A full-Bayesian approach to the groundwater inverse problem for steady state flow

Allan D. Woodbury
Department of Civil Engineering, University of Manitoba, Winnipeg, Manitoba, Canada

Tadeusz J. Ulrych
Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Abstract. A full-Bayesian approach to the estimation of transmissivity from hydraulic head and transmissivity measurements is developed for two-dimensional steady state groundwater flow. The approach combines both Bayesian and maximum entropy viewpoints of probability. In the first phase, log transmissivity measurements are incorporated into Bayes’ theorem, and the prior probability density function is updated, yielding posterior estimates of the mean value of the log transmissivity field and covariance. The two central moments are generated assuming that the prior mean, variance, and integral scales are “hyperparameters”; that is, they are treated as random variables in themselves which is contrary to classical statistical approaches. The probability density functions (pdfs) of these hyperparameters are, in turn, determined from maximum entropy considerations. In other words, pdfs are chosen for each of the hyperparameters that are maximally uncommitted with respect to unknown information. This methodology is quite general and provides an alternative to kriging for spatial interpolation. The final step consists of updating the conditioned natural logarithm transmissivity (ln(T)) field with hydraulic head measurements, utilizing a linearized aquifer equation. It is assumed that the statistical properties of the noise in the hydraulic head measurements are also uncertain. At each step, uncertainties in all pertinent hyperparameters are removed by marginalization. Finally, what is produced is a ln(T) field conditioned on measurements of both hydraulic heads and log transmissivity and covariances of the ln(T) field. In addition, we can also produce resolution matrices, confidence (credibility) limits, and the like for the ln(T) field. It is shown that the application of the methodology yields good estimates of transmissivities, even when hydraulic head measurements are noisy and little or no information is specified on mean values of ln(T), variance of ln(T), and integral scales.

1. Introduction

The challenge of hydrogeological site characterization has attracted a remarkable research effort, particularly over the last 20 years (see reviews by Woodbury [1987], Willis and Yeh [1987], Ginn and Cushman [1990], McLaughlin and Townley [1996, 1997], and Kitamidis [1995, 1997]). These references are by no means exhaustive and serve only to indicate the importance of the overall problem of site characterization. The definition of the problem is quite simple: A three-dimensional “picture” of the heterogeneous subsurface is needed to identify the key controls on the flow processes and contaminant transport. This implies defining the hydraulic parameters such as hydraulic conductivity, storativity, porosity, and the like at each and every point in the aquifer. It is now widely recognized [cf. Gelhar, 1986; Dagan, 1986] that natural heterogeneity and the large spatial variability of the hydraulic conductivity predominantly controls the flow field and hence the spread of contaminants. Aquifers, as a rule, are highly heterogeneous, and hydraulic properties can vary significantly over very short distances. As a result, there may not be enough point measurements of hydraulic parameters for a comprehensive, detailed characterization.

A very interesting example of parametric uncertainty was that summarized by Peck et al. [1988], who visualized a system modeled by finite difference. The problem is three-dimensional, with three layers and a grid size of 30 by 40 in each layer. With a river crossing in 70 cells and 20 wells pumping, this relatively simple example requires 37,820 pieces of data. The reader can easily envision a situation where only a limited number of hydraulic head measurements are available to infer the transmissivity in each cell of this finite difference system. This example helps to illustrate that if one truly wants to characterize the hydraulic parameters of realistic aquifers, an ill-posed problem will result.

It is common in hydrogeologic practice to utilize nonlinear regression methods for parameter estimation. There are currently several private and public domain software packages that are used for this purpose. In order to achieve unique solutions and avoid unstable behavior in the results, it is necessary to parameterize or zone the problem in such a way as to define a smaller (usually much smaller) number of parameters than data points. As stated above, aquifers are highly heterogeneous, and hydraulic properties can vary significantly over very short distances thus making conceptual models based on large zones questionable in our opinion. Another difficulty lies...
in the estimation errors produced from nonlinear regression methods for overparameterized problems. It is important to note that the errors produced are a reflection of the errors in the measurements translated through the appropriate nonlinear kernel that governs the problem along with the regularizer chosen. More will be said on this subject in section 2. These errors do not reflect the spread of reasonably probable models, which is a limitation in the interpretation of estimation errors from nonlinear regression approaches. Therefore it is clear that new methods must be sought to solve the larger class of ill-posed characterization problems.

Alternate methods of inversion based on geostatistics and cokriging have been carried out [Kitanidis and Vomvoris, 1983; Hoeksema and Kitanidis, 1984; Kitanidis, 1996; Dugan, 1985]. At a particular point the unknown log transmissivity perturbations are solved for by weighted linear combinations of observed hydraulic head and transmissivity perturbations. The approach requires a linearized relationship between log transmissivity and hydraulic head perturbations in order to develop a cross-covariance function, which is valid if the variance of log transmissivity is small. As a result, for highly heterogeneous cases, cokriging cannot take full advantage of hydraulic head information collected at sample locations [Yeh et al., 1996; Zhang and Yeh, 1997].

Fundamental to the geostatistical approach are the inherent assumptions regarding the form of the drift and variogram functions and a database in hydraulic heads and hydraulic conductivities (or transmissivities) sufficient to establish covariograms and variograms. However, in geotechnical problems (for example, stability of landslides and design of open-pit mines) few hydraulic conductivity and hydraulic head data are usually available.

On the basis of the discussion above it is of interest to develop alternate methods of inversion that adopt different philosophies of tackling this fundamentally ill-posed problem of site characterization. The reader will note that after reading subsequent sections, the proposed approach shares many of the features and assumptions of the geostatistical method, namely, that we also adopt the small perturbation approach and linearized equations. In addition, one of the primary new aspects of our approach is accounting for the uncertainty in the hyperparameters characterizing the variogram, measurement error, and so on.

