Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis

PETER K. KITANIDIS

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

Linear estimation has found many applications in the inference of spatial functions in surface and subsurface hydrology. The effect of parameter uncertainty is examined in a Bayesian framework with emphasis on the derivation of the Bayesian distribution (and its first two moments) of unknown quantities given some measurements. This distribution accounts not only for natural variability but also for parameter uncertainty. For known covariance parameters the Bayesian distribution is Gaussian (for Gaussian processes) with the mean being a given linear function of the data. This linear estimator is equivalent to the conventional Gaussian conditional mean estimator for a priori known drift coefficients and is the same with kriging for diffuse prior distribution of the drift coefficients; however, the developed procedure is more general. When both drift and covariance function parameters are uncertain, the Bayesian distribution is generally not Gaussian, and the Bayesian conditional mean is a nonlinear estimator. The case of diffuse priors is examined in some detail; it is shown that the posterior distribution of the covariance function parameters is given by the restricted likelihood function, i.e., the likelihood function of generalized increments. The results provide insight into the applicability of maximum likelihood versus restricted maximum likelihood parameter estimation, and conventional linear versus kriging estimation. A more general procedure which includes these methods as special cases is presented.

INTRODUCTION

Hydrologic quantities such as rainfall, transmissivity, piezometric head, and solute concentration vary in space in ways too complex to be represented through simple deterministic functions. In many cases the most appropriate way to represent their spatial variability is in statistical terms, through mean values, variances, correlations, or probabilities of exceedence. In this case the variable of interest is represented as a realization of a random function (a spatial stochastic process). This representation is very general, does not compromise the physical basis of hydrologic models, and allows utilization of any available piece of information by conditioning on prior information or measurements.

The theory of spatial stochastic processes has found many applications in the solution of problems of inference, such as estimation of point values (interpolation) and of areal averages (such as block-averaged transmissivities or mean areal precipitation) from a few sampled values [e.g., *Delhomme*, 1979; *Chua and Bras*, 1982] and the solution of the inverse problem in groundwater modeling [*Hoeksema and Kitanidis*, 1985b]. The method which is almost universally used is to assume that the estimator is a linear function of the measurement and to seek the weights which minimize the estimation variance. Such linear minimum variance estimators assume known values of some parameters but are computationally very efficient and yield not only optimal weights but also the variance of estimation error. This variance depends only on the covariance function of the random function and the location of the samples, making linear estimation a useful tool in network analysis (see the special section on the Chapman Conference on Design of Hydrologic Data Networks, in *Water Resources Research*, volume 15, pp. 1673–1871, 1979). For example, one may design the network of observations which minimizes the estimation variance of an areal mean or point value.

Linear estimation assumes that the mean and the covariance (or only the covariance function, if kriging is used) are

known with certainty. In reality both the functional form and the parameters are often determined from the same sample used for interpolation or averaging. Estimation of parameters from a limited-size sample involves error which is seldom taken into account or even acknowledged. This is usually justified on the basis of the well-established property that given enough data the minimum variance estimates are relatively insensitive to errors in the parameters or the model. Furthermore, in mining exploration, where these methods have been successfully applied, many measurements are available from which variograms may be inferred. In hydrology, estimation must often be based on rather small samples. *Hughes and Lettenmeier* [1981] argued that the optimality of linear estimators is predicated on knowledge of some properties of the random function and, consequently, is questionable when the parameters or the model are partially unknown. These authors then proceeded to evaluate through Monte Carlo simulations the effect of parameter estimation error on kriging estimators. *Kitanidis and Vomvoris* [1983] and *Hoeksema and Kitanidis* [1984, 1985a, b] used a maximum likelihood parameter estimation procedure which also yields some measures of the reliability of parameter estimates. Sensitivity analysis may then be used to evaluate the effects of parameter error on linear estimation.

The problem of parameter estimation is certainly not limited to applications of linear estimation theory. For example, in hydrogeology the properties of the random function representing log permeability are used in the flow and mass transport equations to derive the properties of the functions representing piezometric head or solute concentration. (For example, see *Smith and Freeze* [1979], *Gelhar and Axness* [1983], and *Dagan* [1984].) Ultimately, these parameters must be determined from data.

This paper examines the problem of parameter uncertainty and its effect on the estimation of spatial functions in a Bayesian framework. It may be useful to clarify that the term "Bayesian" here means that the parameters themselves are viewed as random variables and Bayes' theorem is used to revise their probabilities when new information becomes available. Prior information, and in particular subjective opinions, may or may not be available. Bayesian analysis provides a general framework in which parameter uncertainty is recog-

nized and its effect on estimation or decision can be evaluated. There have been several applications of Bayesian analysis in hydrologic time series. For example, *Valdés et al.* [1977] examined the normal multivariate case in the Bayesian generation of synthetic streamflows. However, the author is unaware of any previous similar analysis in the framework of interpolation, averaging, or differencing of spatial processes. Even in the general statistical literature, applications of Bayesian techniques to multivariate problems appear only scarcely. Most published results are related to normal regression [e.g., *Raiffa and Schlaifer*, 1961; *Geisser*, 1965; *Halpern*, 1973].

