi}) \alpha(\boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{39}$$

where $\delta(\ \ )$ is the Dirac delta function, which is introduced to enable $\alpha(\mathbf{x})$ to be written as a functional. The predicted large-scale log conductivities at the M_α log conductivity measurement points $\mathbf{x}_{M_h+1}, \cdots, \mathbf{x}_{M_h+M_\alpha}$ may be assembled in an M_α-dimensional vector functional $\mathcal{F}_\alpha(\alpha)$ (the log conductivity measurement operator) with components given by

$$\alpha_i = \alpha(\mathbf{x}_i) = \mathcal{F}_{\alpha i}(\alpha) = \mathcal{F}_\alpha(\alpha)(\mathbf{x}_i) \tag{40}$$

It follows from (24) that the ith component of the point log conductivity measurement equation may be written as

$$z_{\alpha i} = \alpha_i + v_{\alpha i} = \mathcal{F}_{\alpha i}(\alpha) + v_{\alpha i} \tag{41}$$

where $v_{\alpha i}$ is the composite measurement noise associated with measurement i. As mentioned earlier, nonpoint measurements may be accommodated by integrating the point measurement equations over appropriate time or space intervals.

The head and log conductivity measurement equations may be combined to give a composite expression for all of the $M = M_h + M_\alpha$ measurements used for inverse estimation:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_h \\ \mathbf{z}_\alpha \end{bmatrix} = \begin{bmatrix} \mathcal{F}_h(\alpha) \\ \mathcal{F}_\alpha(\alpha) \end{bmatrix} + \begin{bmatrix} \mathbf{v}_h \\ \mathbf{v}_\alpha \end{bmatrix} = \mathcal{F}(\alpha) + \mathbf{v} \tag{42}$$

This forward equation has the same form as (1).

3.3. Formulation of the Estimation Problem

Our task in this section is to derive an estimator $\mathcal{G}(\mathbf{z})$ which maps the measurement $\mathbf{z}$ to a "good" estimate of the large-scale log conductivity function. Although there are many ways to define "good" estimate, the most convenient choice for present purposes is the maximum a posteriori approach, which is based on a Bayesian interpretation of parameter uncertainty [*Jazwinski*, 1970; *Schweppe*, 1973, pp. 328–334; *Bard*, 1974, pp. 72–73]. In order to introduce this approach we first consider the case where the dimension N of the parameter space is finite, so that the large-scale log conductivity function may be written in the vector form given in (29):

$$\alpha(\mathbf{x}) = \boldsymbol{\phi}^T(\mathbf{x}) \mathbf{a} \tag{43}$$

We focus for the moment on the estimation of the basis function coefficient vector $\mathbf{a}$, recognizing that an estimate of $\alpha(\mathbf{x})$ can always be constructed from (43) once an estimate of $\mathbf{a}$ is available.

The maximum a posteriori estimate $\hat{\mathbf{a}}$ is the value of $\mathbf{a}$ which is most likely, given a particular set of measurements $\mathbf{z}$. To simplify notation we suppose that $\mathbf{a}$ and $\mathbf{v}$ are independent random vectors and that $\mathbf{v}$ is zero mean. In this case the maximum a posteriori estimate is defined more precisely by the following condition:

$\hat{\mathbf{a}}$ is the $\mathbf{a}$ which maximizes

$$p_{a|z}(\mathbf{a}|\mathbf{z}) = \frac{p_{z|a}(\mathbf{z}|\mathbf{a}) p_a(\mathbf{a})}{p_z(\mathbf{z})} = \frac{p_v[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})] p_a(\mathbf{a})}{p_z(\mathbf{z})} \tag{44}$$

where $p_a(\ )$ and $p_{a|z}(\ )$ are the prior and a posteriori probability densities of $\mathbf{a}$, $p_{z|a}(\ )$ is the conditional density of $\mathbf{z}$ given $\mathbf{a}$, and $p_z(\ )$ and $p_v(\ )$ are the measurement and measurement error probability densities. The first equality of (44) is the classical statement of Bayes rule. The second equality follows from (42), (43), and the independence of $\mathbf{a}$ and $\mathbf{v}$. Note that the argument of $\mathcal{F}$ is written as $\boldsymbol{\phi}^T\mathbf{a}$ so that the dependence on $\mathbf{a}$ is explicit. The estimate obtained from (44) is the mode (the "most likely" value) of the a posteriori probability density [*Jazwinski*, 1970]. Since this estimate is conditioned on the measurements it is frequently called the "conditional mode."

In most applications of maximum a posteriori estimation the prior and measurement error probability densities which appear in (44) are assumed to be Gaussian. We follow this precedent here. It should be noted that there is limited experimental evidence suggesting that the first-order probability density of the log hydraulic conductivity is Gaussian [*Freeze*, 1975; *Hoeksema and Kitanidis*, 1985]. While this evidence does not guarantee that higher-order (joint) densities of $\mathbf{a}$ are Gaussian, it does lend credibility to the multivariate Gaussian assumption. All of the groundwater inverse algorithms discussed in subsequent sections are based, either explicitly or implicitly, on the Gaussian assumption.

If $p_a(\ )$ and $p_v(\ )$ are Gaussian with covariance matrices $\mathbf{C}_a$ and $\mathbf{C}_v$, respectively, the a posteriori probability density has the form

$$p_{a|z}(\mathbf{a}|\mathbf{z}) = c(\mathbf{z}) \exp\{-\tfrac{1}{2}[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})]\}$$
$$\cdot \exp\{-\tfrac{1}{2}[\mathbf{a} - \bar{\mathbf{a}}]^T\mathbf{C}_a^{-1}[\mathbf{a} - \bar{\mathbf{a}}]\} \quad (45)$$

where $c(\mathbf{z})$ is a normalization factor that does not depend on $\mathbf{a}$ and $\bar{\mathbf{a}}$ is the prior mean of $\mathbf{a}$. It is useful to consider the form of $p_{a|z}(\ )$ for the special case of a linear $\mathcal{F}(\alpha)$. In this case it can be shown (by "completing the square" in (45)) that $p_{\alpha|z}(\mathbf{a}|\mathbf{z})$ is a multivariate Gaussian probability density function with mean and covariance given by [*Schweppe*, 1973, p. 521]

$$E[\mathbf{a}|\mathbf{z}] = \bar{\mathbf{a}} + \mathbf{C}_{az}\mathbf{C}_z^{-1}[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\bar{\mathbf{a}})] \quad (46)$$

$$E[(\mathbf{a} - \bar{\mathbf{a}})(\mathbf{a} - \bar{\mathbf{a}})^T|\mathbf{z}] = \mathbf{C}_a - \mathbf{C}_{az}\mathbf{C}_z^{-1}\mathbf{C}_{az}^T \quad (47)$$

where $\mathbf{C}_{az}$ is the unconditional cross covariance between $\mathbf{a}$ and $\mathbf{z}$. Since the mode of a Gaussian a posteriori density is equal to its mean, the maximum a posteriori estimate defined by (44) is equal to (46) when $\mathcal{F}(\alpha)$ is linear and $\mathbf{a}$ and $\mathbf{v}$ are jointly Gaussian (see Figure 4a). In fact, in this special case the maximum a posteriori estimate, the conditional mean, and the familiar minimum variance unbiased estimate of Bayesian estimation theory are all the same [*Schweppe*, 1973, p. 329].

When $\mathcal{F}(\alpha)$ is nonlinear the a posteriori density is generally not Gaussian, even when $\mathbf{a}$ and $\mathbf{v}$ are jointly Gaussian, and the a posteriori mean and covariance are no longer given by (46) and (47) [*Schweppe*, 1973; pp. 328–334]. Moreover, the mean and mode of the a posteriori density are no longer equal (see Figure 4b). All of these factors complicate solution of the nonlinear inverse problem. It is worth emphasizing that the maximum a posteriori estimation approach does not require the forward equation to be linear or the a posteriori density to be Gaussian [*Jazwinski*, 1970, pp. 169–174; *Schweppe*, 1973, pp. 391–395; *Bryson and Ho*, 1975, pp. 377–388]. This is in contrast to the impression left by comments of *Gavalas et al.* [1976], *Carrera* [1987], and *Ginn and Cushman* [1990]. The maximum a posteriori approach is actually of most interest in the nonlinear case, when the conditional mean is difficult to derive. The attractive properties of the maximum a posteriori estimate are discussed in more detail by *Bard* [1974, pp. 72–73].

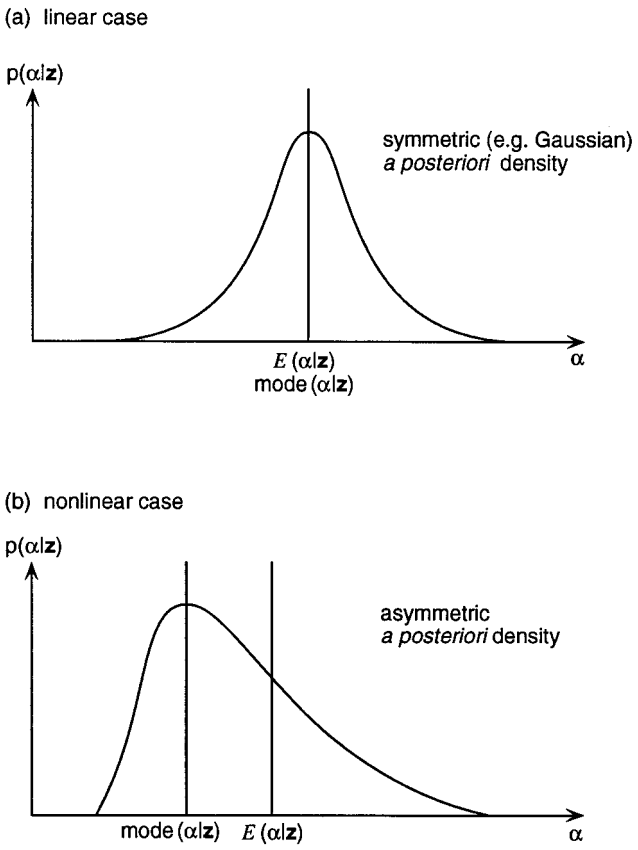

Figure 4. Symmetric and nonsymmetric a posteriori probability density functions.

When $p_{\alpha|z}(\mathbf{a}|\mathbf{z})$ is given by the Gaussian density of (45) it is convenient to find the maximum a posteriori estimate by minimizing $-2 \ln p_{\alpha|z}$, which is a monotonic function of $p_{\alpha|z}$. In this Gaussian case (44) is equivalent to the following weighted least squares criterion [*Schweppe*, 1973, pp. 394–395]:

$\hat{\mathbf{a}}$ is the $\mathbf{a}$ which minimizes

$$J(\mathbf{a}) = [\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})]$$
$$+ [\mathbf{a} - \bar{\mathbf{a}}]^T\mathbf{C}_a^{-1}[\mathbf{a} - \bar{\mathbf{a}}] \quad (48)$$

The solution to this estimation problem is a compromise between the "best fit" estimate (which minimizes only the first term) and the "prior" estimate (which minimizes only the second term). When $\mathcal{F}(\alpha)$ is nonlinear, the maximum a posteriori estimate can be obtained by minimizing $J(\mathbf{a})$ with an iterative search algorithm (see section 3.4 and work by *Gill et al.* [1981]).

The discrete maximum a posteriori estimation criterion given in (48) may be expressed in a more general functional form which accommodates both blocked and geostatistical descriptions of spatial variability. This can be accomplished if $\boldsymbol{\phi}^T\mathbf{a}$ is replaced by α in the first term of (48) and if the identity (A4), derived in Appendix A, is used to write the second term as an integral:

$\hat{\alpha}(\mathbf{x})$ is the $\alpha(\mathbf{x})$ which minimizes

$$\mathcal{J}(\alpha) = [\mathbf{z} - \mathcal{F}(\alpha)]^T \mathbf{C}_v^{-1} [\mathbf{z} - \mathcal{F}(\alpha)] + \int_D \int_D [\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})] C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi}) \cdot [\alpha(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\, d\mathbf{x}\, d\boldsymbol{\xi} \quad (49)$$

where $C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})$ is a scalar "inverse covariance function" which is defined with an identity similar to the one used for the matrix inverse:

$$\int_D \int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi}) C_\alpha(\boldsymbol{\xi}, \boldsymbol{\mu}) \alpha(\boldsymbol{\mu})\, d\boldsymbol{\xi}\, d\boldsymbol{\mu} = \alpha(\mathbf{x}) \quad (50)$$

Equation (50) may be used to derive C_α^{-1} once C_α is specified (an example is provided below). Note that (49) defines a deterministic minimization problem which yields a deterministic solution $\hat{\alpha}(\mathbf{x}) \in A_{ad}$ for any given $\mathbf{z}$ (see the discussion accompanying (17)). The maximum a posteriori approach links this minimization problem to the stochastic description of spatial variability introduced in section 2 [*Wahba*, 1990; *Fitzpatrick*, 1991]. It also provides a rationale for selecting the covariance weighting factors which appear in each term of $\mathcal{J}(\alpha)$.

Although (49) was derived from a finite-dimensional parameterization of $\alpha(\mathbf{x})$ it also applies when $\alpha(\mathbf{x})$ is infinite-dimensional (e.g., when it is a stationary random field). A rigorous demonstration of this result is complicated by the fact that prior and posterior probability densities (countably additive Lebesgue measures) do not exist when α lies in an infinite-dimensional space [*Tarantola*, 1987, chap. 6; *Kuo*, 1975, chap. 1]. Fortunately, it is possible to define an alternative probability measure over the infinite-dimensional parameter space A for the special case of Gaussian random functions. The theoretical basis for defining Gaussian probability measures and related expectations on function spaces is discussed in a number of texts, including those by *Vakhania* [1981] and *Kuo* [1975]. More applied discussions dealing with infinite-dimensional Gaussian extensions of maximum a posteriori estimation are provided by *Kuelbs et al*. [1972], *Larkin* [1972], and *Fitzpatrick* [1991]. *Jazwinski* [1970] presents a heuristic derivation of the infinite-dimensional Gaussian maximum a posteriori estimator which, like more rigorous approaches, yields the estimation criterion given in (49).

Some of the insights provided by a functional formulation of the inverse estimation problem can be appreciated if we consider the structure of (49) when $\alpha(\mathbf{x})$ is a stationary and statistically isotropic exponentially correlated random function of three spatial coordinates. In this case the covariance of α has the following form:

$$C_\alpha(\mathbf{x}, \boldsymbol{\xi}) = \sigma_\alpha^2 \exp\left[-\frac{|\mathbf{x} - \boldsymbol{\xi}|}{\lambda}\right] \quad (51)$$

where σ_α^2 and λ are the variance and correlation scale, respectively. *Tarantola* [1987, pp. 579–584] uses a Fourier transform technique to derive the corresponding inverse covariance function from (50). This inverse covariance is a generalized function (i.e., a weighted sum of derivatives of spatial Dirac delta functions) which may be defined implicitly as

$$\int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi}) \alpha(\boldsymbol{\xi})\, d\boldsymbol{\xi} = \frac{1}{8\pi\sigma_\alpha^2\lambda^3} \cdot [\alpha(\mathbf{x}) - 2\lambda^2\nabla^2\alpha(\mathbf{x}) + \lambda^4\nabla^4\alpha(\mathbf{x})] \quad (52)$$

If we substitute (52) into (49), invoke Green's theorem, and assume that λ is small compared to the dimensions of D (so that we can neglect boundary terms), the resulting expression for $\mathcal{J}(\alpha)$ is [*Tarantola*, 1987, pp. 583–584]

$$\mathcal{J}(\alpha) = [\mathbf{z} - \mathcal{F}(\alpha)]^T \mathbf{C}_v^{-1} [\mathbf{z} - \mathcal{F}(\alpha)] + \frac{1}{8\pi\sigma_\alpha^2\lambda^3} \cdot \int_D [\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})]^2 + 2\lambda^2\nabla[\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})] \cdot \nabla[\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})] + \lambda^4\{\nabla^2[\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})]\}^2\, d\mathbf{x} \quad (53)$$

Since the prior term in the performance index decreases when the spatial derivatives of the estimate are small, the exponential correlation assumption has the effect of imposing a smoothness condition on the inverse problem. The regularization provided by the prior covariance helps to make the inverse problem well-posed by restricting the set of admissible parameter functions [*Tikhonov and Arsenin*, 1977]. A related approach which accomplishes a similar result is described by *Emsellem and Marsily* [1971].

It is instructive to note that an estimate produced by minimizing (53) will always be smoother than a random sample function having the covariance given in (51). In fact, an exponentially correlated sample function would produce an infinitely large regularization term since such functions are not differentiable. We can generalize this result by noting that the infinite-dimensional parameter space A_{ad} of admissible deterministic parameter estimates is always inherently "smoother" than the infinite-dimensional space A of random parameter functions which generate the measurements. More detailed discussions of this point are provided by *Wahba* [1985, 1990] and *Fitzpatrick* [1991].

The example considered above reveals a fundamental connection between stochastic and deterministic approaches to the inverse problem (see work by *Wahba* [1990] and *Bennett* [1992] for detailed discussions). Depending on one's viewpoint, our weighted least squares performance index can be derived probabilistically by introducing Bayesian prior information or it can be derived deterministically by imposing a smoothness condition. In some situations smoothness conditions based on geological arguments may provide better descriptions of log conductivity variability than probabilistic models which are derived from very limited databases. We return to this important topic in section 5.1.

3.4. Solution of the Estimation Problem

In this section we consider how variational methods can be used to solve the functional minimization problem posed in (49). Although there is no general way to find the global minimum of a nonlinear functional, it is possible to find extremal points (local minima, maxima, or saddle points) by imposing the requirement that the first variation of $\mathcal{J}(\alpha)$ about the estimate $\hat{\alpha}(\mathbf{x})$ be 0 [*Courant and Hilbert*, 1953]. This usually leads to a set of nonlinear integral or differential equations which have multiple solutions. The global minimum is defined by the solution which gives the smallest value of $\mathcal{J}(\alpha)$. If the inverse problem is well-posed this global minimum should be unique and relatively insensitive to small fluctuations in the measured data [*Chavent*, 1991].

We begin by deriving an expression for the first variation of $\mathcal{J}(\alpha)$ about a nominal parameter function $\alpha_0(\mathbf{x})$. This expression may be obtained by taking the variation of each term in (49):

$$\delta\mathcal{J}(\alpha_0) = \frac{\partial \mathcal{J}(\alpha_0)}{\partial \alpha}\,\delta\alpha = -2[\mathbf{z} - \mathcal{F}(\alpha_0)]^T\mathbf{C}_v^{-1}\frac{\partial \mathcal{F}(\alpha_0)}{\partial \alpha}\,\delta\alpha + 2\int_D\int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})[\alpha_0(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\delta\alpha(\mathbf{x})\,d\boldsymbol{\xi}\,d\mathbf{x} \tag{54}$$

where $\partial\mathcal{J}(\alpha_0)/\partial\alpha$ and $\partial\mathcal{F}(\alpha_0)/\partial\alpha$ are the functional (or Frechet) derivatives of $\mathcal{J}(\alpha)$ and $\mathcal{F}(\alpha)$, evaluated at $\alpha_0(\mathbf{x})$, and the δ prefix indicates the first variation. These Frechet derivatives are linear operators which map $\delta\alpha(\mathbf{x})$ to the scalar $\delta\mathcal{J}$ and the vector $\delta\mathcal{F}$, respectively. Note that the nominal function $\alpha_0(\mathbf{x})$ and the variation $\delta\alpha(\mathbf{x})$ are constrained to lie in the parameter space A_{ad}. Since the Frechet derivatives are linear functionals they may be expressed as integrals over the kernel functions $g_{\mathcal{J}}$ and $\mathbf{g}_{\mathcal{F}}$ (which remain to be determined):

$$\frac{\partial \mathcal{J}(\alpha_0)}{\partial \alpha}\,\delta\alpha = \int_D g_{\mathcal{J}}(\boldsymbol{\xi}|\alpha_0)\delta\alpha(\boldsymbol{\xi})\,d\boldsymbol{\xi} \tag{55}$$

$$\frac{\partial \mathcal{F}(\alpha_0)}{\partial \alpha}\,\delta\alpha = \int_D \mathbf{g}_{\mathcal{F}}(\boldsymbol{\xi}|\alpha_0)\delta\alpha(\boldsymbol{\xi})\,d\boldsymbol{\xi} \tag{56}$$

By analogy with discrete optimization theory, we call the scalar function $g_{\mathcal{J}}$ the "performance index gradient," and the M-dimensional vector function $\mathbf{g}_{\mathcal{F}}$ the "measurement Jacobian."

Appendix B shows that the log conductivity and head components of the measurement Jacobian for the groundwater problem may be written as (see (B2) and (B6)):

$$g_{\mathcal{F}_{\alpha i}}(\mathbf{x}|\alpha_0) = \delta(\mathbf{x} - \mathbf{x}_i) \tag{57}$$

$$g_{\mathcal{F}_{hi}}(\mathbf{x}|\alpha_0) = -\frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial \alpha}\int_0^T \nabla\zeta_{i0}(\mathbf{x}, \tau)\cdot\nabla h_0(\mathbf{x}, \tau)\,d\tau \tag{58}$$

where $h_0(\mathbf{x}, t)$ is the large-scale head solution derived from $\alpha_0(\mathbf{x})$ and the variable $\zeta_{i0}(\mathbf{x}, t) = G(\mathbf{x}_i, \mathbf{x}, t_i, \tau|\alpha_0)$ is the flow equation Green's function written as a function of only $\mathbf{x}$ and t for measurement i (see (36)). This "adjoint" variable is the solution to the adjoint flow equation given in (B7).