2. Inference Solutions to Ill-Posed Problems

As stated in section 1, groundwater inverse problems are nonlinear and, if conceptualized correctly, most likely highly underdetermined (many more unknowns than data points). The technique that permits the construction of a particular, stable inverse solution by introducing prior information is called regularization. In many problems, regularization is applied in a somewhat arbitrary manner regardless of the nature of the model we are seeking. This is the case of the widely used Levenburg-Marquardt [Press et al., 1992] method which has the advantage of imposing “smoothness” on the model. It is a common way to avoid the amplification of random errors associated with each observation. The well-known prewhitening technique used in spiking and predictive deconvolution is an example of quadratic regularization. In many situations, however, we may wish to use some other type of regularization that permits us to incorporate some relevant information about the model. This prior information may be available in a variety of forms. For example, positivity is a common deterministic constraint that is useful for solving a variety of inverse problems.

R. N. Tikhonov (as cited by Carrera and Neuman [1986]) showed that once an ill-posed problem is properly regularized, it becomes stable. Two things are required for a solution. One is \( \eta \) the regularization weight, and the other is a suitable derivative operator matrix which we will denote as \( % \) [Press et al., 1992, p. 801]. The matrix \( % \) is an arbitrary matrix chosen to reflect geologic “smoothness” in some sense.

Therefore there are two sources of uncertainty: One is the selection of the derivative operator defined by \( % \), and the other is the choice of the regularization parameter \( \eta \). It is well recognized by researchers in geophysical inversion that the choice of \( % \) is arbitrary [Menke, 1984]. For example, Oldenburg [1984] suggests that various choices of \( % \) relate to what is termed construction, namely, finding a variety of models that fit the data. Various choices of \( % \) define the smallest, flattest, smoothest, etc. structure [see also Press et al., 1992, pp. 800–803].

Another difficulty lies in the estimation errors produced from these codes. It is important to note that the errors produced are a reflection of the errors in the measurements translated through the appropriate nonlinear kernel that governs the problem along with the regularizer. These errors do not reflect the spread of reasonably probable models, which is one of the main goals in inverse problems for site characterization [Woodbury and Ulrych, 1998].

Jaynes [1973, 1983] examined these issues quite thoroughly from the perspective of comparing inference to direct inverse solutions. Jaynes showed how inference can convert ill-posed problems of deductive reasoning to well-posed problems of inference and also discussed various aspects related to stability of inference solutions. In our work we have adopted an inference approach (specifically Bayesian) to deal with ill-posed problems.

Much has been written on the subject of Bayesian inference, and different points of view apply (for a review see Woodbury and Ulrych [1998]). The reader will note that we refer to a “full-Bayesian” approach, and this is to signify that the inference problem will consist of both primary parameter and hyperparameter estimation [Mohammad-Djafari, 1996]. This subject will be discussed in section 4. Bayesian inference supposes that an observer can define a prior probability density function (pdf) for some random variable \( m \). This pdf \( p(m) \), can, in principle, be defined on the basis of personal experience or judgment. However, applications of Bayesian probability theory have been hampered by the precise meaning and interpretation of probabilities and the controversy surrounding the appropriate choice of prior pdfs. An orthodox view of probabilities dictates that frequencies measured in an experiment are equated to probabilities and “prior” information is not allowed. An alternative viewpoint of probability, denoted as the Jaynes-Cox viewpoint [Jowitt, 1979], is one in which probabilities are equated with the degree of plausibility of a proposition and may have no frequency interpretation whatsoever. This viewpoint is essentially Bayesian and is readily applicable to the questions that scientists and engineers typically ask. A necessary component of the Jaynes-Cox view is the “principle of maximum entropy” (PME) which replaces the need for subjective prior information in the Bayesian approach and forces all observers who possess common information to produce consistent results [Woodbury and Ulrych, 1998].

Woodbury and Ulrych [1993], Woodbury et al. [1995], and Woodbury [1997] deal with the estimation of appropriate prior
pdfs for hydrogeologic applications. As shown by Woodbury and Ulrych [1993], \( p(m) \) may have the form of a multivariate-truncated exponential distribution. This pdf preserves the statistical independence of the parameters. That is, if no correlation is known beforehand, the maximum entropy principle does not inject any correlation into the result. In this manner, \( p(m) \) has the most freedom in assigning realizations of the process. It is important to note that the above approach (PME) of determining \( p(m) \) is the one which is the most uncommitted with respect to unknown information.

In this paper we present a full-Bayesian approach for log transmissivity determination which combines Bayesian updating of log transmissivities and hydraulic head measurements. In the first phase, log transmissivity measurements are incorporated in Bayes theorem and the prior natural log transmissivity (\( \ln(T) \)) pdf is updated, yielding a posterior conditional mean and covariance values of log transmissivity. The first two central moments are generated assuming that the prior unconditional mean, variance, and integral scales are “hyperparameters”; that is, they are treated as random variables in themselves, contrary to classical statistical approaches. The final step consists of updating the conditioned \( \ln(T) \) field with hydraulic head measurements, utilizing a linearized aquifer equation. In a similar fashion it is assumed that the statistical properties of the noise in the measurements is also uncertain. At each step, uncertainties in all pertinent hyperparameters are removed by marginalization. Finally, what is produced are conditional-expected values and covariances of the \( \ln(T) \) field based on measurements of both hydraulic heads and log transmissivities. In addition, we can also produce resolution matrices and confidence (credibility) limits for the \( \ln(T) \) field.

3. Bayes’ Theorem

Bayes’ rule [e.g., Press, 1989] quantifies how a prior pdf can be changed on the basis of measurements. Simply stated, Bayes’ rule is

\[
\text{posterior} \propto \text{likelihood times prior.}
\]

Consider a vector of observed data \( d^* \). If the conditional pdf of \( d^* \), given \( m \) and some prior information \( I \), is given by \( p(d^* | m, I) \), Bayes’ rule states that

\[
p(m | d^*, I) = \frac{p(d^* | m, I) p(m | I)}{\int p(d^* | m, I) p(m | I) \, dm}.
\]