Among the contributions of this paper is that it provides much insight into the applicability of suboptimal (for imperfectly known parameters) estimators, such as conventional linear minimum variance estimation and kriging. The often misunderstood procedures of kriging and restricted maximum likelihood parameter estimation [*Kitanidis and Lane*, 1985] are given, apparently for the first time, a Bayesian interpretation. The methodology is applicable to processes with constant or spatially variable mean (drift), to stationary or stationary-increment processes, and to cases of a single variable or of multiple related variables (e.g., piezometric head and log transmissivity). To facilitate the analysis and presentation, a concise matrix-vector notation is introduced in section 2, where the general model is defined. Also, linear estimation is reviewed, and the popular kriging equations are given in a concise form. Section 3 reviews the essentials of Bayesian analysis and establishes the notation which is used in subsequent sections. Section 4 derives the posterior distribution of drift coefficients and the Bayesian (predictive) distributions in the case of known covariance function parameters. Section 5 extends the results of section 4 in the case of covariance function being proportional to a partially unknown parameter. Section 6 presents the general case of both drift and covariance function parameters partially unknown, with emphasis on the particular case of negligible prior information about drift coefficients. Chapter 7 presents an extensive discussion of the results of this analysis and its relevance to applications.

2. GENERAL MODEL AND LINEAR ESTIMATORS

The common assumption is that the random function is given by the general linear model:

$$y(\mathbf{x}) = \sum_{i=1}^{p} f_i(\mathbf{x})\beta_i + \varepsilon(\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ is the vector of spatial coordinates of the point where y is sampled; β_i, $i = 1, \cdots, p$ are (generally unknown) parameters; $f_i(\mathbf{x})$, $i = 1, \cdots, p$ are known functions of the spatial coordinates; and $\varepsilon(\mathbf{x})$ is a zero-mean spatial random function. The first term on the right-hand side represents a drift (or trend or "deterministic part"), and the second term represents a zero-mean random field ("stochastic part"). Random measurement error, if it exists, may be represented through a term which may be absorbed in ε. The stochastic part has covariance function $R(\mathbf{u}, \mathbf{v})\theta$, where θ are parameters, defined through

$$E[\varepsilon(\mathbf{u})\varepsilon(\mathbf{v})] = R(\mathbf{u}, \mathbf{v} \mid \theta) \tag{2}$$

The form of the drift and of the covariance function will be assumed known except for the numerical values of some parameters. In principle, uncertainty in the form of the drift and the covariance function may be neglected if a general enough model is assumed. It will be useful to distinguish between the drift coefficients β and the covariance function parameters θ.

At various stages of the analysis they may be assumed perfectly unknown, partially known or deterministic.

From this general representation one may obtain as special cases most of the useful and commonly assumed models. For example, for $p = 1$, $f_i(\mathbf{x}) = 1$, and $R(\mathbf{u}, \mathbf{v}) = r(|\mathbf{u} - \mathbf{v}|)$, equation (1) represents a stationary isotropic field with mean equal to β_1. Polynomials or periodic functions may be represented through the drift. For the time being it will be assumed that R is an ordinary covariance function (positive definite function); however, extension to generalized covariance function (conditionally positive definite functions) will be shown to be straightforward.

Assume that there are n measurements arranged, for the sake of notational convenience, in an $n \times 1$ $\mathbf{y}$ vector. These may be measurements at points, weighed averages over given areas, or gradients of the function. In all these cases, (1) yields the following general relation:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3}$$

where X is a known $n \times p$ matrix of the known functions of the spatial coordinates; $\boldsymbol{\beta}$ is the $p \times 1$ vector of drift coefficients; and $\boldsymbol{\varepsilon}$ is a random vector with zero mean and covariance matrix $Q_{yy}(\theta)$ which is a known function of parameter vector θ. Thus it has been assumed that the deterministic effects are linear in the unknown parameters and the measurements are linearly related to the spatial function (or functions).

In this work, attention will be limited to Gaussian processes. In view of the complexity of Bayesian analysis, Gaussian processes appear the reasonable point to start. In many cases the normality assumption is reasonable, probably after a transformation (e.g., the logarithmic transformation of transmissivity). Furthermore, many of the results developed for the Gaussian case may be given some wide sense interpretation, i.e., in terms of the first few moments.

The prediction problem may be defined as finding the probability density function of $\mathbf{y}_0$, a vector of unknown point or weighted-average values of the function, given observations $\mathbf{y}$. In the same way as with the observations, it is assumed that

$$\mathbf{y}_0 = X_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0 \tag{4}$$

where X_0 is the matrix of deterministic effects for $\mathbf{y}_0$; Q_{yy} is the covariance matrix of $\mathbf{y}$; Q_{00} is the covariance matrix of $\mathbf{y}_0$; and $Q_{oy} = Q_{yo}{}^T$, exponent T denoting transpose, is the cross-covariance matrix between $\mathbf{y}_0$ and $\mathbf{y}$. In practice, the most common procedure is to come up with point estimates, $\hat{\boldsymbol{\beta}}$ and/or $\hat{\theta}$, of the structural parameters which are then assumed perfectly known. Under this assumption the conditional distribution of $\mathbf{y}_0$ given $\mathbf{y}$ is Gaussian with mean