The performance index gradient is obtained by substituting (56) into (54) and applying the definition of (55):

$$g_{\mathcal{J}}(\mathbf{x}|\alpha_0) = -2\mathbf{g}_{\mathcal{F}}^T(\mathbf{x}|\alpha_0)\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\alpha_0)] + 2\int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})]\alpha_0(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\,d\boldsymbol{\xi} \tag{59}$$

It is shown in Appendix B (see B17) that this expression may also be written as

$$g_{\mathcal{J}}(\mathbf{x}|\alpha_0) = 2\frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial \alpha}\int_0^T \nabla\eta_0(\mathbf{x}, t)\cdot\nabla h_0(\mathbf{x}, t)\,dt - 2\sum_{i=M_h+1}^{M_h+M_\alpha}\sum_{j=M_h+1}^{M_h+M_\alpha}[z_{\alpha i} - \alpha_0(\mathbf{x}_i)][\mathbf{C}_{v_\alpha}^{-1}]_{ij}\delta(\mathbf{x} - \mathbf{x}_j) + 2\int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})[\alpha_0(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\,d\boldsymbol{\xi} \tag{60}$$

where η_0 is another adjoint variable (a linear function of the M_h ζ_{i0}'s) which may be obtained by solving (B15). Note that the Jacobian and gradient both depend on the nominal parameter function α_0.

Equations (54) and (55) imply that $\delta\mathcal{J}(\hat{\alpha}) = 0$ only if the gradient function at $\hat{\alpha}(\mathbf{x})$ is 0:

$$g_{\mathcal{J}}(\mathbf{x}|\hat{\alpha}) = -2\mathbf{g}_{\mathcal{F}}^T(\mathbf{x}|\hat{\alpha})\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\hat{\alpha})] + 2\int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})[\hat{\alpha}(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\,d\boldsymbol{\xi} = 0 \tag{61}$$

This integral equation (the Euler-Lagrange equation for our functional minimization problem) is very difficult to solve explicitly if $\mathcal{F}(\alpha)$ depends nonlinearly on α, as it does in the groundwater inverse problem of interest here. It is, however, possible to solve (61) with an iterative algorithm which approximates the nonlinear problem by a sequence of linear problems. On iteration $k + 1$ this algorithm approximates the forward operator by a first-order expansion about $\hat{\alpha}_k$, the estimate obtained from the previous iteration:

$$\mathcal{F}(\alpha) \approx \mathcal{F}(\hat{\alpha}_k) + \frac{\partial \mathcal{F}(\hat{\alpha}_k)}{\partial \alpha}(\alpha - \hat{\alpha}_k) \tag{62}$$

$$k = 0, \cdots, k_{max} - 1$$

where it is understood that the k subscript refers to iteration number. If this linearized expression is substituted for $\mathcal{F}(\hat{\alpha})$ in (61) and the value of $\hat{\alpha}$ in $\mathbf{g}_{\mathcal{F}}(\mathbf{x}|\hat{\alpha})$ is held fixed at $\hat{\alpha}_k$ the extremal solution after iteration $k + 1$ can be written as (see equation (4.108a) of *Tarantola* [1987, p. 244])

$$\hat{\alpha}_{k+1}(\mathbf{x}) = \hat{\alpha}_k(\mathbf{x}) + \int_D [\mathbf{g}_{\mathcal{F}}^T(\mathbf{x}|\hat{\alpha}_k)\mathbf{C}_v^{-1}\mathbf{g}_{\mathcal{F}}(\boldsymbol{\mu}|\hat{\alpha}_k) + C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\mu})]^{-1}\{\mathbf{g}_{\mathcal{F}}^T(\mathbf{x}|\hat{\alpha}_k)\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\hat{\alpha}_k)] - \int_D C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi})[\hat{\alpha}_k(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\,d\boldsymbol{\xi}\}\,d\boldsymbol{\mu} \tag{63}$$

$$k = 0, \cdots, k_{max} - 1$$

The inverse of the square-bracketed function which appears in this equation is defined in a manner analogous to (50).

Equation (63) can be put in an alternative but mathematically equivalent form with the aid of the two operator inversion identities given by *Tarantola* [1987, problem 1.19, p. 158]:

$$\hat{\alpha}_{k+1}(\mathbf{x}) = \bar{\alpha}(\mathbf{x}) + \left[\int_D C_\alpha(\mathbf{x}, \boldsymbol{\mu})\mathbf{g}_{\mathcal{F}}^T(\boldsymbol{\mu}|\hat{\alpha}_k)\,d\boldsymbol{\mu}\right]\cdot\left[\int_D\int_D \mathbf{g}_{\mathcal{F}}(\boldsymbol{\xi}|\hat{\alpha}_k)C_\alpha(\boldsymbol{\xi}, \boldsymbol{\mu})\mathbf{g}_{\mathcal{F}}^T(\boldsymbol{\mu}|\hat{\alpha}_k)\,d\boldsymbol{\xi}\,d\boldsymbol{\mu} + \mathbf{C}_v\right]^{-1}\cdot\left\{[\mathbf{z} - \mathcal{F}(\hat{\alpha}_k)] + \int_D \mathbf{g}_{\mathcal{F}}(\boldsymbol{\xi}|\hat{\alpha}_k)[\hat{\alpha}_k(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\,d\boldsymbol{\xi}\right\} \tag{64}$$

$$k = 0, \cdots, k_{max} - 1$$

This form has the advantage of replacing the functional inversion of (63) by a more computationally convenient matrix in-

version. It is a functional implementation of the widely used Gauss-Newton search algorithm. Although (64) is complex, all of its terms can be readily computed once the flow equation Green's function (or a suitable approximation) is specified.

The primary alternatives to the Gauss-Newton approach for solving the maximum a posteriori estimation problem are gradient-based searches, which include quasi-Newton and conjugate gradient methods [*Gill et al*., 1981; *Tarantola*, 1987]. The functional versions of these iterative algorithms rely on an estimation equation having the following general form:

$$\hat{\alpha}_{k+1}(\mathbf{x}) = \hat{\alpha}_k(\mathbf{x}) + \Delta\hat{\alpha}_k[\mathbf{x}, g_{\mathcal{J}}(\mathbf{x}|\hat{\alpha}_k), \cdots, g_{\mathcal{J}}(\mathbf{x}|\hat{\alpha}_0)] \qquad (65)$$

$$k = 0, \cdots, k_{max} - 1$$

where the update $\Delta\hat{\alpha}_k$ depends on the performance index gradient $g_{\mathcal{J}}(\mathbf{x}|\hat{\alpha}_k)$ evaluated at iteration k (and possibly on gradients from previous iterations) but not on the measurement Jacobian. A simple example is the functional steepest descent algorithm:

$$\Delta\hat{\alpha}_k(\mathbf{x}) = -\mu g_{\mathcal{J}}(\mathbf{x}|\hat{\alpha}_k) \qquad (66)$$

where $g_{\mathcal{J}}(\mathbf{x}|\hat{\alpha}_k)$ is obtained from (60) and μ is a scalar constant that is obtained from a one-dimensional line search [*Tarantola*, 1987, pp. 230–231]. In this case the change in the parameter function is directly proportional to the gradient function.

The similarity between functional estimation algorithms such as (64) and (65) and more familiar discrete estimation algorithms becomes apparent if we replace $\hat{\alpha}_k(\mathbf{x})$ by the finite-dimensional parameterization $\boldsymbol{\phi}^T(\mathbf{x})\hat{\mathbf{a}}_k$ and invoke the identity given in (A4). When this is done, (64) yields the following discrete Gauss-Newton estimation algorithm for the estimate of the basis function coefficient vector $\mathbf{a}$:

$$\hat{\mathbf{a}}_{k+1} = \bar{\mathbf{a}} + \mathbf{C}_a \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\hat{\mathbf{a}}_k)}{\partial \mathbf{a}} \cdot \left[\frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\hat{\mathbf{a}}_k)}{\partial \mathbf{a}} \mathbf{C}_a \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\hat{\mathbf{a}}_k)}{\partial \mathbf{a}} + \mathbf{C}_v\right]^{-1} \cdot \left\{[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\hat{\mathbf{a}}_k)] + \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\hat{\mathbf{a}}_k)}{\partial \mathbf{a}}[\hat{\mathbf{a}}_k - \bar{\mathbf{a}}]\right\} \qquad (67)$$

$$k = 0, \cdots, k_{max} - 1$$

where $\mathbf{C_a}$ is the covariance matrix of $\mathbf{a}$ and $\partial\mathcal{F}(\boldsymbol{\phi}^T\bar{\mathbf{a}})/\partial\mathbf{a}$ is a discrete M by N-dimensional measurement Jacobian matrix which is related to the Jacobian function by the following expression:

$$\frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})}{\partial \mathbf{a}} = \frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})}{\partial \alpha}\frac{\partial \alpha}{\partial \mathbf{a}} = \frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})}{\partial \alpha}\boldsymbol{\phi}^T = \int_D \mathbf{g}_{\mathcal{F}}(\boldsymbol{\xi}|\alpha)\boldsymbol{\phi}^T(\boldsymbol{\xi})\, d\boldsymbol{\xi} \qquad (68)$$

In a similar way, we can derive a discrete gradient-based search algorithm by substituting the basis function expansion for $\hat{\alpha}$ into (65):

$$\hat{\mathbf{a}}_{k+1} = \hat{\mathbf{a}}_k(\mathbf{x}) + \Delta\hat{\mathbf{a}}_k\left[\frac{\partial \mathcal{J}(\boldsymbol{\phi}^T\hat{\mathbf{a}}_k)}{\partial \mathbf{a}}, \cdots, \frac{\partial \mathcal{J}(\boldsymbol{\phi}^T\hat{\mathbf{a}}_0)}{\partial \mathbf{a}}\right] \qquad (69)$$

$$k = 0, \cdots, k_{max} - 1$$

where $\partial\mathcal{J}/\partial\mathbf{a}$ is a discrete one- by N-dimensional performance index gradient vector which is related to the gradient function as follows:

$$\frac{\partial \mathcal{J}(\boldsymbol{\phi}^T\mathbf{a})}{\partial \mathbf{a}} = \frac{\partial \mathcal{J}(\boldsymbol{\phi}^T\mathbf{a})}{\partial \alpha}\frac{\partial \alpha}{\partial \mathbf{a}} = \frac{\partial \mathcal{J}(\boldsymbol{\phi}^T\mathbf{a})}{\partial \alpha}\boldsymbol{\phi}^T = \int_D g_{\mathcal{J}}(\boldsymbol{\xi}|\alpha)\boldsymbol{\phi}^T(\boldsymbol{\xi})\, d\boldsymbol{\xi} \qquad (70)$$

The function $\Delta\hat{\mathbf{a}}_k(\ \ )$ can be identified directly from a discretized version of (65). Note that functional estimates can be obtained from either of the discrete $\hat{\mathbf{a}}$ estimation algorithms by applying the relationship $\hat{\alpha}_k(\mathbf{x}) = \boldsymbol{\phi}^T(\mathbf{x})\hat{\mathbf{a}}_k$.

Appendix B discusses two numerical methods for evaluating the discrete Jacobian and performance index gradient: the adjoint sensitivity approach and the finite difference approach. If an adjoint approach is used, $M_h + 1$ flow equation solutions are required to compute the discrete Jacobian on each iteration, while only two solutions are required to compute the discrete performance index gradient. If a finite difference approach is used, $N + 1$ flow equation solutions are required to evaluate either the Jacobian or the performance function gradient. At first glance an adjoint-based gradient search would also seem to be much more efficient than the Gauss-Newton alternative. Unfortunately, gradient-based search algorithms tend to converge less rapidly and less reliably than Gauss-Newton algorithms, and they are highly dependent on the efficiency and accuracy of the line search needed to derive the step length along the search direction. So, although each iterative update in a gradient search may be less expensive to compute, the overall effort required to obtain an acceptable solution may be greater. The choice between the gradient and Gauss-Newton approaches remains a topic of some controversy since different applications appear to favor different methods. A recent performance comparison provided by *Zou et al*. [1993] shows that the Gauss-Newton approach is probably the best choice if either N or M_h is sufficiently small to make the method computationally feasible. Otherwise, an adjoint-based gradient search algorithm is the best option.

The finite-dimensional discrete forms of the Gauss-Newton and gradient-based maximum a posteriori estimation equations are more convenient than the functional forms for practical numerical computation. The functional forms are, however, more general since they apply for both finite and infinite-dimensional expansions of $\alpha(\mathbf{x})$. Moreover, the discrete algorithms can always be derived from their functional counterparts by substituting a basis function expansion into the appropriate functional update equation. We believe that functional estimation equations play a role in inverse theory similar to the role played by partial differential equations in mathematical modeling. In both cases, it is useful to pose the problem in a continuum form before discretizing. This clearly distinguishes fundamental assumptions about the nature of spatial variability from more pragmatic discretization assumptions introduced for computational reasons.

4. Representative Groundwater Inverse Algorithms

In this section we review a number of inverse algorithms which have been developed from different perspectives but which fit into the maximum a posteriori estimation framework.

The parameterizations adopted in these algorithms can all be expressed probabilistically, either in terms of functional statistics or in terms of basis function expansions and associated discrete statistics. The forward equations all have the form given in (42). The estimation criteria can all be put in the functional form given in (49). Finally, the solution techniques can all be viewed as special cases of the iterative search algorithms described in section 3.4. In the subsections that follow we distinguish linear methods (which provide a solution after one iteration) from nonlinear methods (which search for a solution over many iterations). Our primary objective is to reveal the common origins and assumptions of different methods. We believe that this can help us to develop new inverse techniques as well as to make better use of the ones we have.

4.1. Linear Methods

Bayesian parameter estimates generally depend nonlinearly on measurements when the forward equation is nonlinear in $\alpha(\mathbf{x})$, as in the groundwater flow application of interest here. For example, the Gauss-Newton estimate obtained after the first iteration of (64) is nonlinear because the terms which multiply the measurement residual depend on the measurement vector $\mathbf{z}$. We can obtain an estimator which is linear in $\mathbf{z}$ if we introduce certain approximations in the maximum a posteriori problem formulation. Linear estimators have the merit of being simple to use and reasonably robust, and there is an extensive literature on their properties. For these reasons they deserve serious consideration in applications where their assumptions can be justified.

There are many different ways to derive a linear Bayesian estimator, all giving essentially the same result. One alternative is to replace the forward operator in the original performance index by the following linear approximation:

$$\mathcal{F}(\alpha) \approx \mathcal{F}(\bar{\alpha}) + \frac{\partial \mathcal{F}(\bar{\alpha})}{\partial \alpha}(\alpha - \bar{\alpha}) \tag{71}$$

If the derivation of section 3 is repeated with this substitution the resulting estimate is identical to the one obtained after one iteration of the Gauss-Newton algorithm of (64), with $\hat{\alpha}_0(\mathbf{x})$ set equal to $\bar{\alpha}(\mathbf{x})$ [*Carrera and Glorioso*, 1991]. When the forward operator is truly linear in α the Gauss-Newton search converges to the exact solution in one iteration. When the forward operator is nonlinear, the first iteration solution is an approximation which can be refined by taking additional iterations. In either case, if we stop after the first iteration the resulting linear estimate may be written as

$$\hat{\alpha}(\mathbf{x}) = \bar{\alpha}(\mathbf{x}) + \left[\int_D C_\alpha(\mathbf{x}, \boldsymbol{\mu}) \mathbf{g}_{\mathcal{F}}^T(\boldsymbol{\mu}|\bar{\alpha})\, d\boldsymbol{\mu}\right] \cdot \left[\int_D \int_D \mathbf{g}_{\mathcal{F}}(\boldsymbol{\xi}|\bar{\alpha}) C_\alpha(\boldsymbol{\xi}, \boldsymbol{\mu}) \mathbf{g}_{\mathcal{F}}^T(\boldsymbol{\mu}|\bar{\alpha})\, d\boldsymbol{\xi}\, d\boldsymbol{\mu} + \mathbf{C}_v\right]^{-1} \cdot [\mathbf{z} - \mathcal{F}(\bar{\alpha})] \tag{72}$$

The discrete version of (72) follows from (67):

$$\hat{\mathbf{a}} = \bar{\mathbf{a}} + \mathbf{C}_a \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} \left[\frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} \mathbf{C}_a \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} + \mathbf{C}_v\right]^{-1} \cdot [\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\bar{\mathbf{a}})] \tag{73}$$

Note that both the functional and discrete versions of this first iteration estimate are linear in $\mathbf{z}$ since the coefficients of the measurement residual term do not depend on $\mathbf{z}$.

The linear estimation equations of (72) and (73) may be expressed more concisely if we adopt the following approximation for $\mathbf{z}$:

$$\mathbf{z} = \bar{\mathbf{z}} + \mathbf{z}' \approx \bar{\mathbf{z}} + \frac{\partial \mathcal{F}(\bar{\alpha})}{\partial \alpha}\alpha' + \mathbf{v} = \bar{\mathbf{z}} + \frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}}\mathbf{a}' + \mathbf{v} \tag{74}$$

where $\alpha' = \alpha - \bar{\alpha}$, $\mathbf{a}' = \mathbf{a} - \bar{\mathbf{a}}$, and $\mathbf{z}' = \mathbf{z} - \bar{\mathbf{z}}$ are the fluctuations of α, $\mathbf{a}$, and $\mathbf{z}$ about their respective means. The expressions to the right of the first and second equalities are the functional and discrete versions of the approximation, respectively. The validity of (74) depends both on the structure of $\mathcal{F}(\alpha)$ and on the magnitude of α', as measured by the variance of α. The approximation breaks down for large-scale log conductivity variances much greater than 1.

Equation (74) enables us to write the measurement covariance matrix $\mathbf{C}_z$ and the parameter measurement covariance vector $\mathbf{C}_{\alpha z}(\mathbf{x})$ as follows:

$$\mathbf{C}_z = \overline{\mathbf{z}'\mathbf{z}'^T} = \int_D \int_D \mathbf{g}_{\mathcal{F}}(\boldsymbol{\xi}|\bar{\alpha}) C_\alpha(\boldsymbol{\xi}, \boldsymbol{\mu}) \mathbf{g}_{\mathcal{F}}^T(\boldsymbol{\mu}|\bar{\alpha})\, d\boldsymbol{\xi}\, d\boldsymbol{\mu} + \mathbf{C}_v \tag{75}$$

$$\mathbf{C}_{\alpha z}(\mathbf{x}) = \overline{\alpha'(\mathbf{x})\mathbf{z}'^T} = \int_D C_\alpha(\mathbf{x}, \boldsymbol{\mu}) \mathbf{g}_{\mathcal{F}}^T(\boldsymbol{\mu}|\bar{\alpha})\, d\boldsymbol{\mu} \tag{76}$$

Or, if a finite-dimensional parameterization is introduced, we can use an equivalent discrete representation:

$$\mathbf{C}_z = \overline{\mathbf{z}'\mathbf{z}'^T} = \frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} \mathbf{C}_a \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} + \mathbf{C}_v \tag{77}$$

$$\mathbf{C}_{\alpha z}(\mathbf{x}) = \overline{\alpha'(\mathbf{x})\mathbf{z}'^T} = \boldsymbol{\phi}(\mathbf{x}) \mathbf{C}_a \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} \tag{78}$$

When the appropriate covariance expressions are substituted into either (72) or (73) the resulting expression for the linear estimate of $\alpha(\mathbf{x})$ is

$$\hat{\alpha}(\mathbf{x}) = \bar{\alpha}(\mathbf{x}) + \mathbf{C}_{\alpha z}(\mathbf{x}) \mathbf{C}_z^{-1} [\mathbf{z} - \mathcal{F}(\bar{\alpha})] \tag{79}$$

where it is understood that $\bar{\alpha}(\mathbf{x}) = \boldsymbol{\phi}^T(\mathbf{x})\bar{\mathbf{a}}$ in the discrete case. Equation (79) is equivalent to the expression presented in (46) for the Gaussian conditional mean. This is as expected since the maximum a posteriori estimate (or conditional mode) is equal to the conditional mean in the linear Gaussian case (recall Figure 4). Equation (79) and closely related variants form the basis for all of the linear inverse methods described in the groundwater literature.