In (1), \( p(m | I) \) is the prior probability density of the model parameters, given some form of prior information \( I \), and \( p(d^* | m, I) \) is the likelihood of observing \( d^* \) given the model parameters (hypothesis) and the prior information. This latter term is often referred to as a “direct,” as opposed to a subjective, pdf. The direct pdf is called a “sampling distribution,” when the hypothesis is held constant and one considers different sets of data and a “likelihood” function when the data are held constant and one varies the hypothesis [Brethorst, 1988]. The term on the left-hand side is called the posterior probability density (after measurements are taken into account). Finally, the term in the denominator is a constant that ensures the posterior is normalized but is also the actual pdf of observing a set of data, with the uncertainty in the model parameters taken into account. Equation (1) can be seen as a general solution of the inverse problem. If a likelihood function can be defined (i.e., the forward model exists) and there is a compatibility between observed results and a prior understanding of the model parameters, that is, \( p(d^* | m, I) > 0 \) for some \( m \), where \( p(m | I) > 0 \), then Bayes’ rule implies that the posterior pdf exists and is unique [Tarantola, 1987, p. 53]. The posterior pdf may have more that one likelihood point, however. Indeed, \( p(m | d^*, I) \) may be pathological, as a result of infinite variance, or may be nonnormalizable (the denominator in (1) is unbounded).

4. Bayesian Solution to Linear Interpolation of Log Transmissivity Data

Bayesian updating methods provide an alternate philosophy to kriging for the characterization of input variables of a stochastic mathematical model. In this approach, a priori values of statistical parameters (for instance, mean and covariance) are assumed on subjective grounds or by analysis of a database from a geologically similar area. As measurements become available during site investigations, “updated” estimates of these parameters can be generated. Bryson and Ho [1969], Maybeck [1979], and Woodbury [1989] detail the updating procedure. The updated estimates can be used as input variables in further conditional simulations such as stochastic pore pressure or contaminant transport modeling. Massmann and Freeze [1987] applied such a methodology as a form of conditional simulation to quantify whether hydraulic conductivity measurements are cost-effective in reducing risk for owner/operators of landfill sites. Hachich and Vanmarcke [1983] applied Bayesian updating techniques in the interpolation of hydraulic head fields. Their algorithm first used an a priori hydraulic conductivity distribution to generate a mean hydraulic head field and covariance by a first-order second-moment procedure. Second, they updated this prior hydraulic head field with measurement values using a Bayesian procedure.

Kitanidis [1986] recognized that kriging is a special case of Bayesian estimation and used the Bayesian framework to examine the problem of parameter uncertainty in the estimation of spatial functions, specifically in the identification of trend parameters. Massmann and Freeze [1987, 1989] also compared kriging and Bayesian estimators. The Bayesian interpolation scheme naturally follows from the above discussion of Bayes theorem, and the reader is referred to Woodbury [1989] for further details. Let us assume that measurements of a stochastic random variable \( m \) are made at \( N \) locations in a discretized flow domain. The problem is to interpolate the \( N \) measured values to \( M - N \) other points and closely reproduce the data points themselves at the measurement locations. Therefore the \( N \) measurements of the variable are used to estimate \( M \) values of \( m \) by linear inversion. Here the \( N \) measurements of the variable form a vector \( d^* \) which is referred to as the observed “data.” The \( M \) values of \( m \) are the “model.” The data are mathematically formed as linear combinations of the model and random noise. Hence

\[
d^* = Gm + v,
\]

where

\[
d^* \quad N \times 1 \text{ vector of observed values, the “data”;}
G \quad N \times M \text{ matrix of coefficients;}
m \quad M \times 1 \text{ vector of unknown actual values of the data, the “model”;}
\]
$\mathbf{v}$ is a $N \times 1$ vector of random observational errors.

$G$ is a kernel which transforms data in $\mathbb{R}^N$, the data space, to $\mathbb{R}^M$, the model space, and consists of ones and zeros. Suppose there are $i = 1, \ldots, N$ measurement points and $j = 1 \ldots M$ interpolated points, where $N \ll M$. Where the $i$'th measurement point corresponds to an interpolated point, $G_{ij} = 1$; otherwise, $G_{ij} = 0$.

Let us reexamine the problem posed by (2). A direct and trivial solution of (2) is possible. It is the minimum length solution:

$$\hat{\mathbf{m}} = \mathbf{G}^T(\mathbf{G} \mathbf{G}^T)^{-1} \mathbf{d}^*.$$  

(3)

Note that $\mathbf{d}^* = \mathbf{G} \mathbf{m}_0$ reproduces the data $\mathbf{d}^*$ exactly, hence $\mathbf{m}_0$ consists of values of $\mathbf{d}^*$ at the observation points and zero elsewhere. The general solution $\hat{\mathbf{m}}$ can be represented by

$$\hat{\mathbf{m}} = \mathbf{m}_0 + \mathbf{m}^u,$$  

(4)

where the inner product $(\mathbf{m}_0, \mathbf{m}^u) = 0$. Here $\mathbf{m}^u$ lies in the orthogonal compliment to $\mathbf{m}_0$; hence all solutions (an infinite number) can be represented as linear combinations of the minimum length solution plus arbitrary terms. Note that in the underdetermined problem, there are only sufficient data to resolve that portion of the model in $\mathbf{m}_0$.

In order to obtain a unique solution to the inverse problem, we must have some means of singling out one candidate solution from the infinite number of possibilities. To do this, some information not contained in (2) must be added to the problem to banish all the $\mathbf{m}^u$ terms. One way to do this is to seek a model which is in some sense “small” in length away from a preconceived notion or behavior of the field, which is the classic Bayesian paradigm. Let us assume the following statistics on the model $\mathbf{m}$ and the noise vector $\mathbf{v}$ consistent with the Bayesian framework [Schweppe, 1973; Woodbury, 1989]. The noise $\mathbf{v}$ is random with a mean of zero and a covariance $C_v$:

$$E(\mathbf{v}) = 0,$$  

(5)

$$E(\mathbf{vv}^T) = C_v.$$  

(6)

If $\mathbf{v}$ represents measurement error, say, independent and identically distributed (iid), then $C_v$ is a diagonal matrix of variances of the observations, $\sigma_v^2 I$. The model $\mathbf{m}$ is random and characterized by a prior unconditional mean $\mathbf{s}$ and covariance $C_m$, which physically represents the correlation or spatial variability of the model $\mathbf{m}$:

$$E(\mathbf{m}) = \mathbf{s},$$  

(7)

$$E[(\mathbf{m} - \mathbf{s})(\mathbf{m} - \mathbf{s})^T] = C_m.$$  

(8)

In the cases presented in this paper, $C_m$ is represented by an exponential correlation structure

$$C_m(k, l) = \sigma_m^2 \exp \left[-\frac{(x_k - x_l)^2}{\lambda_i^2} + \frac{(y_k - y_l)^2}{\lambda_j^2}\right],$$  

(9)

where the usual definitions apply in that $\sigma_m^2$ is the variance of, say, $\ln(T)$, $\lambda$ is the integral scale, and $k$ and $l$ refer to two points in question.