$$E(\mathbf{y}_0/\mathbf{y}) = X_0\hat{\boldsymbol{\beta}} + Q_{oy}Q_{yy}{}^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}) \tag{5}$$

and covariance matrix

$$V(\mathbf{y}_0/\mathbf{y}) = Q_{00} - Q_{oy}Q_{yy}{}^{-1}Q_{yo} \tag{6}$$

Throughout this work it will be assumed that Q_{yy} is nonsingular (invertible). The proof of (5) and (6) is available in textbooks [e.g., *Schweppe*, 1973, p. 521]. The conditional mean is a linear function of the measurements while the conditional covariance depends not on the observations but only on their location. Equations (5) and (6) are among the most widely used formulas in applied estimation. They are useful even when the Gaussian assumption does not hold, because for known parameters, (5) gives the minimum variance estimates of $\mathbf{y}_0$ which is a linear function of the observations. It minimizes the conditional mean-square-error matrix, given by (6). (However, in the non-Gaussian case, there may be another

estimator which is a nonlinear function of the observations which has an even smaller $V(y_0/y)$.)

These equations rely on estimates of the drift coefficients which are seldom known a priori and are difficult to obtain for near-nonstationary functions. *Matheron* [1971] has advanced a linear estimation method, known as kriging, which, among other features, does not involve drift coefficient estimates. In Appendix A the kriging estimator is derived in the vector case. In vector-matrix notation,

$$E(y_0/y) = [Q_{0y}Q_{yy}^{-1} + (X_0 - Q_{0y}Q_{yy}^{-1}X)$$
$$\cdot (X^T Q_{yy}^{-1}X)^{-1}X^T Q_{yy}^{-1}]y \qquad (7)$$

with estimation covariance matrix

$$V(y_0/y) = Q_{00} - Q_{0y}Q_{yy}^{-1}Q_{y0} + (X_0 - Q_{0y}Q_{yy}^{-1}X)$$
$$\cdot (X^T Q_{yy}^{-1}X)^{-1}(X_0 - Q_{0y}Q_{yy}^{-1}X)^T \qquad (8)$$

These estimators will be given a Bayesian interpretation in section 4.

3. OVERVIEW OF BAYESIAN ANALYSIS

The Likelihood Function

The likelihood function of the parameters β and θ given the measurements is a multivariate normal with mean $X\beta$ and covariance matrix $Q_{yy}(\theta)$.

$$p(y \mid \beta, \theta) = (2\pi)^{-n/2}|Q_{yy}|^{-1/2}$$
$$\cdot \exp\left[-\tfrac{1}{2}(y - X\beta)^T Q_{yy}^{-1}(y - X\beta)\right] \qquad (9)$$

where two vertical lines denote determinant and the dependence of Q_{yy} on θ is not shown for the sake of notational convenience.

Prior Distribution of the Parameters

Let $p'(\beta, \theta)$ describe the prior joint probability density function of β and θ. Throughout this analysis a prime will denote a prior distribution while a double prime will denote a posterior distribution. The reader may refer to *Berger* [1980] for a good discussion of methods for assessing priors. In practice, it is usual to select functional forms which facilitate the analytical treatment of the problem. The most common and sometimes most appropriate approach [*Raiffa and Schlaifer*, 1961] is to select a prior distribution which is conjugate to the likelihood function in the sense that the posterior distribution has the same form as the prior distribution. Under certain conditions, application of Bayes' theorem then becomes equivalent to the algebraic problem of updating the parameters of the probability density function (pdf) of the parameters. Of course, the main consideration in selecting $p'(\beta, \theta)$ should be that it accurately represents the prior information about the parameters. If the only source of prior information is another sample such as the one at hand, the distribution of the parameters will naturally be conjugate. Furthermore, if little prior information is available (relatively diffuse prior pdf), the exact shape of $p'(\beta, \theta)$ is not of so much importance as long as it is selected so that the posterior distribution is dominated by the likelihood function. At the limit, for completely diffuse or noninformative priors (uniform over the whole range of possible values) the analysis is totally unaffected by the assumed shape of the prior.

Updating of the Distribution of the Parameters

The posterior distribution of the parameters, determined through application of Bayes' theorem, is proportional to the product of the likelihood times the prior distribution:

$$p''(\beta, \theta) = cp(y \mid \beta, \theta)p'(\beta, \theta) \qquad (10)$$

where c is a constant determined so that the posterior is an appropriate probability density function, i.e.,

$$c = \left[\int_\beta \int_\theta p(y \mid \beta, \theta)p'(\beta, \theta) \, d\beta \, d\theta\right]^{-1} \qquad (11)$$

Note that we have used the customary and convenient notation for representing multiple integrals:

$$\int_x f(x) \, dx \doteq \int_{x_1} \int_{x_2} \cdots \int_{x_n} f(x_1, x_2, \cdots, x_n) \, dx_1 \, dx_2 \cdots dx_n \qquad (12)$$

If conjugate priors are used, calculation of the multiple integral (11) may be avoided in certain cases.