The functional covariance expressions given in (75) and (76) are most useful when the large-scale log conductivity is treated as a random field with a specified mean $\bar{\alpha}(\mathbf{x})$ and covariance function $C_\alpha(\mathbf{x}, \boldsymbol{\xi})$. This approach is typically associated with geostatistical descriptions of spatial variability. If we adopt a functional approach we can derive a closed form expression for the head portion of the functional Jacobian if we assume that (1) steady state conditions apply (i.e., $\partial h/\partial t = 0$), (2) the effective conductivity is given by the geometric mean of the

large-scale conductivity so that $[\mathbf{K}_{eff}(\alpha_0)]_{ij} = e^{\alpha_0}\delta_{ij}$, (3) $\bar{\alpha}$ is a constant (i.e., $\nabla\bar{\alpha} = 0$), and (4) the effects of boundary conditions are accounted for by the mean hydraulic gradient, which is assumed to be known. Appendix B shows that $g_{\mathcal{F}_{hi}}(\mathbf{x}|\bar{\alpha})$ has the following form in this special case:

$$g_{\mathcal{F}_{hi}}(\mathbf{x}|\bar{\alpha}) = -\nabla G_{\nabla^2}(\mathbf{x}, \mathbf{x}_i) \cdot \nabla\bar{h}(\mathbf{x}) \tag{80}$$

where G_{∇^2} is the infinite domain Green's function associated with the Laplacian (∇^2) operator. When (80) is substituted into (75) and (76) the estimate obtained from (79) is the same as the Gaussian conditional mean estimator given in (2e–2h) and (3h–3i) of *Dagan* [1985, pp. 66–67]. This simplified maximum a posteriori estimator is also the same as the cokriging estimator proposed by *Gutjahr and Wilson* [1985]. The linear estimation approach has been extended by *Rubin and Dagan* [1987a, b], who use a maximum likelihood technique to estimate a constant random recharge value as well as the mean log conductivity, the components of the mean head gradient, and several statistical parameters associated with the prior covariance $C_\alpha(\mathbf{x}, \boldsymbol{\xi})$. Note that a closed form linear estimator based on (80) and (79) provides a true functional solution to the groundwater inverse problem since it yields a distinct estimate at every point in the computational domain.

Although Dagan's linear inverse algorithm does not require discretization of the log conductivity function, it is instructive to note that the estimate it produces can be obtained by adopting the following M-dimensional basis function expansion for $\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x})$:

$$\alpha(\mathbf{x}) - \bar{\alpha}(\mathbf{x}) = \sum_{i=1}^{M} a_i\phi_i(\mathbf{x}) = \sum_{i=1}^{M} a_i C_{\alpha z_i}(\mathbf{x}) = \mathbf{C}_{\alpha z}^T(\mathbf{x})\mathbf{a} \tag{81}$$

where $\mathbf{a}$ is a zero mean M-dimensional vector of unknown basis function coefficients and the basis function $\phi_i(\mathbf{x})$ is the covariance $C_{\alpha z_i}(\mathbf{x})$ between $\alpha(\mathbf{x})$ and z_i. If this expansion and the linear approximation of (71) are substituted into (49) and the performance index is minimized with respect to $\mathbf{a}$, the result is

$$\hat{\mathbf{a}} = \mathbf{C}_z^{-1}[\mathbf{z} - \mathcal{F}(\bar{\alpha})] \tag{82}$$

When this expression is substituted into (81) the estimate obtained for $\alpha(\mathbf{x})$ is identical to the one given in (79). This suggests that the effective number of unknowns in the linear estimation algorithm of Dagan is equal to M, the number of measurements. However, we know from the discussion following (32) that the Karhunen-Loeve expansion of a stationary random field generally has an infinite number of basis function coefficients. This difference can be explained by noting that the estimate provided in (79) is actually a projection of the infinite-dimensional function $\alpha(\mathbf{x})$ onto the M-dimensional space spanned by the $C_{\alpha z_i}(\mathbf{x})$ functions [*Wahba*, 1990]. As mentioned in section 3.3, this finite-dimensional projection will always be smoother than the actual infinite-dimensional parameter function.

If the simplifying assumptions used to derive (80) are not appropriate we can derive $\mathbf{C}_z$ and $\mathbf{C}_{\alpha z}$ from the discrete expressions given in (77) and (78). *Hoeksema and Kitanidis* [1984] take this approach in their linear geostatistically oriented steady state inverse algorithm. For the purpose of deriving covariances they assume that the log conductivity (or transmissivity) is an intrinsic random field which can be adequately approximated by a discrete block-oriented parameterization:

$$\alpha(\mathbf{x}) = \sum_{i=1}^{N_b} a_{bi}\phi_{bi}(\mathbf{x}) \tag{83}$$

where a_{bi} is the block-averaged log conductivity in block i and the basis functions are defined as follows:

$$\begin{aligned} \phi_{bi}(\mathbf{x}) &= 1 \qquad \text{if } \mathbf{x} \in \text{parameter block } i \\ \phi_{bi}(\mathbf{x}) &= 0 \qquad \text{otherwise} \end{aligned} \tag{84}$$

where $i = 1, \cdots, N_b$ and N_b is the number of grid blocks included in the computational domain. The Hoeksema and Kitanidis algorithm does not need to evaluate all of the discrete block coefficients (the a_{bi}'s) since the required covariances can be derived from the basis functions, the discrete Jacobian, and $\mathbf{C}_a$ (which is obtained by integrating $C_\alpha(\mathbf{x}, \boldsymbol{\xi})$ over each grid block). Once $\mathbf{C}_z$ and $\mathbf{C}_{\alpha z}$ are evaluated, estimates of $\alpha(\mathbf{x})$ can be computed from (79) at as many or as few points as desired. *Hoeksema and Kitanidis* [1984] derived the discrete Jacobian by inverting the coefficient matrix of a discretized version of the linearized steady state flow equation (see (B8)). This operation requires N_b solutions of the discrete flow equation (the same amount of effort as a direct finite difference derivation of the Jacobian). When the domain is large the computational effort required by the algorithm can be reduced significantly if the discrete Jacobian is derived with an adjoint procedure (see Appendix B).

The original Hoeksema and Kitanidis estimation algorithm is expressed in terms of generalized covariances rather than the conventional means and covariances used here (see work by *Marsily* [1986, pp. 312–318] for a discussion of generalized covariance functions). The generalized covariance formulation constrains the estimator to insure that polynomial trends of a given order are always removed from the data. This is useful if the prior mean of the large-scale log conductivity can be represented as a polynomial with unknown coefficients. If polynomial filtering constraints are added to the maximum a posteriori performance index (e.g., by using a Lagrange multiplier approach) it is possible to make (79) look exactly like the *Hoeksema and Kitanidis* [1984] estimation equation. *Carrera and Glorioso* [1991] have demonstrated this for the steady state case. The complete Hoeksema and Kitanidis algorithm also includes a maximum likelihood procedure for estimating the functional form of the generalized covariance. This is an important capability which increases the practical usefulness of the algorithm when there are enough data to justify covariance estimation (see section 5.1).

The Hoeksema and Kitanidis approach to the steady state inverse problem has been extended to the dynamic case by *Sun and Yeh* [1992]. The Sun and Yeh algorithm is also based on (79), with the measurement Jacobian derived from a discrete adjoint technique. *Carrera and Medina* [1994] describe a similar application of the adjoint approach which includes a way to significantly increase computational efficiency when multiple head measurements are taken over time at each sampling location. Nevertheless, $\mathbf{C}_{\alpha z}$ and $\mathbf{C}_z$ can become very large in dynamic problems, particularly if a high-resolution parameterization is used for the large-scale log conductivity. The conceptual and computational issues which affect parameterization decisions are discussed further in section 5.3.

4.2. Nonlinear Methods

Nonlinear inverse methods have been used by hydrologists, chemical engineers, and petroleum engineers since the mid-

1960s. These methods have been popular in part because they mimic the iterative process carried out in manual "model calibration" or "history-matching" exercises. The basic philosophy is to progressively refine an initial or "prior" estimate until the fit between measurements and predictions can no longer be improved. Although nonlinear inversion techniques do not involve as many assumptions as linear methods, they are less likely to provide unique solutions and they can be more difficult to apply in practice. All of the nonlinear inverse algorithms reviewed here minimize some version of the maximum a posteriori performance index given in (49) (functional form) or (48) (discrete form). Since the algorithms we consider all share the same performance index, they differ primarily in the search techniques they use for finding a minimum and in the basis function expansions they use to parameterize $\alpha(\mathbf{x})$.

While there are many ways to find the minimum of a nonlinear performance index, the alternatives most commonly used in groundwater applications are the Gauss-Newton and gradient-based methods discussed in section 3.3. The choice of search technique can have a significant impact on the results obtained in any particular application. Nevertheless, it is difficult to generalize about the performance of different alternatives. Most of the nonlinear inverse methods reviewed here have been tested with different search algorithms and some are even designed to change algorithms as the search proceeds. Generally speaking, standard nonlinear search techniques will converge when the underlying problem is well-posed and will have difficulties when it is not (this statement could, in fact, be viewed as an informal operational definition of well-posedness).

We begin our discussion by considering methods that rely explicitly on Bayesian estimation concepts and then consider, in turn, least squares, maximum likelihood, and pilot point methods. We show that the latter three methods are equivalent to the maximum a posteriori approach when prior information is included. We conclude with a discussion of the extended Kalman filter, which minimizes (48) with a suboptimal recursive algorithm.

4.2.1. Maximum a posteriori methods. A number of investigators have proposed inverse algorithms which are direct applications of the Gaussian maximum a posteriori approach. The basic concepts are described by *Gavalas et al*. [1976] and *Shah et al*. [1978], who use the discrete performance index of (48), and *Reid and McLaughlin* [1994], who use the functional performance index of (49). The parameterizations adopted in all of these studies can be readily written as basis function expansions having the form given in (29). It is instructive to consider the parameterization alternatives in detail since they illustrate some of the computational issues which arise when the maximum a posteriori approach is implemented.

The original version of the *Gavalas et al*. [1976] algorithm approximated the random function $\alpha(\mathbf{x})$ with a blocked parameterization similar to the one used by *Hoeksema and Kitanidis* [1984] (see (83) and (84)). As in the Hoeksema and Kitanidis application, the blocks were chosen to be small "pixels," or grid blocks, rather than regional geological features. This approach provides a high-resolution characterization of parameter variability over the computational domain (see Figure 3). However, there is a price to be paid for such high resolution, especially in the nonlinear case. The nonlinear Gavalas et al. algorithm estimates all of the block parameter values simultaneously (i.e., it must solve a system of N_b nonlinear equations in N_b unknowns, where N_b is the number of blocks in the computational domain). The computational cost of this high-dimensional parameterization becomes prohibitive in three-dimensional applications, where N_b can easily exceed 10^5 [*Ababou et al*., 1989]. Furthermore, such a parameterization is inefficient since most of the estimated block values (those more than few correlation distances away from measurements) are very close to the prior mean.

Gavalas et al. [1976] dealt with the computational disadvantages of a high-resolution blocked parameterization by introducing an alternative parameterization based on an eigenvalue decomposition. This approach expresses the unknown block parameters in (83) as expansions over the eigenvectors of the prior covariance matrix $\mathbf{C}_a$ of the blocked parameter vector $\mathbf{a}_b$:

$$a_{bi} = \sum_{j=1}^{N_e} a_{ej}\Lambda_{ij} \qquad i = 1, \cdots, N_b \tag{85}$$

where a_{bi} is the value of the log conductivity in block i, Λ_{ij} is component i of the jth eigenvector of $\mathbf{C}_a$, a_{ej} is the unknown coefficient of eigenvector j, and N_e is the number of eigenvectors retained in the expansion. Generally speaking, it is possible to capture most of the variability in $\alpha(\mathbf{x})$ by keeping only a small fraction of the N_b eigenvectors of $\mathbf{C}_a$. The eigenvector expansion can be put in the functional form used in our analysis if (83) and (85) are combined:

$$\alpha(\mathbf{x}) = \sum_{j=1}^{N_e} a_{ej}\left[\sum_{i=1}^{N_b} \Lambda_{ij}\phi_{bi}(\mathbf{x})\right] = \sum_{i=1}^{N_e} a_{ej}\phi_{ej}(\mathbf{x}) \tag{86}$$

where $\phi_{bi}(\mathbf{x})$ is defined in (84) and the basis function $\phi_{ej}(\mathbf{x})$ is the known expression in brackets. The eigenvector approach provides a more efficient description of geological variability than the blocked alternative, and it generally yields a better-posed problem since the number of unknowns is much smaller.

Gavalas et al. [1976] solved the maximum a posteriori problem with discrete Gauss-Newton and conjugate gradient algorithms which rely on derivatives derived from an adjoint technique. They provide a detailed comparison of the computational effort and accuracy achieved with their block and eigenvector-based parameterizations. Generally speaking, their results favor an eigenvector parameterization implemented with a Gauss-Newton search. It should be mentioned that a possible alternative to the *Gavalas et al*. [1976] approach is to derive the basis functions $\phi_{ej}(\mathbf{x})$ directly from the $\alpha(\mathbf{x})$ covariance function. This could be accomplished by solving the Karhunen-Loeve eigenvalue equation with a Fourier transform technique (see the discussion accompanying (32)). Although the *Gavalas et al*. [1976] conclusions are based on a simple one-dimensional example, they merit review by anyone interested in applying maximum a posteriori estimation techniques to the groundwater inverse problem.

Reid and McLaughlin [1994] introduced an alternative low-dimensionality parameterization in their functional version of the maximum a posteriori estimation algorithm. This parameterization is suggested by the M-dimensional basis function expansion that can be used to derive the linear *Dagan* [1985] algorithm (see (81)):

$$\hat{\alpha}(\mathbf{x}) - \bar{\alpha}(\mathbf{x}) = \sum_{i=1}^{M} \hat{a}_i\phi_i(\mathbf{x}) = \sum_{i=1}^{M} \hat{a}_i C_{\alpha z_i}(\mathbf{x}) \tag{87}$$

Reid and McLaughlin [1994] substitute this expansion into the functional Gauss-Newton algorithm of (64). This yields the

discrete iterative estimation algorithm of (67) for the M-dimensional coefficient vector $\mathbf{a}$. The Reid and McLaughlin basis/covariance functions (the $C_{\alpha z_i}(\mathbf{x})$'s) are derived from the same functional equation as *Dagan*'s [1985] covariances. The first iteration of their nonlinear search for the optimum gives an estimate $\hat{\mathbf{a}}_1$ identical to the closed form solution obtained by *Dagan* [1985]. On subsequent iterations the discrete M-by-M Jacobian matrix is recomputed with a finite difference approximation and the coefficient estimate $\hat{\mathbf{a}}_k$ is modified, although the basis function vector $C_{\alpha z}(\mathbf{x})$ is not changed. This nonlinear maximum a posteriori approach is very efficient since it estimates only M unknown coefficients (the same as the *Dagan* [1985] linear approach). In fact, the number of coefficients can be reduced still further if an eigenvalue decomposition similar to the one proposed by *Gavalas et al*. [1976] is included in the estimation algorithm. A similar parameterization has been proposed by *Bennett* [1992] for oceanographic applications.

Maximum a posteriori and related regularization techniques have not been applied as extensively in hydrology as in other fields, although the Bayesian philosophy that forms the basis for the maximum a posteriori approach has had a significant impact through the use of geostatistics. In the next subsection we see how Bayesian ideas have influenced nonlinear inverse techniques that rely on non-Bayesian estimation concepts.

4.2.2. Nonlinear least squares and maximum likelihood methods. Nonlinear least squares methods and related regression techniques are widely used to fit experimental models to data [*Bard*, 1974] and have been incorporated into practical groundwater inverse algorithms [*Hill*, 1991]. They differ from the Bayesian methods discussed so far in this paper in that they treat the unknown parameters as deterministic quantities without any particular statistical (or spatial) structure. Least squares estimates minimize the following version of (48):

$$J(\mathbf{a}) = [\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})] \qquad (88)$$

Note that the least squares performance index includes only the first (goodness-of-fit) term in (48). The prior term is omitted, at least in the standard version of nonlinear least squares. The performance index is not derived from probabilistic arguments (as in maximum a posteriori theory) but is simply presented as a reasonable basis for estimation [*Jazwinski*, 1970; pp. 150–159]. The measurement vector $\mathbf{z}$ is assumed to be generated by a forward equation such as (42), with an additive random measurement error $\mathbf{v}$. The measurement error is often assumed to be Gaussian in order to facilitate statistical analysis of results [*Bard*, 1974]. Numerical solutions of the least squares problem are generally found with iterative search algorithms such as the Gauss-Newton or gradient methods discussed earlier.

Nonlinear least squares algorithms were first applied to distributed parameter groundwater problems when numerical models became widely available in the 1960s and 1970s [*Jacquard and Jain*, 1965; *Jahns*, 1966; *McLaughlin*, 1975; *Cooley*, 1977, 1979]. These early algorithms generally provided only for head measurements. *Neuman and Yakowitz* [1979], *Neuman et al*. [1980], and *Neuman* [1980] added log conductivity measurements in the late 1970s. The role of head and log conductivity measurements can be clearly distinguished if the corresponding measurement errors are assumed to be uncorrelated and (88) is expanded as follows:

$$J(\mathbf{a}) = [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{hv}^{-1}[\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})] + [\mathbf{z}_\alpha - \mathcal{F}_\alpha(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{\alpha v}^{-1}[\mathbf{z}_\alpha - \mathcal{F}_\alpha(\boldsymbol{\phi}^T\mathbf{a})] \qquad (89)$$

where the h and α subscripts refer to the head and log conductivity components of the indicated variables.

Neuman and Yakowitz [1979] observed that the least squares estimator of (88) is equivalent to a Gaussian maximum likelihood estimator with known statistical properties. This connection is made more explicit in the subsequent work of *Carrera and Neuman* [1986a, b, c] and *Samper and Neuman* [1989]. The discrete maximum likelihood estimate is the mode of the conditional density $p_{z|\alpha}(\mathbf{z}|\mathbf{a})$, which is commonly called the likelihood function when treated as a function of $\mathbf{a}$ for fixed $\mathbf{z}$. It follows from (42) that the likelihood function is

$$p_{z|\alpha}(\mathbf{z}|\mathbf{a}) = p_v[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})] \qquad (90)$$

where $p_v(\ \ )$ is the probability density of the measurement error $\mathbf{v}$. It is often convenient to derive the maximum likelihood estimate by minimizing $-2 \ln p_{z|\alpha}(\mathbf{z}|\mathbf{a})$, which is a monotonic function of the likelihood. If $p_v(\ \ )$ is Gaussian, this transformed likelihood function is identical to the least squares performance index of (88) and the maximum likelihood estimate is, consequently, the same as the nonlinear least squares estimate.

Since maximum likelihood estimation, like least squares estimation, assumes that $\mathbf{a}$ is an unknown deterministic parameter, the only source of randomness in the formulation is measurement error [*Schweppe*, 1973, pp. 100–104]. This sometimes creates confusion about the implications of certain statistical assumptions made in the maximum likelihood approach. For example, the assumption that the head and log conductivity measurement errors are uncorrelated does not imply that the actual head and log conductivity are uncorrelated. In fact, the concept of correlation does not apply to these variables since they are not random quantities in the maximum likelihood formalism. Only the measurement errors are random. In practice, errors in the measurement process (e.g., errors in reading the depth-to-water in a well) are not large enough to explain observed deviations between measured and predicted heads. In order to be useful, the concept of measurement error must be extended to include such things as scale discrepancies (e.g., differences between model discretization and well scales) or modeling errors (e.g., improperly specified recharge rates) (see the discussion accompanying (24)). This generalization of the measurement error concept is particularly important in least squares and maximum likelihood methods since there is no other way to account for uncertainty.

We can gain some insight about least squares and maximum likelihood solutions to the groundwater inverse problem by noting that they are equivalent to the discrete Gaussian maximum a posteriori solution obtained when the prior covariance $\mathbf{C}_\alpha$ is arbitrarily large [*Schweppe*, 1973, pp. 103, 339]. From the perspective of maximum a posteriori theory an arbitrarily large covariance implies that the prior density for $\mathbf{a}$ is uninformative (i.e., it implies that we know nothing about $\mathbf{a}$). This is unrealistically pessimistic since we can usually put upper and lower bounds on the range of "reasonable" values for parameters such as the large-scale log hydraulic conductivity. In fact, most practical nonlinear regression and maximum likelihood algorithms provide for such constraints in their search algorithms.

It is easy to show that a bounded least squares algorithm is equivalent to a maximum a posteriori estimator which assumes that the measurement error probability density is Gaussian and the prior probability density is uniform between the specified bounds. In this case the maximum a posteriori performance index becomes

$$J(\mathbf{a}) = [\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\boldsymbol{\alpha})]^T\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\boldsymbol{\phi}^T\boldsymbol{\alpha})] - \ln p_\alpha(\mathbf{a}) \quad (91)$$

where the prior density $p_\alpha(\mathbf{a})$ is a positive constant when $\mathbf{a}$ lies within the bounds, and 0 when it lies outside. Since this performance index imposes an infinite penalty on estimates lying outside the permitted interval, it will enforce the constraints perfectly.

In practice, strict bounds may not be the best way to introduce prior information into a least squares or maximum likelihood inverse algorithm. Advocates of the least squares approach have recognized the value of prior geologic information in groundwater applications and have developed a number of ways to accommodate it in their inverse procedures. Examples may be found in work by *Neuman and Yakowitz* [1979], *Clifton and Neuman* [1982], *Cooley* [1982, 1983], and *Carrera and Neuman* [1986a, b, c]. When these investigators use the term "prior information" they are generally referring to either direct or surrogate measurements of log conductivity rather than to Bayesian prior statistics associated with a random parameter. Nevertheless, the maximum likelihood and Bayesian interpretations of prior information often give the same results.