If the combination of forward modeling and measurement errors are assumed Gaussian, then the probability of observing a set of data $\mathbf{d}^*$ given the model parameters is [Tarantola, 1987, p. 68]

$$p(\mathbf{d}^* | \mathbf{m}, I) = \frac{1}{(2\pi)^{N/2}(C_p)^{-1/2} \exp \left[-\frac{1}{2}(\mathbf{d}^* - \mathbf{Gm})^T C_p^{-1}(\mathbf{d}^* - \mathbf{Gm})\right]}.$$  

(10)

Here $N$ is the length of vector $\mathbf{d}^*$. If the prior distribution of the model is also assumed to be Gaussian, then

$$p(\mathbf{m} | I) = \frac{1}{(2\pi)^{N/2} \exp \left[-\frac{1}{2}(\mathbf{m} - \mathbf{s})^T C_q^{-1}(\mathbf{m} - \mathbf{s})\right]}.$$  

(11)

Here Tarantola [1987] illustrates the important result that if the forward modeling is linear, that is, if $\mathbf{d} = \mathbf{Gm}$, and if the above likelihood and prior information are both Gaussian, then the posterior density of $\mathbf{m}$ is Gaussian. This result can be verified by performing the integration in (1) analytically [see Backus, 1988]. The resulting posterior pdf $p(\mathbf{m} | \mathbf{d}^*, I)$ is Gaussian:

$$p(\mathbf{m} | \mathbf{d}^*, I) = \frac{1}{(2\pi)^{N/2} \exp \left[-\frac{1}{2}(\mathbf{m} - \langle \mathbf{m} \rangle)^T C_q^{-1}(\mathbf{m} - \langle \mathbf{m} \rangle)\right]}.$$  

(12)

The first two conditional moments of this pdf are given by Tarantola [1987, equation (1.93)]:

$$\langle \mathbf{m} \rangle = \mathbf{s} + C_p G^T (G C_p G^T + C_p)^{-1}(\mathbf{d}^* - \mathbf{Gs}),$$  

(13)

$$C_q = C_p - C_p G^T (G C_p G^T + C_p)^{-1} G C_p,$$  

(14)

where $\langle \mathbf{m} \rangle$ and $C_q$ are the conditional expected value and covariance of the posterior pdf, respectively.

Such a technique would appear at first glance to offer considerable advantages over kriging in that no prior computation of variograms is necessary. However, a form of the spatial correlation must be assumed. In the traditional implementation of the updating procedure [e.g., Massmann and Freeze, 1987] it may be possible in pathological cases to generate an updated covariance matrix that is not positive-definite because of finite machine precision and numerical roundoff.

In realistic cases more may be known about the form of the underlying pdf of the model (say, multivariate Gaussian) than its magnitude $\sigma_m^2$ or other parameters governing the pdf. It is also assumed that $\sigma_m$ is known. Suppose that the form of the pdf is known but the values of the “hyperparameters” such as $s$, $\sigma_p$, $\sigma_q$, and $\lambda$ are not. In these cases it may be desirable to generate a series of updated (\mathbf{m} values for a wide range of covariances of different magnitudes. It would then remain a problem to choose which candidate solution is “best” in some sense. Massmann and Freeze [1989] propose a method, again based on Bayesian inference, to change the statistical properties of the original field after measurements have been collected but before an update is performed. One could alternatively let the mean, variance, and integral scale be treated as unknown hyperparameters. Both the hyperparameters and the interpolation could then be solved for in one step by a nonlinear maximum likelihood method [Hoeksema and Kitanidis, 1985]. Mohammad-Djafari [1996] also details many strategies with respect to this problem. One interesting technique, as detailed by others [e.g., Kitanidis, 1986; Loredo, 1990; Rubin and Dagan, 1992], treats the hyperparameters as “nuisance” parameters that are “removed” from further consideration by integration over these parameters (marginalization). This point is discussed further below [see also Woodbury and Rubin, 2000].
5. Full-Bayesian Approach to Linear Inversion

Recall the solution to Bayes’ theorem for the case of Gaussian priors and a linear functional transformation, (12), noting that

\[
\langle m \rangle = \int p(m|d^*, I)m\,dm.
\]  

(15)

However, consider that we have a set of hyperparameters \((s, \sigma_\gamma^2, \lambda, \sigma_\mu^2) = u\), then

\[
\langle m \rangle = \int \int p(m|d^*, I, u)p(u)m\,du\,dm.
\]  

(16)

Changing the order of integration results in

\[
\langle m \rangle = \int p(u)\left[\int p(m|d^*, I, u)m\,dm\right]du.
\]  

(17)

However, the term within the brackets is equal to (13), and therefore (15) becomes

\[
\langle m \rangle = \int s + C_pG^T(GCG^T + \sigma_\mu^2I)^{-1}(d^* - Gs)p(u)\,du.
\]  

(18)

If the data errors are iid, then the conditional mean is

\[
\langle m \rangle = \int s + C_pG^T(GCG^T + \sigma_\mu^2I)^{-1}(d^* - Gs)p(u)\,du.
\]  

(19)

Similarly, the covariance \(C_q\) is

\[
C_q = \int C_m + C_pG^T(GCG^T + \sigma_\mu^2I)^{-1}GC_p\,p(u)\,du.
\]  

(20)

Since \(\langle m \rangle\) and \(C_q\) are functions of \(u\) (equation (13)), the integration of (19) and (20) involving the hyperparameters must be carried out through some form of numerical integration. These can be accomplished using the Monte Carlo method and the concept of importance sampling [e.g., Shreider, 1966, p. 100–102; Woodbury and Sudicky, 1992]. The integrals posed by (19) and (20) are evaluated by generating a series of random model vectors using a multivariate random number generator with \(p(u)\) as the pdf. The reader will note that while it is conceptually appealing to evaluate (19) and (20), this approach is restricted to a small number of hyperparameters because of computer storage and computational effort. The fundamental problem is how to specify, in a logical and consistent manner, the prior pdfs for these hyperparameters. This subject is discussed in section 8.