Bayesian or Compound Distribution

The pdf of y_0 given y and the parameters β and θ, $p(y_0 \mid \beta, \theta, y)$, is multivariate normal with mean and covariance function given by (5) and (6), respectively. However, in a Bayesian sense, β and θ may themselves be random variables conditional on prior information and the data, jointly distributed with y_0. The marginal distribution of y_0 given prior information and data is called the Bayesian or compound distribution,

$$\tilde{p}(y_0 \mid y) = \int_\beta \int_\theta p(y_0 \mid \beta, \theta, y)p''(\beta, \theta) \, d\beta \, d\theta \qquad (13)$$

where, in this work, the tilde denotes Bayesian probability distribution or expectation. The essence of this distribution is that it accounts for parameter uncertainty. In the conventional non-Bayesian approach some point estimates, $\hat{\beta}$ and/or $\hat{\theta}$, of the parameters are obtained and then treated as perfectly known; in the Bayesian distribution, uncertainty in the parameters is recognized and "averaged out."

It is worthwhile to note that the Bayesian distribution $\tilde{p}(y_0 \mid y)$ is not necessarily Gaussian, even if $p(y_0 \mid \beta, \theta, y)$ is Gaussian. The Bayesian distribution may be used to obtain point estimators which minimize a given loss function. Of particular interest are the conditional mean and the covariance matrix of y_0 given y (Bayesian distribution). The mean is given from the double expectation theorem:

$$\tilde{E}(y_0 \mid y) = \underset{\beta, \theta}{E} [E(y_0 \mid \beta, \theta, y)]$$
$$= \int_\beta \int_\theta [X_0\beta - Q_{0y}Q_{yy}^{-1}(y - X\beta)]p''(\beta, \theta) \, d\beta \, d\theta \qquad (14)$$

and the covariance matrix is given by another well-known result:

$$\tilde{V}(y_0 \mid y) = \underset{\beta, \theta}{E} [V(y_0 \mid \beta, \theta, y)] + \underset{\beta, \theta}{V} [E(y_0 \mid \beta, \theta, y)]$$
$$= \int_\beta \int_\theta [Q_{00} - Q_{0y}Q_{yy}^{-1}Q_{y0}]p''(\beta, \theta) \, d\beta \, d\theta$$
$$+ \int_\beta \int_\theta [X_0\beta + Q_{0y}Q_{yy}^{-1}(y - X\beta) - \tilde{E}(y_0 \mid y)]$$
$$\cdot [X_0\beta + Q_{0y}Q_{yy}^{-1}(y - X\beta)$$
$$- \tilde{E}(y_0 \mid y)]^T p''(\beta, \theta) \, d\beta \, d\theta \qquad (15)$$

Note that even in the Gaussian case the Bayesian conditional mean generally is a nonlinear function of the observations, and the conditional covariance matrix does depend on

the outcome of the measurements. Because the second term in (15) is nonnegative, the Bayesian covariance is always larger than or equal to the average covariance of the conventional point estimator; the difference is due to accounting for parameter uncertainty in the Bayesian estimator.

4. BAYESIAN ANALYSIS FOR KNOWN COVARIANCE FUNCTION PARAMETERS

The prior distribution of β is assumed Gaussian (conjugate of the likelihood, equation (9)) with mean b' and inverse of covariance matrix, to be called precision matrix, P'. If some combinations of the drift coefficients are perfectly known, the covariance matrix is singular. Even in such cases, prior information is conveniently summarized through a $p \times 1$ vector b' and a positive semidefinite matrix P'. The posterior distribution $p''(\beta)$ is Gaussian with mean vector and precision matrix which may be calculated easily from (10). (Actually, form matching would suffice since a conjugate prior has been selected.) The posterior mean is

$$b'' = (P' + X^T Q_{yy}^{-1} X)^{-1} [X^T Q_{yy}^{-1} y + P'b'] \qquad (16)$$

and the posterior precision matrix is

$$P'' = P' + X^T Q_{yy}^{-1} X \qquad (17)$$

Note that on the one extreme, that of diffuse prior ($P' = 0$), the estimator is the well-known maximum likelihood (also weighted least squares) estimator

$$b'' = (X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1} y \qquad (18)$$

$$P'' = X^T Q_{yy}^{-1} X \qquad (19)$$

On the other extreme, that of prior information dominating the evidence contained in the data in the sense that P' is much "larger" than $X^T Q_{yy}^{-1} X$,

$$b'' = b' \qquad (20)$$

$$P'' = P' \qquad (21)$$

Analytic integration of (13) is possible by making use of the properties of the Gaussian distribution. An even simpler way is to note that since $p(y_0, \beta | y)$ is Gaussian so is its marginal, the Bayesian distribution $\tilde{p}(y_0 | y)$. The mean and variance of $\tilde{p}$ can then be calculated using (14) and (15).

$$\tilde{E}(y_0 | y) = (X_0 - Q_{0y} Q_{yy}^{-1} X)(P' + X^T Q_{yy}^{-1} X)^{-1} P'b'$$

$$+ [Q_{0y} Q_{yy}^{-1} + (X_0 - Q_{0y} Q_{yy}^{-1} X)$$

$$\cdot (P' + X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1}] y \qquad (22)$$

$$\tilde{V}(y_0 | y) = (Q_{00} - Q_{0y} Q_{yy}^{-1} Q_{y0}) + (X_0 - Q_{0y} Q_{yy}^{-1} X)$$

$$\cdot (P' + X^T Q_{yy}^{-1} X)^{-1} (X_0 - Q_{0y} Q_{yy}^{-1} X)^T \qquad (23)$$

The first term in (23) is the result of the randomness of the spatial function (often referred to in the hydrologic literature as "natural" uncertainty), and the second term is the result of uncertainty in the drift coefficients.