This is most clearly demonstrated in the work of *Clifton and Neuman* [1982], who use a block kriging algorithm to estimate log conductivity values in the cells of a model computational grid. The kriged log conductivity estimates are derived from scattered log conductivity measurements and a set of prior statistics (i.e., a variogram or a mean and covariance function). These estimates are incorporated into the least squares/maximum likelihood performance index as if they were actual measurements of log conductivity. It is shown in Appendix C that this two-step operation is equivalent to a discrete maximum a posteriori estimator based on the same prior statistics as the kriging algorithm. Once Bayesian concepts are introduced via kriging (which is itself a maximum a posteriori estimation technique), the nonlinear maximum likelihood estimator acts just like a Bayesian estimator, displaying similar regularization properties and a similar tendency to produce estimates close to the prior mean.

4.2.3. The pilot point method. The pilot point method incorporates prior information in a different way than the methods described above, although it retains many of the traits of maximum a posteriori estimation. The basic idea is to approximate the effective log conductivity by a smooth function which reproduces available log conductivity measurements while giving an acceptable fit to head data. Details are discussed in work by *Marsily* [1984], *Marsily et al.* [1984], and *Certes and Marsily* [1991]. These papers adopt a variogram-based description of spatial variability. In order to maintain consistency with the notation used in preceding sections, we use a covariance-based (but nonstationary) description. Otherwise, the analysis given here is consistent with published explanations of the method.

The log conductivity estimate used in the pilot point method is obtained from a kriging algorithm which interpolates point measurements. Two distinct types of "measurements" are included in this procedure: (1) actual log conductivity measurements $z_{\alpha i}$ inferred from pump tests, soil samples, etc., collected at the locations $\mathbf{x}_i$, where $i = M_h + 1, \cdots, M_h + M_\alpha$, and (2) estimated log conductivity values a_{pi} at a carefully selected set of "pilot points" located at $\mathbf{x}_{pi}$, where $i = 1, \cdots, N$. Recall that M_α and M_h are, respectively, the number of log conductivity and head measurements used for estimation. If we assume for the moment that the pilot point values are known, the kriging (or minimum variance linear unbiased) estimator of $\alpha(\mathbf{x})$ is given by (79), written here as a scalar sum in order to more clearly distinguish the role of the pilot points:

$$\hat{\alpha}(\mathbf{x}) = \bar{\alpha}(\mathbf{x}) + \sum_{i=M_h+1}^{M_h+M_\alpha}\sum_{j=M_h+1}^{M_h+M_\alpha}[\mathbf{C}_{\alpha z_\alpha}(\mathbf{x})]_j[\mathbf{C}_{z_\alpha}^{-1}]_{ji}[z_{\alpha i} - \bar{\alpha}(\mathbf{x}_i)] + \sum_{i=1}^{N}\sum_{j=1}^{N}[\mathbf{C}_{\alpha a_p}(\mathbf{x})]_j[\mathbf{C}_{a_p}^{-1}]_{ji}[a_{pi} - \bar{a}_{pi}] \quad (92)$$

where $\mathbf{C}_{\alpha z_\alpha}(\mathbf{x})$ is the covariance between $\alpha(\mathbf{x})$ and the M_α-dimensional vector $\mathbf{z}_\alpha$ of log conductivity measurements, $\mathbf{C}_{\alpha a_p}(\mathbf{x})$ is the corresponding covariance between $\alpha(\mathbf{x})$ and the N-dimensional vector $\mathbf{a}_p$ of pilot point values, $\mathbf{C}_{z_\alpha}$ is the covariance matrix of $\mathbf{z}_\alpha$, and $\bar{\mathbf{a}}_p$ and $\mathbf{C}_{a_p}$ are the specified mean and covariance of $\mathbf{a}_p$.

Equation (92) can be simplified if we assume both that the log conductivity measurements are noise-free and that the pilot point values are drawn from the same population (have the same statistics) as the true log conductivities. In this case the terms of (92) can be regrouped to conform more closely to the basis function expansion of (29):

$$\hat{\alpha}(\mathbf{x}) - \alpha_0(\mathbf{x}, \mathbf{z}_\alpha) = \sum_{j=1}^{N}\phi_i(\mathbf{x})(a_{pi} - \bar{a}_{pi}) = \boldsymbol{\phi}^T(\mathbf{a}_p - \bar{\mathbf{a}}_p) \quad (93)$$

where $\bar{a}_{pi} = \bar{\alpha}(x_{pi})$ and $\alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$ and $\phi_i(\mathbf{x})$ are given by

$$\alpha_0(\mathbf{x}, \mathbf{z}_\alpha) = \bar{\alpha}(\mathbf{x}) + \sum_{i=M_h+1}^{M_h+M_\alpha}\sum_{j=M_h+1}^{M_h+M_\alpha}[\mathbf{C}_\alpha(\mathbf{x})]_j[\mathbf{C}_\alpha^{-1}]_{ji}[z_{\alpha i} - \bar{\alpha}(\mathbf{x}_i)] \quad (94)$$

$$\phi_i(\mathbf{x}) = \sum_{j=1}^{N}[\mathbf{C}_\alpha(\mathbf{x})]_j[\mathbf{C}_\alpha^{-1}]_{ji} \quad (95)$$

Here $[\mathbf{C}_\alpha(\mathbf{x})]_j$ is the prior covariance between $\alpha(\mathbf{x})$ and $\alpha(\mathbf{x}_j)$ and $[\mathbf{C}_\alpha]_{ji}$ is the prior covariance between $\alpha(\mathbf{x}_j)$ and $\alpha(\mathbf{x}_i)$. These discrete covariances can be obtained by evaluating the prior covariance function $\mathbf{C}_\alpha(\mathbf{x}, \boldsymbol{\xi})$ at the appropriate locations. The function $\alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$ is the estimate of $\alpha(\mathbf{x})$ that would be obtained from a kriging algorithm that uses only actual log conductivity measurements (i.e., which does not use pilot points).

The basis function expansion of (93) demonstrates that the pilot point method uses a log conductivity parameterization which is consistent with (29). The only difference is that the pilot point version is based on an expansion of the function $\alpha(\mathbf{x}) - \alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$ rather than of $\alpha(\mathbf{x})$ itself. Note that $\alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$ depends only on the prior log conductivity mean and covariance functions and the log conductivity measurements. Since these quantities are all known, $\alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$ can be computed before $\mathbf{a}_p$ is estimated from head measurements.

The pilot point method uses a classical nonlinear least squares approach for estimating the unknown pilot point values. The estimation algorithm minimizes the following version of the least squares performance index given in (89):

$$J(\mathbf{a}_p) = \{\mathbf{z}_h - \mathcal{F}_h[\alpha_0 + \boldsymbol{\phi}^T(\mathbf{a}_p - \bar{\mathbf{a}}_p)]\}^T \cdot \mathbf{C}_{hv}^{-1}\{\mathbf{z}_h - \mathcal{F}_h[\alpha_0 + \boldsymbol{\phi}^T(\mathbf{a}_p - \bar{\mathbf{a}}_p)]\} \quad (96)$$

Note that only head measurements are included in this index. The minimum can be found with any of the search algorithms discussed earlier. Although the pilot point basis function expansion depends on the prior covariance of $\alpha(\mathbf{x})$, the performance index in (96) does not include a regularization term. The pilot point estimates are determined solely by their ability to fit head measurements and can deviate arbitrarily far from the log conductivity prior means without any penalty. It might be suspected that this could cause stability problems in some applications. Nevertheless, the method appears to have performed well in several field tests based on reasonably long records of transient head data [*Marsily et al*., 1984; *Certes and Marsily*, 1991; *LaVenue and Pickens*, 1992]. In these applications the number of pilot points was kept much less than the total number of measurements, a condition which helps to make the inverse problem better posed.

A variation on the original pilot point approach has been proposed by *Keidser et al*. [1990] and *Keidser and Rosbjerg* [1991], who introduced a regionalization technique that helps stabilize the estimation algorithm in both synthetic and field tests. Regionalization changes the definitions of $\alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$ and $\boldsymbol{\phi}(\mathbf{x})$ but has no effect on the form of the estimation performance index. *LaVenue and Pickens* [1992] have used adjoint sensitivity techniques to identify the pilot point locations that will provide the greatest improvement in the performance index. An adjoint-based location rule can be applied sequentially if new pilot points are added one-by-one until an acceptable solution is achieved. *Marsily* [1984], *Certes and Marsily* [1991], and *Keidser and Rosbjerg* [1991] all mention the benefits of including prior information and/or parameter constraints in the pilot point approach, but they do not appear to have done this in any of their applications. *LaVenue and Pickens* [1992] include prior information by constraining the pilot point estimates to lie within three error standard deviations of the kriged estimates (the $\alpha_0(\mathbf{x}, \mathbf{z}_\alpha)$'s). This is similar to the maximum a posteriori procedure described in (91).

The strength of the pilot point method lies in its ability to produce smooth log effective conductivity fields which are physically reasonable. The subjectivity involved in the original version of the method can be viewed as an asset, since it provides a flexible way to incorporate qualitative geological information. For example, a pilot point introduced in an area thought to have an anomalously high conductivity can convey quite a bit of information if the corresponding estimate is constrained to lie within a range of high values. We believe that the method's flexibility can be exploited most fully if prior information on the pilot point values is included explicitly, through the use of a Gaussian regularization term. This helps to stabilize the algorithm and makes it easier to control the influence of the pilot points. When prior information is included in this way, the pilot point method is essentially the same as the Gaussian maximum a posteriori estimator discussed earlier.

4.2.4. Extended Kalman filtering. The inverse techniques we have discussed up to this point are "batch" estimation algorithms that combine all available measurements in a single large measurement vector $\mathbf{z}$. Nonlinear batch algorithms update the parameter estimates with the entire measurement vector on each iteration. When measurements are collected at various times, it is possible to develop an iterative inverse algorithm which recursively updates estimates with only the most recent measurements. The most commonly used recursive nonlinear estimator is the extended Kalman filter, a generalization of the ordinary linear Kalman filter [*Jazwinski*, 1970].

In order to see how the extended Kalman filter is structured suppose that the M measurements used to estimate $\alpha(\mathbf{x})$ can be grouped into M_t sets taken at the discrete times $t_1, \cdots, t_{M_t}$. We assume, for simplicity, that each set contains M_x samples so that $M = M_t M_x$. If all of the measurements collected at time t_n are assembled in the M_x-dimensional measurement vector $\mathbf{z}_n$, the composite measurement vector $\mathbf{z}$ may be written as follows:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \cdots \\ \mathbf{z}_{M_t} \end{bmatrix} \tag{97}$$

The Kalman filter processes each of the $\mathbf{z}_n$ separately.

Kalman filters are based on a state equation, which describes how the parameter of interest evolves from one measurement time to another, and a measurement equation, which relates parameters to measurements. Since the large-scale log conductivity is a time-invariant parameter, the unknown parameter value at any given measurement time (say, t_n) is the same as the one at the previous time (t_{n-1}). If we focus on estimation of the discrete parameter vector $\mathbf{a}$ the state equation is simply

$$\mathbf{a}_n = \mathbf{a}_{n-1} \tag{98}$$

where $\mathbf{a}_n$ is the value of $\mathbf{a}$ at time t_n. This recursion is propagated from $n = 1, \cdots, M_t$. The statistics of the uncertain initial condition $\mathbf{a}_0$ are the prior mean and covariance of $\mathbf{a}$, $\bar{\mathbf{a}}$, and $\mathbf{C}_a$. The Kalman filter measurement equation at time t_n is constructed from the portion of the forward equation (42) which applies at this time:

$$\begin{aligned} \mathbf{z}_n &= \mathcal{F}_n(\boldsymbol{\phi}^T\mathbf{a}) + \mathbf{v}_n \\ &\approx \mathcal{F}_n(\boldsymbol{\phi}^T\bar{\mathbf{a}}) + \frac{\partial \mathcal{F}_n(\boldsymbol{\phi}^T\bar{\mathbf{a}})}{\partial \mathbf{a}} (\mathbf{a}_n - \bar{\mathbf{a}}) + \mathbf{v}_n \end{aligned} \tag{99}$$

where the M_x-dimensional vectors $\mathcal{F}_n$ and $\mathbf{v}_n$ are the portions of $\mathcal{F}$ and $\mathbf{v}$ associated with $\mathbf{z}_n$, and the M_x by N-dimensional Jacobian matrix $\partial \mathcal{F}_n / \partial \mathbf{a}$ is the portion of $\partial \mathcal{F} / \partial \mathbf{a}$ associated with $\mathbf{z}_n$. The final equality in (99) is a linear approximation which is obtained by introducing a first-order expansion for the forward operator.

The estimate produced by the extended Kalman filter at t_n is an approximation to the conditional mean of $\mathbf{a}$ given all previous measurements, written as $\hat{\mathbf{a}}_n = E[\mathbf{a}_n | \mathbf{z}_1, \cdots, \mathbf{z}_n]$. This estimate changes over time, even though the actual parameter value does not, because new information becomes available at each measurement time. *Anderson and Moore* [1979, pp. 39–40] and *Bryson and Ho* [1975, pp. 382–385] both show that the conditional mean can be computed with the following recursion when the state and (linearized) measurement equations of (98) and (99) apply and the initial state and measurement error are independent Gaussian variables:

$$\begin{aligned} \hat{\mathbf{a}}_n = \hat{\mathbf{a}}_{n-1} &+ \hat{\mathbf{C}}_{a,n-1} \frac{\partial \mathcal{F}_n^T(\boldsymbol{\phi}^T\hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} \\ &\cdot \left[\frac{\partial \mathcal{F}_n(\boldsymbol{\phi}^T\hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} \hat{\mathbf{C}}_{a,n-1} \frac{\partial \mathcal{F}_n^T(\boldsymbol{\phi}^T\hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} + \mathbf{C}_{vn} \right]^{-1} \\ &\cdot [\mathbf{z}_n - \mathcal{F}_n(\boldsymbol{\phi}^T\hat{\mathbf{a}}_{n-1})] \end{aligned} \tag{100}$$

The conditional covariance matrix which appears in this recursive estimation equation is also obtained from a recursion:

$$\hat{\mathbf{C}}_{an} = \hat{\mathbf{C}}_{a,n-1} - \hat{\mathbf{C}}_{a,n-1} \frac{\partial \mathcal{F}_n^T(\boldsymbol{\phi}^T \hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} \cdot \left[\frac{\partial \mathcal{F}_n(\boldsymbol{\phi}^T \hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} \hat{\mathbf{C}}_{a,n-1} \frac{\partial \mathcal{F}_n^T(\boldsymbol{\phi}^T \hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} + \mathbf{C}_{vn} \right]^{-1} \cdot \frac{\partial \mathcal{F}_n(\boldsymbol{\phi}^T \hat{\mathbf{a}}_{n-1})}{\partial \mathbf{a}} \hat{\mathbf{C}}_{a,n-1} \quad (101)$$

The conditional mean and covariance are initialized at time t_0 at the values of the corresponding unconditional (prior) statistics.

When the forward operator and measurement equation are truly linear in $\mathbf{a}$, the estimate produced by the extended Kalman filtering algorithm after the final measurement update at t_{M_t} is the true conditional mean $E[\mathbf{a}|\mathbf{z}_1, \cdots, \mathbf{z}_{M_t}]$. In this case the extended Kalman filter estimate is identical to the maximum a posteriori estimate since the a posteriori density is Gaussian with a mode which is equal to the mean (recall the discussion accompanying Figure 4). When the forward operator is nonlinear the extended Kalman filter estimate does not necessarily converge to either the conditional mean or the conditional mode. Of course, the hope is that the linearization is sufficiently accurate to insure that the estimate will be "close" to the conditional mean. It is possible to view the temporal sequence of extended Kalman filter updates as an iterative series of reduced-dimensionality linear batch solutions, with the prior mean and covariance for each new solution set equal to the conditional mean and covariance obtained from the previous solution. This can be seen by comparing (100) to (73).

Since the extended Kalman filter uses less information on each iteration than the batch maximum a posteriori algorithm, it is less likely to converge to an acceptable solution. *Ljung* [1979] discusses the convergence problems of the extended Kalman filter and proposes a modified "innovations form" of the algorithm which addresses some of these problems. The innovations form does not appear to have been applied to distributed parameter inverse problems, but it deserves consideration when the total number of states is sufficiently small to make it computationally feasible. *Graham and McLaughlin* [1991] applied a conventional version of the extended Kalman filter to a groundwater transport problem with a few thousand unknowns and a small number of measurements collected at three times. The number of time steps used in this case study is too small to provide an assessment of the filter's convergence properties. We are unaware of other applications of the extended Kalman filter to field problems with a significant number of unknowns.

The extended Kalman filter is most useful in applications where the number of unknowns is relatively small (e.g., where a relatively coarse block parameterization is used to describe spatial variability) and measurements are taken at many times. *Wilson et al*. [1978] and *Townley* [1983] describe successful synthetic experiments with small problems of this type. Although it has some attractive features, we believe that the extended Kalman filter should not be considered a practical inverse method until its convergence properties are better understood and it has been tested more extensively.

4.3. Discussion

We have seen in this section that the linear inverse techniques introduced to the groundwater community by *Dagan* [1985], *Gutjahr and Wilson* [1985], *Hoeksema and Kitanidis* [1984], and *Sun and Yeh* [1992] are all special cases of the Gaussian maximum a posteriori algorithm developed in section 3. The linear estimation approach has the important advantage of providing a unique solution to an approximate version of the maximum a posteriori problem. This uniqueness suggests that the linearized inverse problem should be well-posed. However, even a problem with a unique solution can be ill-posed if the solution is overly sensitive to small changes in the measured data. Such sensitivity can arise in the algorithm of (79) if the matrix $\mathbf{C}_z$ is ill-conditioned (near singular) [*Dietrich and Newsam*, 1989]. Near-singular behavior most often occurs when closely spaced measurements are used to estimate distant log conductivity values or when the problem structure desensitizes the measurements to log conductivity variations in certain portions of the domain. Algorithmic singularity can be avoided, or at least mitigated, if the inverse problem is designed to maximize the sensitivity of measurements to parameters. Whenever possible, the groundwater system should be stressed, more measurements and/or new kinds of measurements should be added, and measurement noise should be properly accounted for (see section 5.3 for further discussion of these points).

The various nonlinear inverse methods used for groundwater problems have gradually evolved beyond the classical least squares perspective to include prior information. This information may enter as an explicit Bayesian probability density, in the form of parameter bounds, or through the use of kriging as an interpolation algorithm. In all of these cases the net result is to stabilize (or regularize) the inverse problem by restricting the set of admissible estimates. Most of the available methods for including prior information are either equivalent to or minor variants on the Gaussian maximum a posteriori approach.

The primary difficulty encountered with nonlinear inverse methods is failure to converge to a reasonable solution. When the forward operator is nonlinear the estimation performance index may have many local minima or may be very flat in the vicinity of a minimum. Iterative search procedures have trouble finding such minima and those that work well in one problem may not work well in others. Such convergence problems are usually caused by an ill-posed problem formulation and can ultimately only be resolved by changing this formulation. As in the case of linear inverse algorithms, convergence can be improved by introducing problem modifications that increase the sensitivity of measurements to parameters.

It is easy to imagine a number of new finite-dimensional parameterizations which could be incorporated into the nonlinear estimation approach. Methods which have not yet been applied to practical groundwater problems include expansions in covariance operator eigenfunctions [*Braud et al*., 1993], splines [*Wahba*, 1990], or wavelets [*Liu*, 1993]. Other alternatives which merit consideration are variants on the pilot point approach and on the "two-scale" approach to prior information proposed by *Cooley* [1982, 1983]. There is no reason why the basis functions used in a nonlinear inverse method need to be the same on every iteration. For example, the basis functions used in the pilot point method could be updated on each iteration of the pilot point algo-

rithm to account for the effect of head conditioning. Although such enhancements are worth investigating, the ultimate success of a parameterization technique can be expected to depend more on its geological credibility than its conceptual elegance. Simple methods may work best if they are able to capture the key features of a site.

5. Related Issues

In this section we briefly discuss a number of topics which are closely related to inverse estimation and are of considerable interest in applications. These topics are all concerned in some way with the evaluation and improvement of estimates obtained from inverse algorithms. Our primary objective is to state the relevant issues and questions. More detailed discussions may be found in the cited references.

5.1. Model Error and Prior Statistics

The Bayesian formulation of the inverse problem presented in section 3 assumes that the prior statistics, measurement error statistics, and forward operator are all known perfectly. These idealized assumptions can be questioned in practical applications, where models can be misapplied and initial conditions, boundary conditions, and forcing terms are usually uncertain. Incorrectly specified model equations and statistics can have an adverse impact on inverse procedures because the parameter estimates are forced to compensate for unacknowledged errors. A frequently encountered example is the tendency of inverse algorithms to generate unrealistic log conductivity estimates when recharge is neglected in the forward model. This happens because the conductivities are adjusted to reproduce head variations which are, in fact, a response to recharge. Although the adjusted conductivities may give a good match to observed heads, they have little value for predictive modeling.