6. Steady State Groundwater Inversion

The basic methodology that we will be following with respect to groundwater inversion is a linearized approach detailed by Hoeksema and Kitsalidis [1984]. The essence of their approach is presented here. The hydraulic head \(\phi\) and the aquifer transmissivity \(T\) satisfy the following partial differential equation in terms of the log transmissivity, \(Y = \ln(T)\),

\[
\partial Y/\partial \phi + \partial^2 Y/\partial x^2 + \partial^2 Y/\partial y^2 = 0
\]  

(21)

Equation (21) is separated into deterministic and stochastic terms. We define \(H\) as the expected value of the hydraulic head field, \(h\) as a zero-mean head perturbation, and \(F\) as a spatially constant unconditional expected value of log transmissivity (which is constant for a zero-order intrinsic function), and \(f\) is a zero-mean log transmissivity perturbation. After some manipulation, the hydraulic head field \(\phi\) becomes the sum of two separate linearly superimposed solutions, one related to the solution to Laplace’s equation

\[
\partial^2 H/\partial x^2 + \partial^2 H/\partial y^2 = 0
\]  

(22)

and the other related to the solution of

\[
\partial^2 h/\partial x^2 + \partial^2 h/\partial y^2 = \partial f/\partial x + \partial f/\partial y
\]  

(23)

Equation (23) is solved by the finite element (FE) method and takes the form

\[
Ah = Bf + Ch_h
\]  

(24)

where \(A, B,\) and \(C\) are constant matrices, \(h\) is a vector of nonboundary hydraulic head perturbations, \(h_h\) is a vector of boundary node head perturbations, and \(f\) is a vector of log transmissivity perturbations at the nodal points. The boundary node perturbations are considered to be known in this exercise and therefore are set to zero. Recall the fundamental properties, \(\phi = H + h\) and \(Y = F + f\), where the boldface type represents discrete quantities expressed at the nodal points. Solving for the hydraulic heads at nodal points yields

\[
\phi - H + A^{-1}BF = A^{-1}BY.
\]  

(25)

Noting that \(B\) acts as a derivative operator and that the derivative of a constant value of log transmissivity is zero, we have

\[
D(\phi - H) = D(A^{-1}B)Y.
\]  

(26)

which is in the same form as \(d^* = Gm\). The matrix \(D\) is a simple Boolean matrix (consisting of ones and zeros) that filters out the computed values of heads at points other than those corresponding to measurement points. Consequently, \(D(\phi - H), D(A^{-1}B),\) and \(Y\) are analogous to \(d^*, G,\) and \(m\), respectively. The second-phase hydraulic head updating can then exactly use the same equations as the first-phase update that uses only transmissivity measurements.

7. Summary of Procedure

The total procedure is as follows. First, a log transmissivity field is envisioned as being statistically distributed as multivariate Gaussian with an exponential correlation structure. However, the spatially constant unconditional mean value of the field, the variance of \(\ln(T)\), and integral scales are unknown. Measurements of \(\ln(T)\) are taken, and the first Bayesian update of this field is performed, incorporating the uncertainty of the unknown hyperparameters, \(s, \sigma_\gamma^2,\) and \(\lambda.\) Next, measurements of hydraulic head are taken into account, and in the second phase a Bayesian update of the \(\ln(T)\) field is again performed using the linearized groundwater flow equation. In this phase, uncertainty in the measurements and any modeling
Plate 1. (a) The natural logarithm transmissivity ($\ln(T)$) field generated from a fictitious aquifer problem. Mean $\ln(T)$ value is 7.0, variance is 0.7, and integral scale is 100 km. (b) SURFER™ updated $\ln(T)$ field from 20 sample points. (c) Bayesian updated $\ln(T)$ field from 20 sample points. (d) Bayesian updated $\ln(T)$ field from 20 sample points of heads and $\ln(T)$. 
error of the heads is taken into account. Section 8 describes how the prior pdfs for all the hyperparameters are determined.

8. Hyperparameter Prior Probabilities

8.1. Complete Ignorance: Scaling Parameter

We first consider the important quest in determining priors that express complete ignorance so that our estimates will not be biased by uncertain knowledge. We quote from Jeffreys’ [1939, p. 39] epic book:

Our first problem is to find a way of saying that the magnitude of a parameter is unknown, when none of the possible values need special attention. Two rules appear to cover the commonest cases. If the parameter may have any value in a finite range, or from $-\infty$ to $+\infty$, its prior probability should be taken as uniformly distributed. If it arises in such a way that it may conceivably have any value from 0 to $\infty$, the prior probability of its logarithm should be taken as uniformly distributed.

Imagine for a moment the concept of a scaling parameter such as the standard deviation $\sigma$. Suppose, however, that we have no knowledge whatsoever about its magnitude; we know only that it runs from zero to infinity. If we adopt a uniform prior for $\sigma$, we face conceptual difficulties in considering power transformations of $\sigma$. Consider the variance $\sigma^2$. Transforming the pdf of the uniform pdf for the standard deviation into a pdf for the variance implies that the variance is distributed as $p(\sigma^2) \propto 1/\sigma$ which is inconsistent with the knowledge that we also know nothing about the variance, knowing only that it too runs from zero to infinity. The whole area of “ignorance” or reference priors has attracted considerable attention within the Bayesian community [e.g., Bretthorst, 1988]. It is generally considered that a reasonable choice for these scaling parameters is [see also Sivia, 1998, p. 105]:

$$p(\sigma) \propto 1/\sigma.$$  

This particular pdf is invariant under power transformations; however, it is “improper” in the sense that it is not normalized. Fortunately, even when it is used in Bayes’ theorem, the posterior pdf is normalized. Sivia [1998, p. 129] discusses these issues in some detail. As a technical convenience, we can adopt a normalized form of the pdf, namely,

$$p(\sigma|I) = \tilde{a}\sigma^{-1},$$  

where $\tilde{a}^{-1} = \text{ln}(U/L)$ and $L$ and $U$ are a lower and upper bound, respectively, for a particular problem. This form is used to set very large upper bounds and small lower bounds as necessary to generate random number sequences out of this pdf. One can typically examine the posterior as $L \to 0$ and $U \to \infty$.