Note that for a priori known drift coefficients the Bayesian estimator is the same with the usual linear minimum variance estimator (or Gaussian conditional mean) given by (5) and (6). On the other extreme, that of diffuse prior of the drift coefficients ($P' = 0$), the Bayesian estimator is identical to kriging (equations (7) and (8)). However, the Bayesian estimator (equations (22) and (23)) is more general and can account for partial prior information about the parameters. Furthermore, it is shown in Appendix B that without any reference to probability distributions, (22) gives the linear minimum variance esti-

mator given the observations and the prior mean and variance of the drift parameters.

5. BAYESIAN ANALYSIS FOR COVARIANCE FUNCTION KNOWN EXCEPT FOR A MULTIPLICATIVE CONSTANT

This is an extension to the case of section 4 such that each covariance matrix may be set equal to a known covariance matrix divided by an unknown parameter. For example,

$$Q_{yy} = \frac{1}{\theta} S_{yy} \qquad Q_{0y} = \frac{1}{\theta} S_{0y} \qquad Q_{00} = \frac{1}{\theta} S_{00} \qquad (24)$$

where the S matrices are known and θ is a parameter to be referred to as the relative precision. Examples include stationary random fields with the covariance function being the exponential with known integral scale or the spherical with known range; and stationary-increment fields with the polynomial generalized covariance with one term (provided, of course, that in the last example only generalized increments are involved). S_{yy} is assumed invertible. The likelihood function, equation (9), may be written in terms of sufficient statistics:

$$p(y | \beta, \theta) \propto \theta^{\mu_s/2} \exp\left[-\frac{\theta}{2} (\beta - b_s)^T H_s (\beta - b_s) \right]$$

$$\cdot \theta^{v_s/2} \exp\left[-\tfrac{1}{2} v_s q_s \theta \right] \qquad (25)$$

where

$$H_s = X^T S_{yy}^{-1} X \qquad (26a)$$

$$H_s b_s = X^T S_{yy}^{-1} y \qquad (26b)$$

$$\mu_s = \text{rank } (H_s) \qquad (26c)$$

$$v_s = n - \mu_s \qquad (26d)$$

$$q_s = (y^T S_{yy}^{-1} y - b_s^T X S_{yy}^{-1} y)/v_s \qquad (26e)$$

For a noninformative prior, the posterior distribution of β and θ is proportional to the likelihood. A clarification is needed at this point. Strictly speaking, the expression of (25) is a proper probability distribution only if H_s is nonsingular, i.e., $\mu_s = p$. This will be the case in most applications. However, sometimes the data contain enough information so that only up to $\mu_s(\mu_s < p)$ linear combinations of the drift coefficients can be estimated from the data. A case in point is estimation of the log transmissivity mean using only piezometric head data. Then the constant term in the mean cannot be estimated from the data. In such a case the first exponent should be expressed in terms of these combinations which can be estimated and which have a nonsingular H_s matrix. Nevertheless, it will be convenient to keep the notation of (25)–(26) with the understanding that they express the pdf of μ_s linear combinations of β. Furthermore, only μ_s combinations of b_s are uniquely defined from (26b).

The conjugate prior distribution is the normal-gamma 2 [see *Raiffa and Schlaifer*, 1961, chapter 13]:

$$p'(\beta, \theta) \propto \theta^{\mu'/2} \exp\left[-\frac{\theta}{2} (\beta - b')^T H'(\beta - b') \right]$$

$$\cdot \theta^{v'/2 - 1} \exp\left(-\tfrac{1}{2} v' q' \theta \right) \qquad (27)$$

where $\propto$ denotes proportional and μ' is the rank of H'. This means that the conditional distribution of β given θ is normal with mean b' and prior covariance $(H'\theta)^{-1}$; and the marginal of θ is gamma 2 with mean $1/q'$ and variance $2/v'q'^2$. The

completely noninformative case corresponds to the limiting case $H' = 0$, $\mu' = 0$, and $v' = 0$.