It is difficult to develop a general way to account for fundamental errors in model structure, such as those that might arise when key processes are ignored or when model boundaries are incorrectly specified [*McLaughlin and Wood*, 1988a, b]. In some cases it may be possible to use hypothesis tests to check the structural validity of a model designed to reproduce field observations [*Bard*, 1974; *Luis and McLaughlin*, 1992]. But these model testing procedures typically depend on statistical assumptions that are difficult to verify in practice. Most investigations of model error consider only errors in specifying the parameters and auxiliary conditions of structurally correct model equations [*Townley and Wilson*, 1985; *Carrera and Neuman*, 1986a, b, c]. As mentioned in section 2, time-invariant sources of uncertainty can be included in an expanded vector of parameter functions. When the uncertain inputs are time-dependent, it may be necessary to use other approaches. For example, *McLaughlin* [1979] and *te Stroet* [1995] have accounted for recharge uncertainty by including "system noise" in recursive inverse algorithms.

Non-Bayesian methods that deal with model error by increasing the number of unknowns risk the danger of making the inverse problem less well-posed by requiring that more information be extracted from a limited number of measurements. Bayesian methods that take the same approach demand greater amounts of prior information. In either case there is a tradeoff between the benefits to be gained by accounting for more sources of model error and the costs imposed by increasing the difficulty of the inverse problem [*McLaughlin and Wood*, 1988a, b]. There are few, if any, studies of this tradeoff in the literature. It should be noted that as more and more sources of model error are included, an inverse algorithm will be less inclined to believe its forward equation. In the extreme case when model errors are thought to be very large, there is no reason to solve the inverse problem at all since the algorithm's only option is to return measured values of the parameter where they are available and prior means elsewhere. Although everyone agrees that model error is an important issue, there is still no general way to deal with it or even to assess its importance.

Since Bayesian prior information is based on a probabilistic model of parameter variability, deficiencies in specified prior statistics can be viewed as a form of model error. Even the modest single-parameter Bayesian inverse problem posed in this paper requires a significant amount of prior statistical information. If we assume (1) that the measurement errors are zero mean and uncorrelated with one another and the log conductivity and (2) that the log conductivity function has the two-scale correlation structure discussed in section 2, we must specify, at a minimum, the measurement error variance, the small-scale log conductivity variance and correlation distance, and the large-scale log conductivity mean, variance, and correlation distance. The small-scale log conductivity statistics are needed to derive the effective conductivity function $\mathbf{K}_{eff}(\alpha)$ while the large-scale statistics are needed to define $\bar{\alpha}$ and C_α. The two basic alternatives for obtaining these measurement and prior statistics are to infer them from available field measurements, using methods such as maximum likelihood [*Schweppe*, 1973] or cross validation [*Wahba*, 1990], and to specify them independently in order to achieve certain smoothing objectives.

The statistical inference approach has been applied to the groundwater inverse problem by *Kitanidis and Vomvoris* [1983], *Hoeksema and Kitanidis* [1984], *Rubin and Dagan* [1987a, b] and *Samper and Neuman* [1989]. These authors assume that the structure of the log conductivity covariance (or variogram) function is known so that the inference procedure can focus on the estimation of the variance and possibly a few variogram coefficients. There are relatively few discussions in the literature of the relationship between the accuracy of estimated sample statistics and the number of measurements used for estimation. *Russo and Jury* [1987a, b] and *Gelhar* [1993] both suggest that large sample sizes (e.g., hundreds) are required to give "acceptable" sample estimates of covariance properties.

An alternative to the inference approach is to treat prior statistics as regularization parameters that control the smoothness of the inverse solution. This interpretation is supported by our earlier (section 3.3) discussion of the regularization term in the performance index and by the analysis presented by *Wahba* [1990]. It is clear that we can force a Bayesian inverse algorithm to estimate larger- or smaller-scale trends by adjusting its regularization parameters, much as we might adjust the window size in a moving window average. One might argue that it is not essential, or even desirable, to estimate regularization parameters from field data. Instead, they should be viewed as inputs that can be used to explicitly define what we mean by "large" and "small" scale. This interpretation is likely to be applied by default whenever field data are too limited to support estimation of the complete suite of sample statistics needed to solve a Bayesian inverse problem.

5.2. Estimation and Prediction Accuracy

Much of the mathematical inverse literature treats parameter estimation as an end in itself and says relatively little about how parameter estimates might be used. By contrast, groundwater hydrologists are often most interested in the prediction of dependent variables such as head or concentration. Parameter estimation is really only a means to that end. In such cases, prediction accuracy may be the most appropriate measure of estimation performance.

Equation (36) suggests that the head prediction at $(\mathbf{x}, t)$ should be computed from $\hat{h}(\mathbf{x}, t) = \mathcal{F}_h(\hat{\alpha})(\mathbf{x}, t)$ (this is, in fact, the maximum a posteriori estimate of $h(\mathbf{x}, t)$ when the prediction operator $\mathcal{F}_h(\mathbf{x})$ is invertible and there is no model error). The accuracy of this prediction is related to the head prediction error $\tilde{h} = h - \hat{h}$ which is, in turn, related to the parameter estimation error $\tilde{\alpha} = \alpha - \hat{\alpha}$. The "size" of these errors can be measured in terms of their variances or, more generally, their covariance functions. This variance-based approach to accuracy evaluation has certain limitations, especially for risk analysis applications concerned with low-probability events, but it provides a useful framework for examining ill-posedness and related experimental design issues.

We consider, for simplicity, the case where the large-scale log conductivity can be characterized by a finite-dimensional parameter vector $\mathbf{a}$. Appendix D shows that the covariance of the discrete parameter estimation error $\tilde{\mathbf{a}} = \mathbf{a} - \hat{\mathbf{a}}$ satisfies the following inequality when (1) the parameter and measurement error vectors are independent and Gaussian (as we have assumed in earlier discussions), (2) the forward operator is approximated by a linear expansion about the known estimate $\hat{\mathbf{a}}$ (see (D7) and (D11)), and (3) the estimate bias is negligible:

$$\mathbf{C}_{\tilde{a}} = E_{z\alpha}\{\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T\} \geq \mathbf{B}^{-1} \tag{102}$$

where $\mathbf{B}$ is the "Bayesian information matrix," defined as follows:

$$\mathbf{B} = \frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\hat{\mathbf{a}})}{\partial \mathbf{a}} \mathbf{C}_v^{-1} \frac{\partial \mathcal{F}(\boldsymbol{\phi}^T\hat{\mathbf{a}})}{\partial \mathbf{a}} + \mathbf{C}_\alpha^{-1} \tag{103}$$

and $\partial \mathcal{F}(\boldsymbol{\phi}^T\hat{\mathbf{a}})/\partial \mathbf{a}$ is the measurement Jacobian matrix evaluated at the known estimate $\hat{\mathbf{a}}$. The inequality in (102) is understood to indicate that the difference between the matrices on the left and right sides is positive semidefinite.

Equation (102) is a Bayesian generalization of the Cramer-Rao bound of maximum likelihood theory. This bound relies on relatively few assumptions and applies for any estimation technique. The two terms in the information matrix expression describe the information provided by measurements and prior information, respectively. The first term will be small if the measurements are insensitive to the parameter vector or if the measurement error is large. The second term will be small if the prior statistics are uninformative. Anything which increases parameter sensitivity or reduces measurement or parameter uncertainty will increase the information gained and reduce the estimation error.

An approximate expression for the prediction error variance can be obtained by combining the Cramer-Rao bound with a linearized version of the head prediction equation. The head prediction error at $(\mathbf{x}, t)$ can be related to the parameter estimation error if the head prediction operator is expanded to first-order about the known parameter estimate $\hat{\mathbf{a}}$:

$$\tilde{h}(\mathbf{x}, t) = h(\mathbf{x}, t) - \hat{h}(\mathbf{x}, t) = \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})(\mathbf{x}, t) - h(\mathbf{x}, t)$$
$$\approx \frac{\partial \mathcal{F}_h(\boldsymbol{\phi}^T\hat{\mathbf{a}})(\mathbf{x}, t)}{\partial \mathbf{a}} (\mathbf{a} - \hat{\mathbf{a}}) \tag{104}$$

where $\partial \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a}(\mathbf{x}, t)/\partial \mathbf{a}$ is a prediction Jacobian vector which specifies the sensitivity of $h(\mathbf{x}, t)$ at $(\mathbf{x}, t)$ to each component of $\mathbf{a}$. The components of this vector, like those of the measurement Jacobian matrix, may be expressed as integrals over the flow equation Green's function (see (56) and (58)). In the case of the prediction Jacobian the Green's function is evaluated at the prediction point $(\mathbf{x}, t)$ rather than the measurement point $(\mathbf{x}_i, t_i)$.

A lower bound on the head prediction error variance at $(\mathbf{x}, t)$ is obtained by squaring each side of (104), taking the expectation, and substituting (102) for the estimation error covariance:

$$\sigma_h^2(\mathbf{x}, t) \approx \frac{\partial \mathcal{F}_h^T(\boldsymbol{\phi}^T\hat{\mathbf{a}})(\mathbf{x}, t)}{\partial \mathbf{a}} \mathbf{C}_{\tilde{\alpha}} \frac{\partial \mathcal{F}_h(\phi^T\hat{\mathbf{a}})(\mathbf{x}, t)}{\partial \mathbf{a}}$$
$$\geq \frac{\partial \mathcal{F}_h^T(\boldsymbol{\phi}^T\hat{\mathbf{a}})(\mathbf{x}, t)}{\partial \mathbf{a}}$$
$$\cdot \left[\frac{\partial \mathcal{F}^T(\boldsymbol{\phi}^T\hat{\mathbf{a}})}{\partial \mathbf{a}} \mathbf{C}_v^{-1} \frac{\partial \mathcal{F}(\phi^T\hat{\mathbf{a}})}{\partial \mathbf{a}} + \mathbf{C}_\alpha^{-1} \right]^{-1}$$
$$\cdot \frac{\partial \mathcal{F}_h(\boldsymbol{\phi}^T\hat{\mathbf{a}})(\mathbf{x}, t)}{\partial \mathbf{a}} \tag{105}$$

This expression indicates that the prediction error depends on both the measurement and prediction Jacobians. All of the quantities appearing in the prediction error variance bound can be evaluated once the final estimate $\hat{\mathbf{a}}$ has been computed. Some of the computational issues involved in deriving the Jacobians are discussed in Appendix B.

The Cramer-Rao estimation and prediction error bounds provide convenient measures of the performance of an inverse procedure since they account for all of the important factors which influence prediction accuracy. The extra effort required to derive the bound is usually minimal if a Gauss-Newton or so-called variable metric search algorithm is used since the information matrix is evaluated as part of the search procedure anyway [*Tarantola*, 1987, pp. 243–248]. For these reasons, we believe that the lower covariance bound (or information matrix) should be routinely evaluated in groundwater inverse algorithms.

5.3. Identifiability and Experimental Design

Although identifiability is a recurrent theme in both theoretical and applied discussions of the groundwater inverse problem, there is no generally accepted definition of the term. For present purposes, we can say that a parameter is identifiable if it can be estimated with "reasonable" accuracy from available measurements. In order for a parameter to be identifiable in this qualitative sense, the underlying inverse problem should be well-posed. That is why most theoretical studies of identifiability concentrate on the classical issues of existence, uniqueness, and stability [*Banks and Kunisch*, 1989]. These studies typically take a deterministic perspective which assumes that the measured dependent variable function (e.g., head) is known perfectly everywhere and that the identifiable parameter must be estimated without error. *Chavent* [1991]

describes an alternative deterministic approach which accounts for data limitations and estimation error.

The accuracy bounds discussed in the previous subsection suggest that parameter identifiability can also be addressed from a probabilistic perspective. One option is to specify that a parameter vector is identifiable if the prediction error variance bound is smaller than some threshold. The value of the threshold selected is probably less important than a recognition of the factors that control the bound. These factors are (1) the parameterization adopted for $\alpha(\mathbf{x})$, as reflected in the basis function vector $\boldsymbol{\phi}(\mathbf{x})$ and the prior statistics of $\mathbf{a}$; (2) the sampling times and locations of the measurements used to estimate $\mathbf{a}$, as reflected in the forward operator $\mathcal{F}(\boldsymbol{\phi}^T\mathbf{a})$ and the corresponding measurement Jacobian; (3) measurement uncertainty, as reflected in the measurement error covariance matrix $\mathbf{C}_v$; (4) the forcing and boundary conditions applied to the groundwater system, as reflected in the forward operator; and (5) the locations and times where predictions are required, as reflected in the prediction Jacobian. A practical and comprehensive theory of identifiability should account for all of these factors.

Although it is difficult to check inverse problems in advance to see if a particular set of parameters is identifiable, it is possible to develop some general guidelines which suggest how we can improve parameter identifiability in practice. All of these share the common principle that the inverse problem should be constructed to maximize the sensitivity of measured variables to estimated parameters. Sensitivity can be increased by changing the problem parameterization, by adding or improving measurements, or by stressing the groundwater system. These options are discussed further in the following paragraphs.

One of the most difficult challenges of inverse estimation is the need to find parameterizations which are simple enough to yield well-posed problems but complex enough to capture the spatial heterogeneity found at real field sites. The blocked parameterization technique mentioned in the beginning of this paper generally yields a small number of unknown parameters which are associated with well-defined geological features. If the heads at measurement points are sensitive to these blocked parameters and the number of unknowns is much smaller than the number of measurements, it is likely that the resulting inverse problem will be well-posed. A classic example is estimation of effective hydraulic conductivities in a system composed of a few distinct geological layers, using transient pump test data [*Carrera and Neuman*, 1986c]. If the layers are properly delineated, the prior values for the layer conductivities are reasonably accurate, and the forcing applied during the pump test is significant most inverse algorithms will give reasonable results.

The primary problem with the blocked approach is its dependence on block geometry, which usually is assumed to be known perfectly. If the actual spatial distribution of hydraulic conductivity is not blocked or if the block boundaries are incorrectly specified, the inverse algorithm may be forced to generate unrealistic estimates in order to provide a good fit to head measurements. So, although the inverse problem may be well-posed in the sense that it yields a stable solution, the estimates it provides may not properly characterize the subsurface environment.

The geostatistical approach to parameterization avoids the problem of specifying block geometry by treating the effective log conductivity as a random field which can take on a different value at every point in the domain of interest. This does not imply that the geostatistical approach is actually able to estimate an infinite number of independent parameters from a finite number of measurements. The correlation properties of the random field impose spatial structure (or smoothness) on the estimates. Generally speaking, prior correlation scales that are small compared to the sample spacing yield inverse estimates that are close to the prior mean except in the immediate vicinity of measurements. Conversely, correlation scales that are large compared to the sample spacing yield smooth estimates which vary gradually over the region of interest.

This discussion suggests that the blocked and geostatistical approaches are appropriate in different situations. The choice between the two alternatives depends both on the geological setting and on the configuration of the measurements available for estimation. The maximum a posteriori formulation of the inverse problem summarized earlier in this paper is sufficiently general to permit the use of hybrid parameterizations which combine the blocked and geostatistical approaches. A simple way to accomplish this is to allow the prior mean of the random log conductivity function to take on different (unknown) values in different blocks. Since the mean values in the specified blocks appear linearly in the basis function expansion, their values may be estimated as additional parameters.

It is sometimes useful to view the acquisition of measurements for inverse estimation as an experiment. This is particularly true when there is an opportunity to specify sampling locations or to carry out hydraulic tests. The discussion of parameter accuracy presented in section 5.2 suggests that experimental designs which increase the sensitivity of measurements to parameters will have a beneficial effect on inverse estimation. The sensitivity of measurements to the discrete parameter vector $\mathbf{a}$ is described by the measurement Jacobian $\partial\mathcal{F}^T(\boldsymbol{\phi}^T\hat{\mathbf{a}})/\partial\mathbf{a}$. When the forward operator is linear this Jacobian depends on the sampling times and locations but not on the actual values of the samples. Consequently, we can identify an "optimal" sampling strategy before any measurements are actually collected [*Delhomme*, 1978]. This can be done by comparing, for example, the prediction error variance obtained with a number of different candidate strategies and selecting the strategy which gives the best accuracy for a given cost. Similar comments apply to the identification of optimal pump test strategies.

When the forward operator is nonlinear, the situation becomes more complicated since the measurement Jacobian depends on the actual measured values. In this case we cannot identify an "optimal" experimental design a priori but must construct it gradually, over a number of sequential sampling rounds (or iterations). On the first iteration the Jacobian can be evaluated about the prior mean and a few sampling times and locations selected. After the samples are collected and the inverse estimate is computed the Jacobian can be reevaluated about the new estimate and then used to design the next sampling round. This process continues until either the prediction error is reduced to an acceptable level or experimental resources are exhausted. The advantage of an iterative approach is that it allows the experimental design to evolve in response to new information. The disadvantage is that it requires more time and is more difficult to implement than a "one-shot" design. There are a number of variants on the iterative (or sequential) approach to experimental design. Examples of two applications to groundwater solute transport

modeling are provided by *Graham and McLaughlin* [1989, 1991].

It may be that the major contribution of a statistically based approach to experimental design is the emphasis it puts on the role of measurement sensitivity. In many situations where parameters are difficult to identify it is easier to improve sensitivity by introducing new kinds of information than by simply adding more measurements of the same variable. This is particularly true in the groundwater flow inverse problem. Since steady state heads are relatively insensitive to spatial variations in hydraulic conductivity, estimation performance may not improve dramatically when more head measurements are included. A smaller number of measurements of other variables such as groundwater solute concentration may prove to be more valuable [*McLaughlin et al*., 1993; *Reid*, 1996]. Another attractive option is to supplement traditional borehole data with synoptic geophysical measurements. Although geophysical data cannot generally be used to measure log conductivity directly, ground-penetrating radar, seismic, and electrical resistivity observations can, in some cases, identify the boundaries of lithologic features that have hydrologic significance [*Han et al*., 1986; *Hyndman et al*., 1994]. Such information can be incorporated into either blocked or geostatistical inverse parameterizations. There is considerable benefit to be gained by including many different types of information in the inverse estimation process. The information matrix approach provides a quantitative way to evaluate this benefit.

5.4. Discussion

Our survey of "related issues" suggests that the practical problems of inverse estimation extend well beyond the task of selecting an appropriate estimation algorithm. Any field application requires a number of different design decisions that can have a dramatic effect on the success of an inverse estimation effort. These include (1) selection of the parameters to be estimated, (2) selection of a parameterization technique, (3) specification of prior statistics or regularization parameters, (4) selection of the kinds of measurements to be used for estimation, and (5) design of sampling strategies and hydraulic tests. When we consider the mathematical similarity of the inverse methods in current use, these site-specific design factors appear even more important. We are convinced that the most beneficial advances in inverse estimation will be those which provide better ways to account for the special hydrogeologic characteristics of individual sites, either through more flexible methods for describing spatial variability or through the ability to incorporate a wider range of field information.

6. Summary and Conclusions

The inverse problem considered in this paper is concerned with the estimation of physical properties (such as the log hydraulic conductivity) which vary over space. If such properties are treated as random functions it is possible to develop a general inverse theory which can accommodate many different descriptions of hydrogeologic variability. These include blocked descriptions which divide the region of interest into distinct geological units, geostatistical descriptions which treat physical properties as random fields, and hybrid approaches which combine aspects of each.

Functional approaches to the groundwater inverse problem seek "optimum" estimates which are spatial functions rather than vectors of discrete parameters. Recent developments in probability theory establish the mathematical basis for a functional definition of optimality [*Larkin*, 1972; *Fitzpatrick*, 1991]. In particular, it is possible to define a "most probable" or maximum a posteriori estimate if the unknown random function is Gaussian. This estimate can be obtained by minimizing a "regularized" version of a classical least squares performance index. The regularization term in the performance index is generally a spatial integral which weights derivatives of the estimated parameter function. The minimization procedure favors estimates which give reasonable fits to measurements while remaining smooth enough to keep the derivative terms small. Although it is possible to apply a regularized least squares approach to the inverse problem without relying on probabilistic arguments, the maximum a posteriori viewpoint provides a convenient basis for relating regularization parameters to observable hydrogeological properties such as the variance and correlation scale of the log hydraulic conductivity.

When any inverse approach, including the maximum a posteriori approach, is applied to field problems, care must be taken to insure that the estimation objectives are consistent with the information available. One way to do this is to adopt a two-scale description of variability which clearly distinguishes large-scale trends that can be estimated from available measurements from small-scale fluctuations. In this case, point observations are assumed to be the sum of the large-scale trend, the small-scale fluctuation, and a random noise term which accounts for instrument and recording errors. In groundwater applications the large-scale piezometric head is the solution to an "upscaled" groundwater equation which depends on an effective hydraulic conductivity tensor. No attempt is made to predict smaller-scale head fluctuations. The effective conductivity is a function of the large-scale log conductivity and the variance and correlation scale of the small-scale log conductivity. When the small-scale log conductivity statistics are given, the only remaining unknown is the large-scale log conductivity, which is the parameter estimated in the inverse procedure. This two-scale approach to the groundwater inverse problem permits the scale of estimation to be defined explicitly, through specification of the prior log conductivity statistics.