8.2. Lower and Upper Bounds With an a Priori Mean

Woodbury and Ulrych [1993] deal with the estimation of appropriate prior pdfs for hydrogeologic applications, and the essence of their results is briefly repeated here. Specifically, it is acknowledged that hydrologic databases are such that reasonable upper and lower bounds are obtainable for virtually all model parameters. This information implies that our base level of knowledge is a joint boxcar pdf (uniform distribution between an upper and lower bound). Suppose additional information $I$ such as a geological interpretation, an informed guess, limited tests, or calibration becomes available. This “new” prior estimate of the model is defined here as $s$. Assume $s$ is the expected value vector of a pdf $p(m)$ which is chosen so that it has minimum relative entropy with respect to a boxcar pdf, subject to the expected value constraints $s_n$. As shown by Woodbury and Ulrych [1993], $p(m)$ has the form

$$p(m|I) = \prod_{n=1}^{M} \frac{-\beta_n}{\exp(-\beta_n U) - 1} \exp(-\beta_n m_n),$$  

which is a multivariate truncated exponential. The $\beta_n$ are Lagrange multipliers which must be determined from the upper and lower bounds and the expected value constraints [see Woodbury and Ulrych, 1993, Appendix B]. By definition, this equation satisfies the expected value constraints

$$\int p(m|I)m dm = s.$$  

Thus, unless further constraints are included, the expected values $p(m)$ yield the initial model estimates $s$. It is important to note that the above approach of determining $p(m)$ is the one which is the most uncommitted with respect to unknown information.

8.3. Mean and Variance Specified

Given prior information $I$ that consists of the mean value $\bar{s}$ of a random variable $x$ and the second-central moment $\sigma^2$, an application of the maximum entropy principle leads to a Gaussian distribution for $x$ [Bretthorst, 1988; Kapur, 1989]:

$$p(x|\sigma, s, I) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-s)^2}{2\sigma^2}\right).$$  

(29)

Here (29) is the least informative prior probability density for a variable that is consistent with the given second moment. Even if the second moment of a variable is not known, the central limit theorem leads to the Gaussian form [Jaynes, 1983].

8.4. Summary: Prior pdfs

Summarizing this section, we adopt three types of pdfs as priors in the inverse procedure.

1. The $\ln(T)$ field is assumed to be a stochastic random variable with constant unconditional mean $s$ and covariance $C_p$. However, the parameters associated with this variable (as typical in the Bayesian paradigm) are also random. Note $C_p$ is a function of the hyperparameters $\lambda$ and $\sigma^2$. Therefore, although the $\ln(T)$ field is viewed as being multivariate Gaussian, the actual parameters of that Gaussian distribution are possibly unknown.

2. The $\ln(T)$ field is measured at various points to produce a vector $d^*$. There are measurement/modeling errors attached to each one of these points, but as a whole the errors are assumed to be iid and specified as $d, I$. Note that $s_d$ is also treated as a hyperparameter in that we may not be able to quantify the errors in any specific way. These are therefore distributed according to the Jeffreys [1939] prior pdf.

3. In the absence of any other information the prior unconditional mean may be considered to be a location parameter, that is, a uniform pdf between two bounds. However, $s$ can also be estimated from the measurements of $\ln(T)$ that are available. Typically, the prior mean value estimated this way is normally distributed, regardless of the underlying distribution of the variable itself, as a consequence of the central limit
4. It may or may not be feasible to estimate \( \lambda \) from variogram data, as the number of pairs of points may actually be quite small. Therefore \( \lambda \) may be distributed as a truncated exponential or treated as a scaling parameter depending on the information provided. Note that the exponential distribution defaults to uniform if \( s = (L + U)/2 \).

5. The hydraulic head \( h^* \) is sampled at various locations. Note that \( \sigma_h \) is also treated as a hyperparameter in that we may not be able to quantify the errors in any specific way. These can also be described with the Jeffreys prior pdf.

9. Noise-Free Simulations and Exact Hyperparameters

The two-dimensional numerical application of the full-Bayesian approach to the inverse problem will be tested in this section through simulation. A log transmissivity field is first generated along with its associated hydraulic head field. This generation is accomplished by using a multivariate random number generator, with known statistical properties. The hydraulic head values are solved for using the finite element procedure, utilizing linear basis functions and triangular elements. Next, random samples of the computed hydraulic head field are taken from the fictitious aquifer. The number of samples are initially 10, then 20, 40, and finally 80. There are 1089 nodes and 2048 elements in the FE grid, with 961 degrees of freedom. For the four cases above this gives 951, 941, 921, and 881 unknown ln(\( T \)) values to be solved for in the finite element grid. The ln(\( T \)) values correspond to point values that also correspond to hydraulic head points. Thus, at every measured head location, there also exists a transmissivity measurement. Details of the grid, statistical parameters, etc. are described below.

In the second step the point values of ln(\( T \)) are used in the Bayesian updating procedure. In this first phase of analysis the statistical properties of the aquifer are assumed to be known, that is, determined from the basis of substantial testing in some other area. In addition, the sampled ln(\( T \)) points are assumed to be noise free. Recall that the Bayesian procedure requires a prior pdf and uses the measurement points to condition this pdf. In the noise-free case the ln(\( T \)) measurement points are reproduced exactly.