The posterior distribution $p''(\beta, \theta)$ would have the same form as the prior, with parameters:

$$H'' = H' + H_s \tag{28a}$$

$$H''b'' = H'b' + H_s b_s \tag{28b}$$

$$\mu'' = \text{rank } (H'') \tag{28c}$$

$$v'' = \mu' + v' + \mu_s + v_s - \mu'' \tag{28d}$$

$$q'' = \frac{v'q' + v_s q_s + (b_s - b'')^T H_s (b_s - b')}{v''} \tag{28e}$$

For an informative prior, the Bayesian distribution of y_0 given observations y may be calculated as follows. The Bayesian distribution of $\tilde{p}(y_0 | \theta, y)$ is Gaussian with mean and covariance function given by (22) and (23). The posterior marginal distribution of θ, $p''(\theta)$, is gamma 2. The Bayesian distribution of y_0 given only y,

$$\tilde{p}(y_0 | y) = \int_\theta \tilde{p}(y_0 | \theta, y) p''(\theta) \, d\theta \tag{29}$$

is multivariate Student t [see *Raiffa and Schlaifer*, 1961, p. 256] with mean

$$\tilde{E}(y_0/y) = (X_0 - S_{0y}S_{yy}^{-1}X)(H'')^{-1}H'b'$$
$$+ [S_{0y}S_{yy}^{-1} + (X_0 - S_{0y}S_{yy}^{-1}X)(H'')^{-1}X^T S_{yy}^{-1}]y \tag{30}$$

and covariance matrix

$$\tilde{V}(y_0 | y) = [S_{00} - S_{0y}S_{yy}^{-1}S_{y0} + (X_0 - S_{0y}S_{yy}^{-1}X)(H'')^{-1}$$
$$\cdot (X_0 - S_{0y}S_{yy}^{-1}X)^T]q'' \frac{v''}{v'' - 2} \tag{31}$$

Consider, for example, the common case of negligible prior information. Then the conditional mean of the Bayesian distribution is given by the kriging estimator. The covariance is given by the covariance of the kriging estimator increased by a factor of $(\eta - p)/(\eta - p - 2)$, provided that the unknown quantity $(1/\theta)$ is substituted by q_s, its minimum variance unbiased quadratic estimator [*Kitanidis*, 1985, Equation 29]. For a priori unknown θ but perfectly known β the multiplicative factor is $n/(n - 2)$.

Thus we have shown that when the covariance function is known except for a multiplicative constant, the Bayesian distribution associated with predictions is multivariate Student t with mean the same as in the known-covariance case. The estimation matrix is equal to the one given by the known-covariance case increased by a factor of $v''/(v'' - 2)$, where v'' is the number of degrees of freedom associated with the calculation of the variance; v'' takes values in accordance with (28).

6. BAYESIAN ANALYSIS USING THE POSTERIOR MARGINAL PDF OF COVARIANCE PARAMETERS

In the cases examined in sections 4 and 5, analytical solutions could be obtained by using the generally unrestrictive assumption of conjugate priors and taking advantage of the fact that the likelihood function could be described through a fixed number of sufficient statistics. In simple terms, sufficient statistics are variables which fully summarize the information contained in the data and whose number does not increase with the number of observations. When conjugate priors are used, the problem of deriving the posterior pdf becomes equivalent to an algebraic problem of updating the sufficient statistics. However, the likelihood function of parameters of most

covariance functions used in hydrology, such as the exponential, spherical, Bessel, and multiparameter polynomial, does not admit a fixed number of sufficient statistics. Thus the use of conjugate priors for θ does not necessarily reduce the computational burden of integrating the equations of section 3. No general method is available for the analytical derivation of either the posterior distribution of the parameters or the Bayesian distribution of the predicted quantities. In many cases, numerical methods must be used.

Let $p'(\beta, \theta) = p'(\beta | \theta)p(\theta)$ where the prior conditional distribution $p'(\beta | \theta)$ is assumed Gaussian with mean $b'(\theta)$ and precision matrix (inverse of the covariance matrix) $P'(\theta)$. The posterior distribution is

$$p''(\beta, \theta) = p''(\beta | \theta)p''(\theta) \propto |Q_{yy}|^{-1/2}$$
$$\exp\left[-\tfrac{1}{2}(y - X\beta)^T Q_{yy}^{-1}(y - X\beta)\right]|P'|^{1/2}$$
$$\exp\left[-\tfrac{1}{2}(\beta - b')^T P'(\beta - b')\right]p'(\theta) \tag{32}$$

It is obvious that the posterior distribution of β conditional on θ, $p''(\beta | \theta)$, is Gaussian with mean and covariance matrix given by (16) and (17) except, of course, that they are now functions of θ. The posterior marginal of θ, consisting of the remaining terms of (32), is

$$p''(\theta) \propto |Q_{yy}|^{-1/2}|P'|^{1/2}|(P' + X^T Q_{yy}^{-1}X)^{-1}|^{1/2}$$
$$\cdot \exp\left\{-\tfrac{1}{2}[y^T(Q_{yy}^{-1} - Q_{yy}^{-1}X(P' + X^T Q_{yy}^{-1}X)^{-1}X^T Q_{yy}^{-1})y\right.$$
$$- 2b'^T P'(P' + X^T Q_{yy}^{-1}X)^{-1}X^T Q_{yy}^{-1}y + b'^T P'b'$$
$$\left.- b'^T P'(P' + X^T Q_{yy}^{-1}X)^{-1}P'b']\right\}p'(\theta) \tag{33}$$