The solution of the functional least squares problem is formally equivalent to the solution of the classical discrete parameter least squares problem. Consequently, most of the solution algorithms and analytical tools described in texts such as those by *Schweppe* [1973], *Bard* [1974], and *Gill et al*. [1981] apply directly to the functional case. This idea is developed in detail by *Tarantola* [1987]. The functional alternative is attractive primarily because it clearly distinguishes assumptions about natural variability from discretization assumptions introduced to obtain numerical solutions. The functional estimation equations and related solution algorithms can be clearly stated in an integral or differential form which applies for a wide range of different discretization schemes. Any particular discretization can be introduced by replacing the log conductivity function by a finite-dimensional expansion over a set of specified basis functions. Classical numerical methods can then be used to solve for the unknown basis function coefficients.

When this perspective is taken it becomes apparent that most common groundwater inverse algorithms are special cases of the functional maximum a posteriori estimator. The primary factors distinguishing these algorithms are the particular parameterization adopted (e.g., blocked versus geostatis-

tical) and the type of search procedure used to find the least squares solution (e.g., single versus multiple iteration). All of the alternatives have merit, and it is not possible to say that any single option is "best." Particular inverse methods are, however, more appropriate in some situations than in others. When geological structure is apparent and formation boundaries are distinct, a blocked approach to parameterization is probably the best choice. When hydrogeologic properties appear to vary in a more or less random fashion and there is no clearly defined structure, a geostatistical parameterization may be preferable. Single-iteration linear methods may be more appropriate when time is limited and convenient analytical approximations can be used to simplify the inverse solution. Multi-iteration nonlinear methods may be more appropriate in complex problems, especially those that require estimation of several different parameters (e.g., log conductivity, recharge rates, and boundary conditions).

In any case, we should recognize that inverse estimation is rarely the final step in a groundwater flow analysis. It is important to know how good the inverse estimates are and to evaluate the effect of estimation uncertainty on model predictions. A Bayesian approach to the inverse problem provides a convenient way to account for multiple sources of uncertainty and to consider the many different factors that affect prediction accuracy. These include (1) the parameterization adopted; (2) the sampling times, locations, and accuracy of the measurements used for estimation; (3) the forcing and boundary conditions applied to the groundwater system; and (4) the locations and times where predictions are required. An integrated approach to the inverse problem should consider all of these factors and should design the inverse experiment so as to maximize the sensitivity of measurements to estimated parameters. Although theoretical insights can help us understand how to analyze and improve inverse algorithms, the best way to insure success is to incorporate as much information as possible about the specific application of interest into the problem formulation. If we take care in formulating the problem and have realistic expectations, inverse procedures can yield useful estimates which are geologically meaningful and consistent with observed data.

Appendix A: Alternative Forms for the Prior Term of the Gaussian Maximum a Posteriori Performance Index

The inverse covariance $C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})$ of α is defined by the following identity (see (50)):

$$\int_D \int_D C_\alpha(\mathbf{x}, \boldsymbol{\mu}) C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) \alpha(\boldsymbol{\xi})\, d\boldsymbol{\mu}\, d\boldsymbol{\xi} = \alpha(\mathbf{x}) \tag{A1}$$

where $C_\alpha(\mathbf{x}, \boldsymbol{\xi})$ is the covariance between $\alpha(\mathbf{x})$ and $\alpha(\boldsymbol{\xi})$. When the dimension N of the parameter space A is finite, (29) may be substituted into (A1) to give

$$\int_D \int_D \boldsymbol{\phi}^T(\mathbf{x}) \mathbf{C}_a \boldsymbol{\phi}(\boldsymbol{\mu}) C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) \boldsymbol{\phi}^T(\boldsymbol{\xi}) \mathbf{a}\, d\boldsymbol{\mu}\, d\boldsymbol{\xi} - \boldsymbol{\phi}^T(\mathbf{x}) \mathbf{a}$$

$$= \boldsymbol{\phi}^T(\mathbf{x}) \left[\mathbf{C}_a \int_D \int_D \boldsymbol{\phi}(\boldsymbol{\mu}) C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) \cdot \boldsymbol{\phi}^T(\boldsymbol{\xi})\, d\boldsymbol{\mu}\, d\boldsymbol{\xi} - \mathbf{I} \right] \mathbf{a} = 0 \tag{A2}$$

where $\mathbf{I}$ is the N-dimensional identity matrix and (31) has been used to relate the covariance function $C_\alpha(\mathbf{x}, \boldsymbol{\mu})$ to the covariance matrix $\mathbf{C}_\alpha$. Since the basis functions are linearly independent, this identity can hold only if the matrix expression in brackets is 0. It follows that the integral is equal to the inverse of the discrete covariance matrix:

$$\int_D \int_D \boldsymbol{\phi}(\boldsymbol{\mu}) C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) \boldsymbol{\phi}^T(\boldsymbol{\xi})\, d\boldsymbol{\mu}\, d\boldsymbol{\xi} = \mathbf{C}_a^{-1} \tag{A3}$$

If we postmultiply each side of this matrix identity by the mean-removed parameter vector $[\mathbf{a} - \bar{\mathbf{a}}]$ and premultiply by $[\mathbf{a} - \bar{\mathbf{a}}]^T$ the result is

$$\int_D \int_D [\mathbf{a} - \bar{\mathbf{a}}]^T \boldsymbol{\phi}(\boldsymbol{\mu}) C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) \boldsymbol{\phi}^T(\boldsymbol{\xi}) [\mathbf{a} - \bar{\mathbf{a}}]\, d\boldsymbol{\mu}\, d\boldsymbol{\xi}$$

$$= \int_D \int_D [\alpha(\boldsymbol{\mu}) - \bar{\alpha}(\boldsymbol{\mu})] C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) [\alpha(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})] \cdot d\boldsymbol{\mu}\, d\boldsymbol{\xi} = [\mathbf{a} - \bar{\boldsymbol{\alpha}}]^T \mathbf{C}_a^{-1} [\mathbf{a} - \bar{\boldsymbol{\alpha}}] \tag{A4}$$

In the first equality, (29) is used to replace $\boldsymbol{\phi}^T(\mathbf{x})\mathbf{a}$ and its mean with $\alpha(\mathbf{x})$ and its mean, respectively. Equation (A4) establishes a correspondence between the quadratic forms which appear in the prior terms of the functional and discrete versions of the Gaussian maximum a posteriori performance index.

Although (A4) holds only for finite N it is possible to derive a similar identity for the infinite-dimensional case by using the Karhunen-Loeve expansion for $\alpha(\mathbf{x})$ and replacing matrix operations with infinite summations (see (30) and (31)). In this case, (A4) may be written as

$$\int_D \int_D [\alpha(\boldsymbol{\mu}) - \bar{\alpha}(\boldsymbol{\mu})] C_\alpha^{-1}(\boldsymbol{\mu}, \boldsymbol{\xi}) [\alpha(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\, d\boldsymbol{\mu}\, d\boldsymbol{\xi} = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} [\mathbf{a}_i - \bar{\mathbf{a}}_i]^2 \tag{A5}$$

where λ_i is the eigenvalue associated with the ith Karhunen-Loeve basis function.

Appendix B: Computation of the Measurement Jacobian and Performance Index Gradient

This appendix describes how the measurement Jacobian used in the Gauss-Newton search and the performance index gradient used in gradient-based searches can be derived for groundwater flow applications. Computational details are discussed in the cited references.

The functional measurement Jacobian $\mathbf{g}_{\mathscr{F}}$ provides information about the sensitivity of predictions to changes in the large-scale log conductivity function. This Jacobian may be derived from the forward equation which forms the basis for the groundwater inverse problem. It is useful to consider the log conductivity and head components of the Jacobian separately since they are treated rather differently. Component i of the log conductivity portion of $\mathbf{g}_{\mathscr{F}}$ is associated with a point log conductivity measurement taken at location $\mathbf{x}_i$. It can be identified by noting that the first variation of (39) at $\mathbf{x}_i$ is

$$\delta\alpha(\mathbf{x}_i) = \frac{\partial \mathcal{F}_{\alpha i}}{\partial \alpha}\,\delta\alpha = \int_D \delta(\mathbf{x}_i - \boldsymbol{\xi})\delta\alpha(\boldsymbol{\xi})\,d\boldsymbol{\xi} \qquad \text{(B1)}$$

It follows from the definition of (56) that

$$g_{\mathcal{F}_{\alpha i}}(\boldsymbol{\xi}|\alpha) = g_{\mathcal{F}_{\alpha i}}(\boldsymbol{\xi}) = \delta(\mathbf{x}_i - \boldsymbol{\xi}) \qquad \text{(B2)}$$

The head components of the Jacobian are more difficult to obtain since they must be derived from the groundwater equation given in (35).

Component i of the head component of $\mathbf{g}_{\mathcal{F}}$ is defined in (56). If we take the first variation of (37) about some nominal parameter function $\alpha_0(\mathbf{x})$ and apply this definition, the result is

$$\delta h(\mathbf{x}_i, t_i) = \frac{\partial \mathcal{F}_{hi}(\alpha_0)}{\partial \alpha}\,\delta\alpha = \int_D g_{\mathcal{F}_{hi}}(\boldsymbol{\xi}|\alpha_0)\delta\alpha(\boldsymbol{\xi})\,d\boldsymbol{\xi} \qquad \text{(B3)}$$

Our goal is to find an explicit integral expression for $\delta h(\mathbf{x}_i, t_i)$ which can be used to identify $g_{\mathcal{F}_{hi}}(\mathbf{x}|\alpha_0)$. A differential expression for δh at any $\mathbf{x}$ and t can be obtained by taking the first variation of (35) about $\alpha_0(\mathbf{x})$:

$$S\frac{\partial \delta h}{\partial t} - \nabla\cdot\mathbf{K}_{eff}(\alpha_0)\nabla\delta h = \nabla\cdot\frac{\partial \mathbf{K}_{eff}(\alpha_0)}{\partial\alpha}\,\delta\alpha\nabla h_0 \qquad \text{(B4)}$$

$$\mathbf{x}\in D,\ 0 < t \le T$$

$$\delta h(\mathbf{x}, t) = 0 \qquad \mathbf{x}\in D,\ t = 0$$

$$\delta h(\mathbf{x}, t) = 0 \qquad \mathbf{x}\in \partial D_d,\ 0 \le t \le T$$

$$-\mathbf{K}_{eff}(\alpha_0)(\mathbf{x})\nabla\delta h(\mathbf{x}, t)\cdot\mathbf{n} = -\frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial\alpha}\,\delta\alpha(\mathbf{x})\nabla h_0(\mathbf{x}, t)$$

$$x\in \partial D_n,\ 0 \le t \le T$$

where $h_0(\mathbf{x}, t)$ is the solution to (35) when $\alpha(\mathbf{x}) = \alpha_0(\mathbf{x})$ and $\partial\mathbf{K}_{eff}(\alpha_0)/\partial\alpha$ is the derivative of the effective conductivity tensor taken with respect to α and evaluated at α_0. Equation (B4) may be confirmed from a first-order perturbation analysis of (35).

The solution to (B4) may be expressed in terms of the flow equation Green's function evaluated at $\alpha_0(\mathbf{x})$:

$$\delta h(\mathbf{x}, t) = \int_0^T\int_D G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha_0)\cdot\left[\nabla\cdot\delta\alpha(\boldsymbol{\xi})\,\frac{\partial \mathbf{K}_{eff}(\alpha_0)(\boldsymbol{\xi})}{\partial\alpha}\,\nabla h_0(\boldsymbol{\xi}, \tau)\right]d\boldsymbol{\xi}\,d\tau$$

$$= -\int_D \frac{\partial \mathbf{K}_{eff}(\alpha_0)(\boldsymbol{\xi})}{\partial\alpha}\int_0^T \nabla G(\mathbf{x}, \boldsymbol{\xi}, t, \tau|\alpha_0)\cdot\nabla h_0(\boldsymbol{\xi}, \tau)\delta\alpha(\boldsymbol{\xi})\,d\tau\,d\boldsymbol{\xi} \qquad \text{(B5)}$$

The final equality in (B5) is obtained by applying Green's theorem and neglecting the effect of $\alpha(\mathbf{x})$ fluctuations on the specified flux boundary. This approximation simplifies our presentation but is not essential.

If we compare (B5) with (B3) it is apparent that component i of the head portion of the measurement Jacobian vector is given by

$$g_{\mathcal{F}_{hi}}(\mathbf{x}|\alpha_0) = -\frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial\alpha}\int_0^T \nabla\zeta_{i0}(\mathbf{x}, \tau)\cdot\nabla h_0(\mathbf{x}, \tau)\,d\tau \qquad \text{(B6)}$$

where the adjoint variable $\zeta_{i0}(\mathbf{x}, t) = G(\mathbf{x}_i, \mathbf{x}, t_i, \tau|\alpha_0)$ is the flow equation Green's function written as a function of only $\mathbf{x}$ and t for measurement i. The Green's function/adjoint variable is defined to be the function which satisfies the following "adjoint flow equation" [*Courant and Hilbert*, 1953; *Greenberg*, 1971]:

$$-S\frac{\partial\zeta_{i0}(\mathbf{x}, t)}{\partial t} - \nabla\cdot\mathbf{K}_{eff}(\alpha_0)\nabla\zeta_{i0}(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_i, t - t_i) \qquad \text{(B7)}$$

$$\mathbf{x}\in D,\ 0 < t \le T$$

with homogeneous terminal and boundary conditions:

$$\zeta_{i0}(\mathbf{x}, t) = 0 \qquad \mathbf{x}\in D,\ t = T$$

$$\zeta_{i0}(\mathbf{x}, t) = 0 \qquad \mathbf{x}\in \partial D_d,\ 0 \le t \le T$$

$$-\mathbf{K}_{eff}(\alpha_0)(\mathbf{x})\nabla\zeta_{i0}(\mathbf{x}, t)\cdot\mathbf{n} = 0 \qquad \mathbf{x}\in \partial D_n,\ 0 \le t \le T$$

where $\delta(\mathbf{x} - \mathbf{x}_i, t - t_i)$ is a space-time Dirac delta function.

In some specialized situations it is possible to derive the measurement Jacobian in closed form. For example, if we assume that (1) steady state conditions apply (i.e., $\partial h/\partial t = 0$), (2) the effective conductivity is given by the geometric mean of the large-scale conductivity so that $[\mathbf{K}_{eff}(\alpha_0)]_{ij} = e^{\alpha_0}\delta_{ij}$, (3) α_0 is a constant, and (4) the effect of boundary conditions is accounted for by the known head gradient $\nabla h_0(\mathbf{x})$, then (B4) simplifies to

$$-\nabla^2\delta h = \nabla\cdot\delta\alpha\nabla h_0 \qquad \mathbf{x}\in D,\ 0 < t \le T \qquad \text{(B8)}$$

Consequently, (B5) becomes

$$\delta h(\mathbf{x}, t) = -\int_D \nabla G_{\nabla^2}(\mathbf{x}, \boldsymbol{\xi})\cdot\nabla h_0(\boldsymbol{\xi})\delta\alpha(\boldsymbol{\xi})\,d\boldsymbol{\xi} \qquad \text{(B9)}$$

where $G_{\nabla^2}(\mathbf{x}, \boldsymbol{\xi})$ is the well-known infinite domain (or "free space") Green's function associated with the Laplacian operator ∇^2 (see (3g) of *Dagan* [1985]). The adjoint variable may be identified by comparing (B5) and (B9):

$$\zeta_{i0}(\mathbf{x}) = G(\mathbf{x}_i, \mathbf{x}|\alpha_0) = \exp(-\alpha_0)G_{\nabla^2}(\mathbf{x}_i, \boldsymbol{\xi}) \qquad \text{(B10)}$$

So $g_{\mathcal{F}_{hi}}$ is given by

$$g_{\mathcal{F}_{hi}}(\boldsymbol{\xi}|\alpha_0) = -\nabla G_{\nabla^2}(\mathbf{x}_i, \boldsymbol{\xi})\cdot\nabla h_0(\boldsymbol{\xi}) \qquad \text{(B11)}$$

This Jacobian function is used in the linear inverse procedure proposed by *Dagan* [1985] (see section 4). Unfortunately, (B11) cannot be used in a nonlinear Gauss-Newton search because the nominal function $\alpha_0(\mathbf{x})$ is not constant after the first iteration of the search. In nonlinear applications the Jacobian must be evaluated numerically (see below).

The functional gradient $g_{\mathcal{J}}$ provides information about the sensitivity of the performance index to changes in the large-scale conductivity. It is related to the measurement Jacobian through (59):

$$g_{\mathcal{J}}(\mathbf{x}|\alpha_0) = -2\mathbf{g}_{\mathcal{F}}^T(\mathbf{x}|\alpha_0)\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\alpha_0)] + 2\int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})[\alpha_0(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\,d\boldsymbol{\xi} \qquad \text{(B12)}$$

If we distinguish the head and log conductivity parts of the measurement residual term, substitute (B2) and (B6) in this expression, and assume that the measurement errors $\mathbf{v}_h$ and $\mathbf{v}_\alpha$ are independent (in order to simplify notation), the result is

$$g_{\mathcal{J}}(\mathbf{x}|\alpha_0) = 2 \sum_{i=1}^{M_h} \sum_{j=1}^{M_h} [z_{hi} - h_0(\mathbf{x}_i, t_i)][\mathbf{C}_{v_h}^{-1}]_{ij} \frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial \alpha} \cdot \int_0^T [\nabla \zeta_{j0}(\mathbf{x}, t) \cdot \nabla h_0(\mathbf{x}, t)]\, dt - 2 \sum_{i=M_h+1}^{M_h+M_\alpha} \sum_{j=M_h+1}^{M_h+M_\alpha} [z_{\alpha i} - \alpha_0(\mathbf{x}_i)][\mathbf{C}_{v_\alpha}^{-1}]_{ij} \delta(\mathbf{x} - \mathbf{x}_j) + 2 \int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})[\alpha_0(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\, d\boldsymbol{\xi} \tag{B13}$$

This equation expresses $g_{\mathcal{J}}$ in terms of a weighted sum of Green's function gradients (the $\nabla \zeta_{j0}$'s) evaluated at all the head measurement points. But the individual Green's functions do not need to be evaluated in order to compute $g_{\mathcal{J}}$. Suppose that we define a new adjoint variable $\eta_0(\mathbf{x}, t)$ which is a weighted sum of the ζ_{j0}'s:

$$\eta_0(\mathbf{x}, t) = \sum_{i=1}^{M_h} \sum_{j=1}^{M_h} [z_{hi} - h_0(\mathbf{x}_i, t_i)][\mathbf{C}_{v_h}^{-1}]_{ij} \zeta_{j0}(\mathbf{x}, t) \tag{B14}$$

Since the adjoint flow equation (B7) is linear in ζ_{j0}, the superposition principle applies, and $\eta_0(\mathbf{x}, t)$ satisfies a related adjoint equation which is identical to (B7) except for the right-hand side. This η_0 adjoint equation is

$$-S \frac{\partial \eta_0}{\partial t} - \nabla \cdot \mathbf{K}_{eff}(\alpha_0) \nabla \eta_0 = \sum_{i=1}^{M_h} \sum_{j=1}^{M_h} [z_{hi} - h_0(\mathbf{x}_i, t_i)] \cdot [\mathbf{C}_{v_h}^{-1}]_{ij} \delta(\mathbf{x} - \mathbf{x}_j, t - t_j) \tag{B15}$$

$$\mathbf{x} \in D,\ 0 < t \le T$$

with homogeneous terminal and boundary conditions:

$$\eta_0(\mathbf{x}, t) = 0 \qquad \mathbf{x} \in D,\ t = T$$

$$\eta_0(\mathbf{x}, t) = 0 \qquad \mathbf{x} \in \partial D_d,\ 0 \le t \le T$$

$$-\mathbf{K}_{eff}(\alpha_0) \nabla \eta_0(\mathbf{x}, t) \cdot \mathbf{n} = 0 \qquad \mathbf{x} \in \partial D_n,\ 0 \le t \le T$$

If (B14) is substituted into (B13) the performance function gradient may be written as

$$g_{\mathcal{J}}(\mathbf{x}|\alpha_0) = 2 \frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial \alpha} \int_0^T [\nabla \eta_0(\mathbf{x}, t) \cdot \nabla h_0(\mathbf{x}, t)]\, dt - 2 \sum_{i=M_h+1}^{M_h+M_\alpha} \sum_{j=M_h+1}^{M_h+M_\alpha} [z_{\alpha i} - \alpha_0(\mathbf{x}_i)][\mathbf{C}_{v_\alpha}^{-1}]_{ij} \delta(\mathbf{x} - \mathbf{x}_j) + 2 \int_D C_\alpha^{-1}(\mathbf{x}, \boldsymbol{\xi})[\alpha_0(\boldsymbol{\xi}) - \bar{\alpha}(\boldsymbol{\xi})]\, d\boldsymbol{\xi} \tag{B16}$$

This expression shows that we need only know η_0, rather than all M_h ζ_{j0}'s, in order to find the performance index gradient. The gradient may be derived from one solution of the nominal flow equation (to give $h_0(\mathbf{x}, t)$) and one solution of the adjoint equation (B15) (to give $\eta_0(\mathbf{x}, t)$).

Both (B7), the ζ_{i0} adjoint equation, and (B15), the η_0 adjoint equation, have the same general form as the original groundwater flow equation, except that they must be integrated backward, rather than forward, in time, and the right-hand side (forcing) terms are different. Consequently, the numerical solution algorithm used to solve the flow equation can also be used, with minor modification, to solve either of the adjoint equations. When the adjoint computation of the measurement Jacobian (B6) or performance index gradient (B16) is carried out, care must be taken to insure that the spatial gradients and integrations required are performed in a manner consistent with the numerical discretization used in the forward/adjoint solution algorithm [*Chavent*, 1991].