In the third step the point values of hydraulic head are used in the last Bayesian update. These values are also taken as being noise free in this phase. This last update utilizes the linearized aquifer equations and the corresponding linear inverse (26). Finally, the a posteriori ln(\( T \)) field is produced which is the conditional expected value of the pdf

\[
\langle m \rangle = \int p(m|Y^*, \phi^*, I) m \, dm.
\]

The aquifer postulated as a test case is square, 33 rows by 33 columns, with each side being 300 km long. This test case is very similar to that described by Hoeksema and Kitanidis [1984]. The heads are prescribed on every boundary. The statistical properties of ln(\( T \)) are as follows: mean value of 7.0, standard deviation of 0.70, and integral scales in \( x \) and \( y \) of 100 km. Note that the grid sizes are much smaller than the integral scale of the process (9 km).

The first test of the numerical procedure is to generate a realization of the ln(\( T \)) field and solve both the traditional aquifer equation (21) as well as the linearized form (23). In this way we can check that the linearization (normally restricted to \( \sigma_h^2 < 1.0 \)) produces a reasonably accurate field. The ln(\( T \)) field is shown in Plate 1a, and the hydraulic heads are shown in Figure 1. Notice that the linearized solution reproduces the “true” hydraulic heads very well. The standard deviation (RMS) error of the true field against the linearized solution is 0.133, which is less than 0.07% of the maximum hydraulic head difference across the aquifer.

We now show results from two of the simulations performed, that is, from 20 and 80 sample point examples. In each case we also present a “blind” kriging example in which the sample points are interpolated over a 50 by 50 grid with SURFER™, and the results contoured. In this way we can point out if any improvements over a standard interpolation (something a practitioner would use) are realized. Plate 1b depicts this interpolated field. Plate 1c shows the first step, the Bayesian update of the ln(\( T \)) field based on the 20 point sample.

The last step is shown on Plate 1d, which is conditioned by the hydraulic heads. Comparing Plates 1a, 1c, and 1d, we see an improvement in the resolution of the “true” field. The same basic exercise is performed with 80 sample points, and these are shown in Plates 2a–2c. Especially with the last example, with only 80 points the Bayesian procedure produced a ln(\( T \)) map very close to the “true” field.

The results of this section can be best summarized in tabular form (Table 1), including runs that were not presented graphically. As shown, in every case the addition of hydraulic head data aid considerably in the refinement of the ln(\( T \)) field. At a glance one can see the conditioning effects of ln(\( T \)) and hydraulic head data as illustrated in Plate 2d in which the standard deviations in ln(\( T \)) are plotted.

10. Noise-Free Simulations and Uncertain Hyperparameters

In this section the full-Bayesian update is carried out with both ln(\( T \)) and hydraulic head data, where the hyperparameters are treated as unknowns. The hydraulic head and ln(\( T \)) data are noise-free. The example chosen is the 80 point case as described in section 9. It is assumed that a modeler has access to the 80 point sample for heads and ln(\( T \)) values, and statistical analysis is then performed. The 80 point sample has a ln(\( T \)) mean of 6.71 and a variance of 0.41. For 80 points the standard deviation of the mean is 0.07, and the standard deviation of the variance is 0.06 (assumed Gaussian). The process integral scale is more problematic to estimate [Woodbury and Sudicky, 1991]. Plate 3a is a plot of the experimental semivariogram from the 80 point sample. Here the classic estimator is used with a basic lag of 12 km and one direction with a 360° spread. The bold long-dashed line is the “correct” or true variogram in this case. This is a good example of the difficulties in estimating the pertinent statistical parameters from variograms. A practitioner would likely have great difficulty in uniquely determining these quantities, let alone decide that a trend was not present. The solid green line on Plate 3a is produced from an exponential variogram model with a variance of 0.41 (the variance of the sampled data) and a calibrated integral scale of 100 km.

At this point the reader will note that more sophisticated methods can be used to estimate both the variance and integral scales [e.g., Woodbury and Sudicky, 1991]. However, without
Plate 2. (a) SURFER™ updated ln(T) field from 80 sample points. (b) Bayesian updated ln(T) field from 80 sample points. (c) Bayesian updated ln(T) field from 80 sample points of head and ln(T). (d) Standard deviations of the interpolated ln(T) field from 80 sample points.
Plate 3. (a) Experimental semivariogram from the 80 point case. Bold green line is a calibrated fit of an exponential variogram function with a variance of 0.41 and integral scale of 100 km. See legend for other parameters. (b) The $\ln(T)$ field for high variance (2.25) example. (c) Bayesian updated $\ln(T)$ field from 80 sample points, high variance problem. (d) Bayesian updated $\ln(T)$ field from 80 sample points of head and $\ln(T)$, high variance problem.
loss of generality we have chosen to estimate the integral scales based on the method suggested by Rehfedt [1988]. This could be an example of what a practicing hydrogeologist might do and is applicable as an example of a highly uncertain estimate of the integral scale and its effect on the Bayesian update. Upper and lower (99%) bounds on the variance can be obtained from the sample, as indicated above. These are 0.60 and 0.22, respectively. Then, fixing each of these quantities, a calibrated integral scale for each can be produced, which is 150 km and 50 km, respectively. This establishes upper and lower bounds on the integral scale. However, the integral scale pdf cannot be assumed to be Gaussian because the variance, integral scale, and exponential model for the variogram are not linearly related. For the purposes of the Bayesian update of the \( \ln(T) \) field, \( \lambda \) is assumed to be exponentially distributed with a mean of 100 km, lower bound of 50 km, and an upper bound of 150 km.

The results of combined \( \ln(T) \) and hydraulic head data are not shown in this case as they are nearly identical to the plot of Plate 2c. The RMS errors are only slightly higher in this case, 0.250 as opposed to 0.247 (see Table 1), in spite of the large uncertainty in the hyperparameters. This is, indeed, an encouraging result.