Turning our attention to the prediction problem, note that the Bayesian of y_0 can be written as

$$\tilde{p}(y_0 | y) = \int_\theta \tilde{p}(y_0 | \theta, y)p''(\theta) \, d\theta \tag{34}$$

where $\tilde{p}(y_0 | \theta, y)$ is Gaussian with mean $\tilde{E}(y_0 | \theta, y)$ and covariance $\tilde{V}(y_0 | \theta, y)$ given by (22) and (23). The Bayesian estimator may then be written as

$$\tilde{E}(y_0 | y) = \int_\theta \tilde{E}(y_0 | \theta, y)p''(\theta) \, d\theta \tag{35}$$

with covariance matrix

$$\tilde{V}(y_0 | y) = \int_\theta \{\tilde{V}(y_0 | \theta, y) + [\tilde{E}(y_0/y) - \tilde{E}(y_0 | \theta, y)]$$
$$\cdot [\tilde{E}(y_0/y) - \tilde{E}(y_0 | \theta, y)]^T\}p''(\theta) \, d\theta \tag{36}$$

Equations (33), (35), and (36) are easier to use than (10), (14), and (15) since the uncertainty in β has already been accounted for. They are useful in the approximate derivation of Bayesian estimators for some of the most common covariance functions.

The analysis of this section provides insight into the problem of suboptimal estimation of spatial functions. In most applications, parameters are first estimated, and then a linear estimator is used. A procedure common in applications of universal kriging is to assume a covariance function to estimate drift coefficients, then use these estimates to come up with a covariance function, and keep iterating between estimates of drift coefficients and covariance function parameters until some convergence criterion is met. Then linear estimation, such as kriging, is used. Another procedure [*Dagan*, 1985] is to maximize the likelihood function, equation (9), to estimate both β and θ. *Kitanidis and Vomvoris* [1983] suggested that this procedure is appropriate if good prior estimates of drift coefficients are available; if no prior information is avail-

able, *Kitanidis and Vomvoris* [1983], *Hoeksema and Kitanidis* [1985a, b], and *Kitanidis and Lane* [1985] recommended estimation of covariance parameters through maximization of the restricted likelihood, i.e., the likelihood of data increments which do not depend on drift coefficients. It will be shown that these results are particular cases of those derived here.

Since the posterior marginal $p''(\theta)$ summarizes all prior and data information about θ, it is obvious that it should be the one from which to obtain point estimates of covariance function parameters. For example, one may obtain the maximum a posteriori (MAP) estimates of θ, i.e., the values which maximize the expression of (33). These values may then be used in the linear estimator of (22) and (23) for predictions. This is a practical and rather general suboptimal procedure which can account for prior information in both drift coefficients and covariance parameters. To answer the question of how various suboptimal procedures compare with one another and with the exact Bayesian procedure, one must resort to extensive computational experiments. Such a study is beyond the scope of this paper and will be presented elsewhere.

Note that for a priori known drift coefficients, i.e., taking the limit for $P' \to I \cdot \infty$, the posterior marginal of θ tends to

$$p''(\theta) \propto |Q_{yy}|^{-1/2} \exp\left[-\tfrac{1}{2}(y - X\mathbf{b}')^T Q_{yy}^{-1}(y - X\mathbf{b}')\right] p''(\theta)$$
(37)

The marginal likelihood thus has the same form as (9) except that $\mathbf{b}'$ represents a priori known drift coefficients, not parameters estimated from the sample.

In many applications there is practically no prior information about drift coefficients. This important particular case will now be examined in detail. For $P' \to 0$, the posterior marginal of θ is

$$p''(\theta) \propto |Q_{yy}|^{-1/2} |(X^T Q_{yy}^{-1} X)^{-1}|^{1/2} \exp\left\{-\tfrac{1}{2} y^T [Q_{yy}^{-1}\right.$$
$$\left. - Q_{yy}^{-1} X (X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1}] y\right\} p''(\theta) \qquad (38)$$

One may readily verify that if an arbitrary drift were introduced in the data, i.e., y were replaced by $y + X\mathbf{b}$ where $\mathbf{b}$ is any $p \times 1$ vector representing drift coefficients, the argument of the exponent would not be affected. Consequently, $p''(\theta)$ depends only on increments of the data which are unaffected by drift coefficients. Such increments are known in the geostatistical literature as generalized or authorized increments. An important result of this analysis is that from a strictly Bayesian viewpoint, only generalized increments are relevant in the calculation of covariance function parameters when no prior information is available about the drift coefficients. This property is not limited to any particular class of covariance functions or any particular point estimator. However, when only authorized increments are used, the class of permissible covariance functions may be expanded to include conditionally positive definite functions. Such covariance functions are called generalized (see, for example, *Kitanidis* [1983]).