This point can be illustrated if we suppose, for example, that the forward and Jacobian adjoint equations are discretized over space with a finite element technique based on the following basis function expansion applied over a common computational grid:

$$h_0(\mathbf{x}, t) = \sum_m \psi_m(\mathbf{x}) h_{m0}(t)$$

$$\zeta_{i0}(\mathbf{x}, t) = \sum_n \psi_n(\mathbf{x}) \zeta_{in0}(t) \tag{B17}$$

where $h_{m0}(t)$ and $\zeta_{in0}(t)$ are the values of the head and adjoint functions at grid nodes m and n, respectively, and $\psi_m(\mathbf{x})$ and $\psi_n(\mathbf{x})$ are the corresponding nodal finite element basis functions. The m and n summations are taken over all nodes in the computational grid. The discrete flow and adjoint solvers compute the $h_{m0}(t)$'s and the $\zeta_{in0}(t)$'s for a given set of finite element basis functions. Component (i, j) of the head portion of the discrete measurement Jacobian can then be computed by substituting (B6) and (B17) into (68):

$$\frac{\partial \mathcal{F}_{hi}(\boldsymbol{\phi}^T \mathbf{a})}{\partial a_j} = \int_D g_{\mathcal{F}_{hi}}(\boldsymbol{\xi}|\alpha) \phi_j(\boldsymbol{\xi})\, d\boldsymbol{\xi} = -\int_D \frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial \alpha} \phi_j(\boldsymbol{\xi}) \cdot \left[\int_0^T \nabla \zeta_{i0}(\mathbf{x}, \tau) \cdot \nabla h_0(\mathbf{x}, \tau)\, d\tau \right] d\boldsymbol{\xi} = -\sum_m \sum_n \left[\int_D \frac{\partial \mathbf{K}_{eff}(\alpha_0)(\mathbf{x})}{\partial \alpha} \phi_j(\boldsymbol{\xi}) \nabla \psi_m(\boldsymbol{\xi}) \cdot \nabla \psi_n(\boldsymbol{\xi})\, d\boldsymbol{\xi} \right] \left[\int_0^T \zeta_{in0}(\tau) h_{m0}(\tau)\, d\tau \right] \tag{B18}$$

where $i = 1, \cdots, M_h$, $j = 1, \cdots, N$. Note that the spatial discretization permits the time and space integrals in (B18) to be separated. The space integral depends only on the effective conductivity function, the parameter basis functions (the ϕ_j's), and the finite element basis functions (the ψ_m's and ψ_n's). If the spatial integration is carried out numerically the gradients of the finite element basis functions are unambiguously defined at the internal Gauss points used in the quadrature scheme,

even if the finite element basis functions are discontinuous at the element boundaries. Equation (B18) defines a consistent procedure for computing the discrete Jacobian from a discretized solution of the ζ_{i0} functional adjoint equation. Similar methods may be used to derive the discrete performance function gradient from a discretized solution of the η_0 functional adjoint equation.

It should be noted that a set of discrete adjoint and Jacobian or gradient equations can be derived by starting with the discrete maximum a posteriori performance index and a discretized version of the groundwater flow equation [*Sun*, 1994]. In this case the derivation is based on algebraic rather than variational principles. *Sun and Yeh* [1992] show that the algebraic adjoint approach produces exactly the same results as the discretized variational approach outlined in (B18). We prefer the variational alternative because it clearly demonstrates the connection between the Green's function and the adjoint variable and because it yields functional estimation algorithms which apply for a wide range of different parameterization techniques. *Chavent* [1991] provides a useful discussion of some of the computational issues which arise in practical applications of adjoint methods. Adjoint-based Gauss-Newton and gradient search algorithms are popular in optimal control, meteorology, and oceanography (see the bibliography by *Courtier et al*. [1993]). Applications in groundwater hydrology are discussed by *Neuman* [1980], *Townley and Wilson* [1985], *Carrera and Neuman* [1986b], and *Sun* [1994], among others.

It is also possible to derive the discrete Jacobian and performance index gradient with straightforward finite difference approximations. In this case element (i, j) of the discrete Jacobian is approximated by

$$\frac{\partial \mathcal{F}_{hi}(\boldsymbol{\phi}^T\mathbf{a}_0)}{\partial a_j} \approx \frac{\mathcal{F}_{hi}[\boldsymbol{\phi}^T(\mathbf{a}_0 + \mathbf{u}_j\Delta a_j)] - \mathcal{F}_{hi}[\boldsymbol{\phi}^T(\mathbf{a}_0)]}{\Delta a_j} \tag{B19}$$

where $i = 1, \cdots, M_h$, $j = 1, \cdots, N$, and $\mathbf{u}_j$ is a vector with a 1 in element j and zeros elsewhere. Here it is understood that the forward operator represents a numerical solution of the original flow equation, evaluated at $(\mathbf{x}_i, t_i)$. Finite difference computation of all N columns of $\partial\mathcal{F}_h/\partial\mathbf{a}$ requires $N + 1$ flow solutions, as compared to $M_h + 1$ flow solutions if the adjoint-based approach is used. Clearly, the finite difference approach for computing the discrete measurement Jacobian is more efficient when $N < M_h$ while the adjoint method is more efficient when $N > M_h$.

In a similar way, element j of the discrete performance index gradient can be derived from the following finite difference approximation:

$$\frac{\partial \mathcal{J}(\boldsymbol{\phi}^T\mathbf{a}_0)}{\partial a_j} \approx \frac{\mathcal{J}[\boldsymbol{\phi}^T(\mathbf{a}_0 + \mathbf{u}_j\Delta a_j)] - \mathcal{J}[\boldsymbol{\phi}^T(\mathbf{a}_0)]}{\Delta a_j} \tag{B20}$$

where $j = 1, \cdots, N$ and it is understood that the performance index $\mathcal{J}(\alpha)$ is evaluated from a numerical solution of the original flow equation. The adjoint approach for computing the performance index gradient is much more efficient than the finite difference approach since it requires only two flow solutions, as compared to $N + 1$ for the finite difference technique.

Appendix C: Maximum Likelihood Estimation with Surrogate Log Conductivity Measurements

Clifton and Neuman [1982] describe a generalized least squares/maximum likelihood inverse procedure which derives surrogate log effective conductivity measurements from a block kriging algorithm. Here we consider a version of the Clifton and Neuman procedure which uses kriging with a known mean and covariance (rather than kriging with a variogram). Our analysis assumes that $\alpha(\mathbf{x})$ is a random function which can be approximated by the finite-dimensional expansion presented in (29). Although Clifton and Neuman adopted the blocked discretization described in (84) the results derived in this appendix apply for any finite-dimensional parameterization.

We assume that the M_α log conductivity measurements used for kriging are point observations located at $\mathbf{x}_1, \cdots, \mathbf{x}_{M_\alpha}$. If the definition of (39) is invoked the log conductivity forward equation for the kriging problem may be written as

$$\mathbf{z}_\alpha = \mathcal{F}_\alpha(\boldsymbol{\phi}^T\mathbf{a}) + \mathbf{v}_\alpha = \boldsymbol{\Phi}\mathbf{a} + \mathbf{v}_\alpha \tag{C1}$$

where $\boldsymbol{\Phi}$ is an M_α-by-N matrix with row i equal to $\boldsymbol{\phi}^T(\mathbf{x}_i)$. The elements of this matrix identify the components of $\mathbf{a}$ associated with each measurement location.

Since the log conductivity portion of the forward operator is linear an optimal (maximum a posteriori or best linear unbiased) estimate of $\mathbf{a}$ may be obtained from the following expression:

$$\mathbf{z}_{Ka} = \hat{\mathbf{a}}_K = \bar{\mathbf{a}} + \mathbf{C}_a\boldsymbol{\Phi}^T[\boldsymbol{\Phi}\mathbf{C}_a\boldsymbol{\Phi}^T + \mathbf{C}_{v\alpha}]^{-1}[\mathbf{z}_\alpha - \boldsymbol{\Phi}\bar{\mathbf{a}}] \tag{C2}$$

where the K subscript indicates that the variables in question are produced by a kriging algorithm. Equation (C2) may be obtained by inserting the definition of $\mathcal{F}_\alpha$ from (C1) into (73). The prior mean $\bar{\mathbf{a}}$ and covariance matrix $\mathbf{C}_a$ of the discrete parameter vector are assumed to be known.

The maximum likelihood portion of the *Clifton and Neuman* [1982] algorithm assembles the N kriged estimates in a vector of uncertain surrogate measurements with its own forward equation (compare to (C1)):

$$\mathbf{z}_{Ka} = \mathcal{F}_{Ka}(\boldsymbol{\phi}^T\mathbf{a}) + \mathbf{v}_{Ka} = \mathbf{a} + \mathbf{v}_{Ka} \tag{C3}$$

where $\mathbf{v}_{Ka}$ is the kriging estimation error. Note that $\mathbf{a}$ is treated as an unknown deterministic parameter in the maximum likelihood approach. Equations (C2) and (C3) may be used to show that $\mathbf{v}_{Ka}$ is zero mean (i.e., the estimate is unbiased) with a covariance equal to [*Schweppe*, 1973, p. 96]

$$\mathbf{C}_{Ka} = [\mathbf{C}_a^{-1} + \boldsymbol{\Phi}^T\mathbf{C}_{v\alpha}^{-1}\boldsymbol{\Phi}]^{-1} \tag{C4}$$

If the surrogate measurements and measurement error covariance obtained from the kriging algorithm are inserted into the second term of the maximum likelihood performance index of (89), the result is

$$\begin{aligned} J_{mle}(\mathbf{a}) &= [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{hv}^{-1}[\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})] \\ &\quad + [\mathbf{z}_{Ka} - \mathcal{F}_{Ka}(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{Ka}^{-1}[\mathbf{z}_{Ka} - \mathcal{F}_{Ka}(\boldsymbol{\phi}^T\mathbf{a})] \\ &= [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{hv}^{-1}[\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})] \\ &\quad + [\mathbf{z}_{Ka} - \mathbf{a}]^T\mathbf{C}_{Ka}^{-1}[\mathbf{z}_{Ka} - \mathbf{a}] \end{aligned} \tag{C5}$$

where (C3) is used to write the forward operator in the kriging error term as $\mathcal{F}_{Ka}(\boldsymbol{\phi}^T\mathbf{a}) = \mathbf{a}$.

Now consider an alternative maximum a posteriori performance index which includes both the log conductivity measurement $\mathbf{z}_\alpha$ and prior information about $\mathbf{a}$, where $\mathbf{a}$ is now treated as a random vector with known prior statistics:

$$J_{map}(\mathbf{a}) = [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{hv}^{-1}[\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})] + [\mathbf{z}_\alpha - \mathcal{F}_\alpha(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{v\alpha}^{-1}[\mathbf{z}_\alpha - \mathcal{F}_\alpha(\boldsymbol{\phi}^T\mathbf{a})] + [\mathbf{a} - \bar{\mathbf{a}}]^T\mathbf{C}_a^{-1}[\mathbf{a} - \bar{\mathbf{a}}] = [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{hv}^{-1} \cdot [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})] + [\mathbf{z}_\alpha - \boldsymbol{\Phi}\mathbf{a}]^T\mathbf{C}_{v\alpha}^{-1}[\mathbf{z}_\alpha - \boldsymbol{\Phi}\mathbf{a}] + [\mathbf{a} - \bar{\mathbf{a}}]^T\mathbf{C}_\alpha^{-1}[\mathbf{a} - \bar{\mathbf{a}}] \tag{C6}$$

Here (C1) is used to write the forward operator in the log conductivity error term as $\mathcal{F}_\alpha(\boldsymbol{\phi}^T\mathbf{a}) = \boldsymbol{\Phi}\mathbf{a}$. This performance index is equal to (C5) to within a additive constant.

The equivalence of (C5) and (C6) may be demonstrated by expanding and regrouping the last two terms in (C6). We begin by noting that these terms constitute a quadratic form in $\mathbf{a}$ which may be written as

$$[\mathbf{z}_\alpha - \boldsymbol{\Phi}\mathbf{a}]^T\mathbf{C}_{v\alpha}^{-1}[\mathbf{z}_\alpha - \boldsymbol{\Phi}\mathbf{a}] + [\mathbf{a} - \bar{\mathbf{a}}]^T\mathbf{C}_a^{-1}[\mathbf{a} - \bar{\mathbf{a}}] = [\mathbf{a} - \mathbf{b}]^T\mathbf{B}^{-1}[\mathbf{a} - \mathbf{b}] + \mathbf{d} = \mathbf{a}^T\mathbf{B}^{-1}\mathbf{a} - 2\mathbf{b}^T\mathbf{B}^{-1}\mathbf{a} + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{b} + \mathbf{d} \tag{C7}$$

where $\mathbf{b}$, $\mathbf{B}$, and $\mathbf{d}$ are unknown coefficients which do not depend on $\mathbf{a}$ and which can be identified by comparing like terms in (C7). In particular, the quadratic coefficient $\mathbf{B}^{-1}$ and the linear coefficient $\mathbf{b}^T\mathbf{B}^{-1}$ can be identified after expanding the left side of (C7):

$$\mathbf{B}^{-1} = \mathbf{C}_a^{-1} + \boldsymbol{\Phi}^T\mathbf{C}_{v\alpha}^{-1}\boldsymbol{\Phi} = \mathbf{C}_{Ka}^{-1} \tag{C8}$$

$$\mathbf{b}^T\mathbf{B}^{-1} = \bar{\mathbf{a}}^T\mathbf{C}_a^{-1} + \mathbf{z}_\alpha^T\mathbf{C}_{v\alpha}^{-1}\boldsymbol{\Phi} \tag{C9}$$

Equation (C8) may be inverted to give an expression for $\mathbf{B}$. Equation (C9) can then be postmultiplied by $\mathbf{B}$ and transposed to give

$$\mathbf{b} = [\mathbf{C}_a^{-1} + \boldsymbol{\Phi}^T\mathbf{C}_{v\alpha}^{-1}\boldsymbol{\Phi}]^{-1}[\mathbf{C}_a^{-1}\bar{\mathbf{a}} + \boldsymbol{\Phi}^T\mathbf{C}_{v\alpha}^{-1}\mathbf{z}_\alpha] = \bar{\mathbf{a}} + \mathbf{C}_a\boldsymbol{\Phi}^T[\boldsymbol{\Phi}\mathbf{C}_a\boldsymbol{\Phi}^T + \mathbf{C}_{v\alpha}]^{-1}[\mathbf{z}_\alpha - \boldsymbol{\Phi}\bar{\mathbf{a}}] = \mathbf{z}_{Ka} \tag{C10}$$

where the second equality follows from the matrix inversion lemma given by *Schweppe* [1973, p. 496]. If we substitute (C8) and (C10) into (C7) and then substitute the result into (C6) we can write the maximum a posteriori performance index as

$$J_{map}(\mathbf{a}) = [\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{hv}^{-1}[\mathbf{z}_h - \mathcal{F}_h(\boldsymbol{\phi}^T\mathbf{a})] + [\mathbf{z}_{Ka} - \mathcal{F}_{Ka}(\boldsymbol{\phi}^T\mathbf{a})]^T\mathbf{C}_{Ka}^{-1}[\mathbf{z}_{Ka} - \mathcal{F}_{K\alpha}(\boldsymbol{\phi}^T\mathbf{a})] + \mathbf{d} \tag{C11}$$

This equation is the same as (C5) except for presence of the constant $\mathbf{d}$, which does not affect the minimization operation since it does not depend on $\mathbf{a}$.

The equivalence of (C5) and (C6) indicates that a maximum likelihood estimation algorithm which uses kriging to produce surrogate measurements of the unknown parameter $\mathbf{a}$ is equivalent to a maximum a posteriori algorithm which provides for prior information in a Bayesian fashion. The surrogate measurements can be viewed as "sufficient statistics" which contain all the information conveyed by the actual point measurements and the prior statistics of $\mathbf{a}$ [*Sorenson*, 1980, pp. 79–86].

Appendix D: Bayesian Generalization of the Cramer-Rao Bound on the Estimation Error Covariance

A Bayesian generalization of the discrete parameter Cramer-Rao bound may be derived by following the general procedure outlined by *Schweppe* [1973, pp. 372–375], but allowing the parameter $\mathbf{a}$ to be random. We begin by defining the estimator bias $\mathbf{b}(\mathbf{a})$ for a given $\mathbf{a}$:

$$\mathbf{b}(\mathbf{a}) = \mathbf{a} - E_{z|a}[\hat{\mathbf{a}}(\mathbf{z})] = \int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}]p_{z|a}(\mathbf{z}|\mathbf{a})\, d\mathbf{z} \tag{D1}$$

where $\hat{\mathbf{a}}(\mathbf{z})$ is the estimate, written as a function of $\mathbf{z}$ only, $E_{z|a}[\ \ ]$ and $p_{z|a}(\mathbf{z}|\mathbf{a})$ are the conditional expectation and probability density of $\mathbf{z}$ given $\mathbf{a}$, and the integral is understood to be taken over each of the N components of $\mathbf{z}$. If both sides of (D1) are differentiated with respect to $\mathbf{a}$ and the terms are rearranged the result may be written as

$$\int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}]\left[\frac{\partial \ln p_{z|a}(\mathbf{z}|\mathbf{a})}{\partial \mathbf{a}}\right]^T p_{z|a}(\mathbf{z}|\mathbf{a})\, d\mathbf{z} = -I + \frac{\partial \mathbf{b}(\mathbf{a})}{\mathbf{a}} \tag{D2}$$

where I is the N-dimensional identity matrix.

We can use Bayes' rule to replace $p_{z|a}(\mathbf{z}|\mathbf{a})$ by $p_{za}(\mathbf{z}, \mathbf{a})/p_a(\mathbf{a})$ in the $\ln p_{z|a}(\mathbf{z}|\mathbf{a})$ derivative and then multiply both sides of (D2) by $p_a(\mathbf{a})$ to obtain

$$\int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}]\left[\frac{\partial \ln p_{za}(\mathbf{z}, \mathbf{a})}{\partial \mathbf{a}}\right]^T p_{za}(\mathbf{z}, \mathbf{a})\, d\mathbf{z} = -p_a(\mathbf{a})I + p_a(\mathbf{a})\frac{\partial \mathbf{b}(\mathbf{a})}{\partial \mathbf{a}} + \int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}] \cdot \left[\frac{\partial \ln p_a(\mathbf{a})}{\partial \mathbf{a}}\right]^T p_{z|a}(\mathbf{z}|\mathbf{a})p_a(\mathbf{a})\, d\mathbf{z} \tag{D3}$$

The integral on the right side of this equation can be simplified by writing it as an expectation and invoking the bias definition of (D1):

$$\int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}]\left[\frac{\partial \ln p_a(\mathbf{a})}{\partial \mathbf{a}}\right]^T p_{z|a}(\mathbf{z}|\mathbf{a})p_a(\mathbf{a})\, d\mathbf{z} = \{\mathbf{a} - E_{z|a}[\hat{\mathbf{a}}(\mathbf{z})]\}\left[\frac{\partial \ln p_a(\mathbf{a})}{\partial \mathbf{a}}\right]^T p_a(\mathbf{a}) = \mathbf{b}(\mathbf{a})\left[\frac{\partial p_a(\mathbf{a})}{\partial \mathbf{a}}\right]^T \tag{D4}$$

When (D4) is substituted into (D3) and the two derivative terms on the right side containing $\mathbf{b}(\mathbf{a})$ are combined the result is

$$\int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}]\left[\frac{\partial \ln p_{za}(\mathbf{z}, \mathbf{a})}{\partial \mathbf{a}}\right]^T p_{za}(\mathbf{z}, \mathbf{a})\, d\mathbf{z} = -p_a(\mathbf{a})I + \frac{\partial}{\partial \mathbf{a}}[\mathbf{b}(\mathbf{a})p_a(\mathbf{a})] \tag{D5}$$

If we integrate both sides of (D5) over $\mathbf{a}$, the last (derivative) term on the right side vanishes, provided that $\mathbf{b}(\mathbf{a})$ is finite when any of the components of $\mathbf{a}$ is infinite. The result is a Bayesian generalization of *Schweppe*'s equation (12.3.5) [1973, p. 374]:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [\mathbf{a} - \hat{\mathbf{a}}]\left[\frac{\partial \ln p_{za}(\mathbf{z}, \mathbf{a})}{\partial \mathbf{a}}\right]^T p_{za}(\mathbf{z}, \mathbf{a})\, d\mathbf{z}\, d\mathbf{a} = -I \tag{D6}$$

Schweppe [1973, pp. 374–375] shows how Schwartz's inequality can be applied to (D6) to give the following lower bound on the mean-squared estimation error:

$$E_{za}\{\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T\} = E_z\{E_{a|z}[\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T]\} \geq \mathbf{B}^{-1} \tag{D7}$$

where $\tilde{\mathbf{a}} = \mathbf{a} - \hat{\mathbf{a}}$ and the "information matrix" $\mathbf{B}$ is defined as

$$\mathbf{B} = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{\partial}{\partial \mathbf{a}}\frac{\partial}{\partial \mathbf{a}}\ln p_{za}(\mathbf{z}, \mathbf{a})\right]p_{za}(\mathbf{z}, \mathbf{a})\, d\mathbf{z}\, d\mathbf{a}$$

$$= -E_{za}\left\{\frac{\partial}{\partial \mathbf{a}}\frac{\partial}{\partial \mathbf{a}}\ln p_{za}(\mathbf{z}, \mathbf{a})\right\} \tag{D8}$$

The inequality in (D7) is interpreted as specifying that the difference between the matrices on the left and right sides is positive semidefinite. If the bias is sufficiently small, the mean-squared error is equal to the estimation error covariance and the bound of (D7) can be written:

$$E_{za}\{\tilde{\mathbf{a}}\tilde{\mathbf{a}}^T\} \approx E_{za}\{[\tilde{\mathbf{a}} - E_{za}(\tilde{\mathbf{a}})][\tilde{\mathbf{a}} - E_{za}(\tilde{\mathbf{a}})]^T\} = \mathbf{C}_{\tilde{a}} \geq \mathbf{B}^{-1} \tag{D9}$$

This approximation improves as the number of measurements (M) increases if the maximum a posteriori estimator is used since the bias of this estimator asymptotically approaches 0 [*Bard*, 1974, p. 73]. Equation (D9) is the Bayesian version of the Cramer-Rao bound on the estimation error covariance.