### 11. Noisy Simulations and Uncertain Hyperparameters

In this section the full-Bayesian update is carried out with both \( \ln(T) \) and hydraulic head data, and, in this case, the hyperparameters are treated as unknowns. Also, a high standard deviation in \( \ln(T) \) is simulated. Here the statistical properties of \( \ln(T) \) are as follows: mean value of 7.0, variance of 2.25, and integral scales in \( x \) and \( y \) of 100 km. See Plate 3b. Since this simulation involves a high variability in log transmissivity, an immediate question is then posed, How good is the linearized solution which is a key to implementing the Bayesian solution?

Figure 2 shows a plot of the approximate and “true” hydraulic head data. As shown and as can be quantitatively verified, the approximation is good even at a variance of \( \ln(T) \) of 2.25. The fit of the line here should be linear for a good approximation, and it is; a best fitting regression line has a slope of 1.002 and an RMS error of 0.59 m or just 0.29% of the max-

![Figure 1](image.png)

**Figure 1.** Hydraulic head fields generated with aquifer equation (solid lines) and linearized form (dashed lines).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>RMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>B10</td>
<td>0.472</td>
</tr>
<tr>
<td>H10</td>
<td>0.385</td>
</tr>
<tr>
<td>B20</td>
<td>0.405</td>
</tr>
<tr>
<td>H20</td>
<td>0.321</td>
</tr>
<tr>
<td>B40</td>
<td>0.304</td>
</tr>
<tr>
<td>H40</td>
<td>0.254</td>
</tr>
<tr>
<td>B80</td>
<td>0.295</td>
</tr>
<tr>
<td>H80</td>
<td>0.247</td>
</tr>
</tbody>
</table>

**Table 1.** Statistics of Interpolation and Inversion

RMS Refers to the square root of the mean square difference between the interpolated field and the “true” field. B10, B20, B40, and B80 refer to a Bayesian update for the 10, 20, 40, and 80 point sample, respectively; H10 and so on refer to the \( \ln(T) \) field updated from the hydraulic head and \( \ln(T) \) interpolated fields.
The hydraulic head data are assumed to be noisy, with the noise variance unknown and therefore distributed as a Jeffreys pdf. The results of the ln(T) update are shown on Plates 3c and 3d. Comparing Plates 3c and 3d with Plate 3b, one can see that hydraulic head conditioning does affect, and actually improves, the resolution of the reconstructed ln(T) field.

12. Summary and Conclusions

This paper presents a methodology for the spatial inversion of transmissivity from hydraulic head and transmissivity measurements for the two-dimensional steady state groundwater flow case. The methodology used is based on a full-Bayesian approach [Woodbury and Rubin, 2000]. First, the ln(T) data are interpolated from a set of sparse measurements to nodal points in a finite element grid. The full-Bayesian approach assumes that the ln(T) field is multivariate Gaussian with an exponential correlation structure. However, the parameters governing this stochastic process are assumed to be unknown and are treated as hyperparameters. The pdfs of these hyperparameters are, in turn, determined from maximum entropy and consistency considerations. In other words, pdfs are chosen for each of the hyperparameters that are maximally uncommitted with respect to unknown information. The ln(T) interpolation follows along the approach determined by Woodbury [1989] with the exception that the hyperparameters are essentially viewed as being “nuisance” parameters whose effects in the Bayesian update are removed by marginalization. This Bayesian methodology is quite general and provides an alternative to kriging for spatial interpolation.

Second, the partial differential equation relating hydraulic head to ln(T) perturbations for a two-dimensional confined aquifer with a given head along the boundary is linearized, yielding a linear inverse problem for the unknown ln(T) fluctuations. A set of hydraulic head measurements are taken, and the linear inverse problem is solved via another Bayesian update, using the ln(T) update in the previous step as a prior probability. The hydraulic head measurements are also assumed to be “noisy,” and these effects are removed by marginalization. The combined two-step Bayesian update produces conditional expected values of ln(T) over the aquifer and estimation errors.

The procedure is applied to a series of test cases in which the actual values of ln(T) and hydraulic head are generated with known values of the stochastic parameters. The test aquifer is similar to that described by Hoeksema and Kitanidis [1984]. It is square, with 300 km sides and constant hydraulic head boundaries. In a first series of tests, 10, 20, 40, and 80 ln(T) and hydraulic head values are used in the two-step procedure with a ln(T) standard deviation of 0.7. There are 951, 941, 921, and 881 unknown ln(T) values solved for. In this series the hyperparameters are assumed to be known, and the Bayesian updating produced is used. The use of the hydraulic head data is shown to improve the ln(T) estimates in comparison to simply interpolating the sparse ln(T) data alone.

The procedure is also applied to the case in which 80 ln(T) and hydraulic head measurements are available; however, the statistical parameters are unknown. Basic statistics of the sampled data are used to derive pdfs of the hyperparameters. Variograms are also used to partially define upper and lower limits of integral scales. The results of combined ln(T) and hydraulic head data are nearly identical to the certain-hyperparameter case (the RMS errors are 0.250 as opposed to 0.247, see Table 1) in spite of the large uncertainty in the hyperparameters. This is, indeed, an encouraging result.

One final simulation is carried out on a highly variable field with ln(T) variance of 2.25. This value is quite high, and one would normally question the validity of the linearization inherent in the approach. However, the approximation is shown to produce unbiased but noisy heads. The Bayesian approach produced a reasonable facsimile to the true picture even when the hydraulic head data are noisy.

Acknowledgments. The authors would like to express their sincere appreciation to T.-C. (Jim) Yeh for a series of very helpful and stimulating discussions of cokriging. Three anonymous reviewers also provided good suggestions to improve the manuscript. This work was supported by a grant to the authors by the Natural Sciences and Engineering Research Council (NSERC). Commercial products mentioned in this paper are for identification purposes only and do not represent endorsement by the authors.

References


Figure 2. Approximate and “true” hydraulic heads for the high variance in ln(T) example.

T. J. Ulrych, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4 (ulrych@eos.ubc.ca)
A. D. Woodbury, Department of Civil Engineering, 342 Engineering Building, 15 Gilson Street, University of Manitoba, Winnipeg, Mani- toba, Canada R3T 5V6 (woodbur@cc.umanitoba.ca)

(Received June 29, 1999; revised March 27, 2000; accepted March 28, 2000.)