If the parameter and measurement error are independent Gaussian vectors, the log of $p_{za}(\mathbf{z}, \mathbf{a})$ becomes

$$\begin{aligned}\ln p_z\mathbf{a}(\mathbf{z}, \mathbf{a}) &= \ln\left[p_{z|a}(\mathbf{z}|\mathbf{a})p_a(\mathbf{a})\right] \\ &= \ln p_v[\mathbf{z} - \mathcal{F}(\mathbf{\Phi}^T\mathbf{a})] + \ln p_a[a] \\ &= \gamma - \tfrac{1}{2}[\mathbf{z} - \mathcal{F}(\mathbf{\Phi}^T\mathbf{a})]^T\mathbf{C}_v^{-1}[\mathbf{z} - \mathcal{F}(\mathbf{\Phi}^T\mathbf{a})] \\ &\quad - \tfrac{1}{2}[\mathbf{a} - \bar{\mathbf{a}}]^T\mathbf{C}_a^{-1}[\mathbf{a} - \bar{\mathbf{a}}]\end{aligned} \tag{D10}$$

where γ is a constant and we have used Bayes' rule and the forward equation (42) to express $\ln p_{za}(\mathbf{z}, \mathbf{a})$ in terms of the known probability densities of $\mathbf{v}$ and $\mathbf{a}$. If we approximate the nonlinear forward operator $\mathcal{F}(\mathbf{\Phi}^T\mathbf{a})$ by a first-order Taylor series about a specified nominal parameter value $\mathbf{a}_0$ (see (62)) the second derivative of (D10) becomes

$$\frac{\partial}{\partial \mathbf{a}}\frac{\partial}{\partial \mathbf{a}}\ln p_{za}(\mathbf{z}, \mathbf{a}) \approx \frac{\partial \mathcal{F}^T(\mathbf{a}_0)}{\partial \mathbf{a}}\mathbf{C}_v^{-1}\frac{\partial \mathcal{F}(\mathbf{a}_0)}{\partial \mathbf{a}} + \mathbf{C}_a^{-1} \tag{D11}$$

Note that the second derivative of $\ln p_{za}(\mathbf{z}, \mathbf{a})$ is deterministic (does not depend on either $\mathbf{a}$ or $\mathbf{z}$) when a linearized Gaussian approximation is adopted. In this case the lower bound on the estimation error covariance becomes

$$\mathbf{C}_{\tilde{a}} \geq \mathbf{B}^{-1} = \left[\frac{\partial \mathcal{F}^T(\mathbf{a}_0)}{\partial \mathbf{a}}\mathbf{C}_v^{-1}\frac{\partial \mathcal{F}(\mathbf{a}_0)}{\partial \mathbf{a}} + \mathbf{C}_a^{-1}\right]^{-1} \tag{D12}$$

In practice, the nominal parameter value $\mathbf{a}_0$ is usually set equal to the best available estimate.

Acknowledgments. We appreciate the valuable comments and suggestions provided by Ghislain de Marsily, Lynn Gelhar, Lynn B. Reid, Mark Matthews, Mark Lukas, Claude Dietrich, and two anonymous reviewers. The first author would also like to thank the CSIRO Division of Water Resources (Australia), the TNO Institute of Applied Geological Sciences (Netherlands), and the U.S. National Science Foundation (grants EAR-9218602 and BES-9222116) for their support.

References

Ababou, R., D. McLaughlin, L. W. Gelhar, and A. F. B. Tompson, Numerical simulation of three dimensional saturated flow in randomly heterogeneous porous media, *Transp. Porous Media*, *4*(6), 549–565, 1989.

Anderson, B., and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, N. J., 1979.

Anger, G., *Inverse Problems in Differential Equations*, Plenum, New York, 1990.

Banks, H. T., and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhauser, Boston, 1989.

Bard, Y., *Nonlinear Parameter Estimation*, Academic, San Diego, Calif., 1974.

Bear, J., *Hydraulics of Groundwater*, McGraw-Hill, New York, 1979.

Beckie, R., A. A. Aldama, and E. F. Wood, The universal structure of the groundwater flow equations, *Water Resour. Res.*, *30*(5), 1407–1420, 1994.

Bencala, K. E., and J. H. Seinfeld, Distributed parameter filtering: Boundary noise and discrete observations, *Int. J. Syst. Sci.*, *10*, 735–749, 1979.

Bennett, A. F., *Inverse Methods in Physical Oceanography*, Cambridge Univ. Press, New York, 1992.

Braud, I., C. Obled, and A. Phamdinhtuan, Empirical orthogonal function (eof) analysis of spatial random fields: Theory, accuracy of the numerical approximations and sampling effects, *Stochastic Hydrol. Hydraul.*, *7*(2), 146–160, 1993.

Bryson, A., and Y. Ho, *Applied Optimal Control*, Ginn/Blaisdell, Waltham, Mass., 1975.

Carrera, J., State of the art of the inverse problem applied to the flow and solute transport problems, in *Groundwater Flow and Quality Modeling*, *NATO ASI Ser.*, pp. 549–585, 1987.

Carrera, J., and L. Glorioso, On geostatistical formulations of the groundwater flow inverse problem, *Adv. Water Resour.*, *14*(5), 273–283, 1991.

Carrera, J., and A. Medina, An improved form of adjoint-state equations for transient problems, in *Computational Methods in Water Resources X*, pp. 199–206, Kluwer Acad., Norwell, Mass., 1994.

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, 1986a.

Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 2, Uniqueness, stability, and solution algorithms, *Water Resour. Res.*, *22*(2), 211–227, 1986b.

Carrera, J., and S. P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 3, Application to synthetic and field data, *Water Resour. Res.*, *22*(2), 228–242, 1986c.

Certes, C., and G. de Marsily, Application of the pilot point method to the identification of aquifer transmissivities, *Adv. Water Resour.*, *14*(5), 284–299, 1991.

Chavent, G., On the theory and practice of non-linear least-squares, *Adv. Water Resour.*, *14*(2), 55–63, 1991.

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., A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 1, Theory and numerical properties, *Water Resour. Res.*, *13*(2), 318–324, 1977.

Cooley, R. L., A method of estimating parameters and assessing reliability for models of steady state groundwater flow, 2, Application of statistical analysis, *Water Resour. Res.*, *15*(3), 603–617, 1979.

Cooley, R. L., Incorporation of prior information on parameters into nonlinear regression groundwater flow models, 1, Theory, *Water Resour. Res.*, *18*(4), 965–976, 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.

Courant, R., and D. Hilbert, *Methods of Mathematical Physics*, Wiley-Intersci., New York, 1953.

Courtier, J. P., J. Dreber, R. Errico, J.-F. Louis, and T. Vukicevic, Important literature on the use of adjoint, variational methods and the kalman filter in meteorology, *Tellus*, *45A*, 342–357, 1993.

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

Daley, R., *Atmospheric Data Analysis*, Cambridge Univ. Press, New York, 1991.

Delhomme, J. P., Kriging in the hydrosciences, *Adv. Water Resour.*, *1*(5), 251–266, 1978.

Dietrich, C. R., and G. N. Newsam, A stability analysis of the geostatistical approach to aquifer identification, *Stochastic Hydrogeol. Hydraul.*, *3*(4), 293–316, 1989.

Dietrich, C. R., and G. N. Newsam, Sufficient conditions for identifying transmissivity in a confined aquifer, *Inverse Probl.*, *6*, L21–L28, 1990.

Durlovsky, L., Numerical calculations of equivalent grid block permeability tensors for heterogeneous porous media, *Water Resour. Res.*, *27*(5), 699–708, 1991.

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

Fitzpatrick, B. G., Bayesian analysis in inverse problems, *Inverse Probl.*, *7*, 675–702, 1991.

Freeze, R. A., A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media, *Water Resour. Res.*, *11*(5), 725–741, 1975.

Gavalas, G. R., P. C. Shah, and J. H. Seinfeld, Reservoir history matching by Bayesian estimation, *Trans. Soc. Pet. Eng.*, *261*, 337–350, 1976.

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

Gelhar, L. W., and C. L. Axness, Three-dimensional stochastic analysis of macrodispersion in a stratified aquifer, *Water Resour. Res.*, *19*(1), 161–180, 1983.

Gill, P. E., W. Murray, and M. H. Wright, *Practical Optimisation*, Academic, San Diego, Calif., 1981.

Ginn, T. R., and J. H. Cushman, Inverse methods for subsurface flow: A critical review of stochastic techniques, *Stochastic Hydrol. Hydraul.*, *4*, 1–26, 1990.

Graham, W. D., and D. McLaughlin, Stochastic analysis of nonstationary subsurface solute transport, 2, Conditional moments, *Water Resour. Res.*, *25*(11), 2331–2355, 1989.

Graham, W. D., and D. McLaughlin, A stochastic model of solute transport in groundwater: Application to the Borden, Ontario, tracer test, *Water Resour. Res.*, *27*(6), 1345–1359, 1991.

Greenberg, M. D., *Application of Green's Functions in Science and Engineering*, Prentice-Hall, Englewood Cliffs, N. J., 1971.

Gutjahr, A., and J. L. Wilson, Co-kriging and stochastic differential equations, in *The Stochastic Approach to Subsurface Flow*, pp. 2–22, Ecole des Mines de Paris, Montvillargenne, 1985.

Hadamard, J., *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Dover, Mineola, N. Y., 1952.

Han, D., A. Nur, and D. Morgan, Effect of porosity and clay content on wave velocity in sandstones, *Geophysics*, *51*, 2093–2107, 1986.

Hill, M. C., A computer program (MODFLOWP) for estimating parameters of a transient, three-dimensional groundwater flow model using nonlinear regression, *U.S. Geol. Surv. Open-File Rep. 91-484*, 1991.

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, Analysis of the spatial structure of properties of selected aquifers, *Water Resour. Res.*, *21*(4), 563–572, 1985.

Hyndman, D. W., J. M. Harris, and S. M. Gorelick, Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, *30*(7), 1965–1978, 1994.

Indelman, P., and G. Dagan, Upscaling of permeabilities of anisotropic heterogeneous formations, 1, The general framework, *Water Resour. Res.*, *29*(4), 917–924, 1993.

Jacquard, P., and C. Jain, Permeability distribution from field pressure data, *Trans. Soc. Pet. Eng.*, *5*(4), 281–294, 1965.

Jahns, H. O., A rapid method for obtaining a two-dimensional reservoir description from well pressure response data, *Trans. Soc. Pet. Eng.*, *237*, 315–327, 1966.

Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic, San Diego, Calif., 1970.

Keidser, A., and D. Rosbjerg, A comparison of four inverse approaches to groundwater flow and transport parameter identification, *Water Resour. Res.*, *27*(9), 2219–2232, 1991.

Keidser, A., D. Rosbjerg, K. Hogh Jensen, and K. Bitsch, A joint kriging and zonation approach to inverse groundwater modelling, in *ModelCARE90: Calibration and Reliability in Groundwater Modelling*, *IAHS Publ. 195*, 171–183, 1990.

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*(13), 677–690, 1983.

Kleinecke, D., The use of linear programming for estimating geohydrologic parameters of groundwater basins, *Water Resour. Res.*, *7*(2), 367–374, 1971.

Kubrusly, C. S., Distributed parameter system identification a survey, *Int. J. Control*, *26*(4), 509–535, 1977.

Kuelbs, J., F. M. Larkin, and J. A. Williamson, Weak point distributions on reproducing kernel Hilbert spaces, *Rocky Mt. J. Math.*, *2*(3), 369–378, 1972.

Kuiper, L., A comparison of several methods for the solution of the inverse problem in two-dimensional steady state groundwater flow modeling, *Water Resour. Res.*, *22*(5), 705–714, 1986.

Kuo, H. H., *Gaussian Measures in Banach Spaces*, Springer-Verlag, New York, 1975.

Larkin, F. M., Gaussian measure in Hilbert space and applications in numerical analysis, *Rocky Mt. J. Math.*, *2*(3), 379–421, 1972.

LaVenue, A. M., and J. F. Pickens, Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model, *Water Resour. Res.*, *28*(6), 1543–1570, 1992.

Liu, J., A multiresolution method for distributed parameter estimation, *SIAM J. Sci. Comput.*, *14*(2), 389–405, 1993.

Ljung, L., Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems, *IEEE Trans. Autom. Control*, *24*, 36–50, 1979.

Lorenc, A. C., Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*, 1177–1194, 1986.

Luis, S., and D. McLaughlin, A stochastic approach to model validation, *Adv. Water Resour.*, *15*, 15–32, 1992.

de Marsily, G., Spatial variability of properties in porous media: A stochastic approach, in *Fundamentals of Transport in Porous Media*, edited by J. Bear and M. Y. Corapcioglu, pp. 719–769, Martinus Nijhoff, Boston, 1984.

de Marsily, G., *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*, Academic, San Diego, Calif., 1986.

de Marsily, G., G. Lavedan, M. Bocuher, and G. Fasanino, Interpretation of interference tests in a well filed using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for Natural Resources Characterization*, edited by G. Verly, M. David, A. G. Journel, and A. Marechal, pp. 831–849, D. Reidel, Norwell, Mass., 1984.

McLaughlin, D., Investigation of alternative procedures for estimating ground-water basin parameters, Water Resour. Eng., Walnut Creek, Calif., 1975.

McLaughlin, D., Hanford groundwater modeling—a numerical comparison of Bayesian and Fisher parameter estimation techniques, *Consult. Rep. RHO-C-24*, Rockwell Hanford Oper., Richland, Wash., 1979.

McLaughlin, D., Recent advances in hydrologic data assimilation, *U.S. Natl. Rep. Int. Union Geod. Geophys. 1991–1994*, *Rev. Geophys.*, *33*, 977–984, 1995.

McLaughlin, D., and E. Wood, A distributed parameter approach for evaluating the accuracy of groundwater model predictions, 1, Theory, *Water Resour. Res.*, *24*(7), 1037–1047, 1988a.

McLaughlin, D., and E. Wood, A distributed parameter approach for evaluating the accuracy of groundwater model predictions, 2, Application to groundwater flow, *Water Resour. Res.*, *24*(7), 1048–1060, 1988b.

McLaughlin, D., W. Kinzelbach, and F. Ghassemi, Modelling subsurface flow and transport: Where do we stand?, *Hydrogeologie*, *4*, 269–280, 1993.

Nelson, R. W., In-place measurement of permeability in heterogeneous media, 1, Theory of a proposed method, *J. Geophys. Res.*, *65*(6), 1753–1758, 1960.

Nelson, R. W., In-place measurement of permeability in heterogeneous media, 2, Experimental and computational considerations, *J. Geophys. Res.*, *66*(8), 2469–2478, 1961.

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

Neuman, S. P., A statistical approach to the inverse problem of aquifer hydrology, 3, Improved solution method and added perspective, *Water Resour. Res.*, *16*(2), 331–346, 1980.

Neuman, S. P., and S. Yakowitz, A statistical approach to the inverse problem of aquifer hydrology, 1, Theory, *Water Resour. Res*., *15*(4), 845–860, 1979.

Neuman, S. P., G. E. Fogg, and E. A. Jacobson, A statistical approach to the inverse problem of aquifer hydrology, 2, Case study, *Water Resour. Res*., *16*(1), 33–58, 1980.

Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1984.

Polis, M. P., The distributed system parameter identification problem: A survey of recent results, in *IFAC 3rd Symposium*, *Control of Dist. Perm. Systems*, pages 45–58, Toulouse, France, 1982.

Rajaram, H., and D. McLaughlin, Identification of large-scale spatial trends in hydrologic data, *Water Resour. Res*., *27*(10), 2411–2423, 1991.

Reid, L. B., A functional inverse approach for three-dimensional characterization of subsurface contamination, Ph.D. thesis, Dep. of Civ. and Environ. Eng., Mass. Inst. of Technol., Cambridge, 1996.

Reid, L. B., and D. McLaughlin, Estimating continuous aquifer properties from field measurements: The inverse problem for groundwater flow and transport, in *Computational Methods in Water Resources X*, pp. 777–784, Kluwer Acad., Norwell, Mass., 1994.

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.

Rubin, Y., and J. J. Gomez-Hernandez, A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations, *Water Resour. Res*., *25*(4), 405–419, 1989.

Russo, D., and W. A. Jury, A theoretical study of the estimation of the correlation scale in spatially variable fields, 1, Stationary fields, *Water Resour. Res*., *23*(7), 1257–1268, 1987a.

Russo, D., and W. A. Jury, A theoretical study of the estimation of the correlation scale in spatially variable fields, 2, Nonstationary fields, *Water Resour. Res*., *23*(7), 1269–1279, 1987b.

Samper, F. J., and S. P. Neuman, Estimation of spatial covariance structures by adjoint state maximum likelihood cross validation, 1, Theory, *Water Resour. Res*., *25*(3), 351–362, 1989.

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

Scott, D. M., An evaluation of flow net analysis for aquifer identification, *Groundwater*, *30*(5), 755–764, 1992.

Shah, P. C., G. R. Gavalas, and J. H. Seinfeld, Error analysis in history matching: The optimum level of parameterization, *Trans. Soc. Pet. Eng*., *263*, 219–228, 1978.

Sorenson, H. W., *Parameter Estimation, Principles and Problems*, Marcel Dekker, New York, 1980.

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

Sun, N.-Z., and W. W. G. Yeh, A stochastic inverse solution for transient groundwater flow: Parameter identification and reliability analysis, *Water Resour. Res*., *28*(12), 3269–3280, 1992.

Tarantola, A., *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier, New York, 1987.

te Stroet, C. B. M., Calibration of stochastic groundwater models, Ph.D. thesis, Tech. Univ. of Delft, Delft, Netherlands, 1995.

Tikhonov, A. N., and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Halsted Press/Wiley, New York, 1977.

Townley, L. R., Numerical models of groundwater flow: Predictions and parameter estimation in the presence of uncertainty, Ph.D. thesis, Mass. Inst. of Technol., Cambridge, 1983.

Townley, L. R., and J. L. Wilson, Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow, *Water Resour. Res*., *21*(12), 1851–1860, 1985.

Townley, L. R., and J. L. Wilson, Estimation of boundary values and identification of boundary type for numerical models of aquifer flow, *Transp. Porous Media*, *4*, 567–584, 1989.

Vakhania, N. N., *Probability Distributions on Linear Spaces*, North-Holland, New York, 1981.

Wahba, G., Variational methods for multidimensional inverse problems, in *Advances in Remote Sensing Retrieval Methods*, edited by A. Deepak, H. E. Fleming, and M. T. Chahine, pp. 385–410, A. Deepak, Hampton, Va., 1985.

Wahba, G., Spline models for observational data, Soc. for Ind. and Appl. Math., Philadelphia, Pa., 1990.

Wilson, J. L., P. Kitanidis, and M. Dettinger, State and parameter estimation in groundwater models, in *Applications of Kalman Filter to Hydrology, Hydraulics, and Water Resources*, edited by C.-L. Chiu, pp. 657–680, Stochastic Hydraul. Program, Dep. of Civ. Eng., Univ. of Pittsburgh, Pittsburgh, Pa., 1978.

Yakowitz, S., and L. Duckstein, Instability in aquifer identification: Theory and case study, *Water Resour. Res*., *16*(6), 1045–1064, 1980.

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

Zou, X., I. M. Navon, M. Berger, K. H. Phua, T. Chlick, and F. X. Le Diment, Numerical experience with limited memory quasi Newton and truncated Newton methods, *SIAM J. Opt*., *3*(3), 582–608, 1993.

D. McLaughlin, Parsons Laboratory, Massachusetts Institute of Technology, Rm. 48-209, Cambridge, MA 02139. (e-mail: dennism@mit.edu)

L. R. Townley, CSIRO Division of Water Resources, Private Bag P.O. Wembley WA 6014, Australia. (e-mail: lloyd@per.dwr.csiro.au)

(Received January 31, 1995; revised December 5, 1995; accepted January 4, 1996.)