It is of interest to examine the relation between the marginal likelihood

$$L(\theta \mid y) \propto |Q_{yy}|^{-1/2} |(X^T Q_{yy}^{-1} X)^{-1}|^{1/2}$$
$$\cdot \exp\left\{-\tfrac{1}{2} y^T [Q_{yy}^{-1} - Q_{yy}^{-1} X (X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1}] y\right\}$$
(39)

and the likelihood of any complete set of generalized increments. Consider the transformation

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} T_1 \\ (X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1} \end{bmatrix} y = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix} y = Ty \quad (40)$$

where T_1 is any $(n - p) \times n$ matrix of rank $(n - p)$ such that $T_1 X = 0$. Consequently, $z_1 = T_1 y$ is an $(n - p)$ vector of generalized increments. Because T_1 has rank $(n - p)$, $(X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1}$ has rank p, and T_1 is orthogonal to T_2 with respect to a nonsingular matrix Q_{yy} (i.e., $T_1 Q_{yy} T_2^T = 0$), T has n linearly independent rows and consequently is nonsingular. Its inverse is

$$T^{-1} = [Q_{yy} T_1^T (T_1 Q_{yy} T_1^T)^{-1} : X] \qquad (41)$$

as one may readily verify that this expression satisfies the relation $TT^{-1} = I$. Since $T^{-1}T = I$, another useful relation follows:

$$Q_{yy}^{-1} - Q_{yy}^{-1} X (X^T Q_{yy}^{-1} X)^{-1} X Q_{yy}^{-1} = T_1^T (T_1 Q_{yy} T_1^T)^{-1} T_1$$
(42)

Also note that

$$T Q_{yy} T^T = \begin{bmatrix} T_1 Q_{yy} T_1^T & 0 \\ 0 & (X^T Q_{yy}^{-1} X)^{-1} \end{bmatrix} \qquad (43)$$

Taking determinants of both sides of (43),

$$|T|^2 |Q_{yy}| = |T_1 Q_{yy} T_1^T| |(X^T Q_{yy}^{-1} X)^{-1}| \qquad (44)$$

From (42), (43), and (44) the marginal likelihood may be written as

$$L(\theta \mid y) \propto |T|(2\pi)^{-(n-p)/2} |(T_1 Q_{yy} T_1^T)|^{-1/2}$$
$$\cdot \exp\left[-\tfrac{1}{2} z_1^T (T_1 Q_{yy} T_1^T)^{-1} z_1\right] = |T| p(z_1 | \theta) \qquad (45)$$

where $p(z_1 | \theta)$ is the likelihood of θ given the set of $(n - p)$ generalized increments. Taking the derivative of $|T|$ with respect to any covariance parameter θ_j,

$$\frac{\partial |T|}{\partial \theta_j} = |T| \operatorname{Tr}\left[T^{-1} \frac{\partial T}{\partial \theta_j}\right]$$
$$= |T| \operatorname{Tr}\left\{X \frac{\partial}{\partial \theta_j} [(X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1}]\right\} = 0 \qquad (46)$$

Thus $|T|$ is not a function of the parameters. Consequently, it has been shown that the likelihood of any complete set of generalized increments is proportional to the marginal likelihood of the covariance function parameters. This result, which appears to be original, provides additional theoretical support to methods which use generalized increments, such as restricted maximum likelihood estimation [*Hoeksema and Kitanidis*, 1985a, b; *Kitanidis*, 1985].

One may select T_1 as follows:

$$T_1 = W[I - X(X^T V X)^{-1} X^T V] \qquad (47)$$

where V is any nonsingular symmetric matrix and W is any $(n - p) \times n$ matrix of rank $n - p$. One can verify that $T_1 X = 0$ and T_1 is $(n - p) \times n$ generally of rank $n - p$. For example,

$$V = I_n \qquad (48)$$

$$W = [I_{n-p} : 0] \qquad (49)$$

where I_n is the $n \times n$ identity matrix and I_{n-p} is the $(n - p) \times (n - p)$ identity matrix. In this particular selection the meaning of W is that it eliminates the last p rows of the projection matrix $[I - X(X^T X)^{-1} X^T]$. This simple method of constructing generalized (drift parameter free) increments has been used by *Kitanidis* [1983] and *Hoeksema and Kitanidis* [1985a, b].

7. Discussion and Conclusions

The problem of uncertain parameters and its effect on inference of spatial functions has been examined in the framework

of Bayesian analysis. Parameters are treated as random variables with probability distributions reflecting what is known about them. The analysis shows how prior information about the parameters, if available, may be combined in the analysis with information contained in the sample. The estimation problem is seen as that of finding the Bayesian distribution of quantities of interest, such as point values or block averages of the function, given some measurements. The Bayesian distribution accounts not only for the randomness of the spatial function but also for parameter uncertainty.

It is appropriate at this point to recapitulate the assumptions made in this analysis and to discuss briefly how they may limit the validity of the conclusions. In this paper, attention has been limited to Gaussian processes. In many practically encountered cases this is a reasonable assumption, probably after suitable variable transformations. A case in point is the logarithm of transmissivity which, according to much evidence, is nearly normally distributed. Most of the results of the analysis would still be valid to a great extent even if the normality assumption is not strictly true; as pointed out in the text, many of them can be derived by using only the first few moments instead of invoking any distributional assumptions. However, it must be recognized that application of this analysis as well as linear minimum variance estimation theory to highly non-Gaussian data, especially those which contain "outliers," can be inappropriate. Another assumption which has been made is that the prior distribution of the drift coefficients conditional on the parameters of the covariance function is Gaussian. In most cases this assumption, which does not affect the analysis in the case of diffuse priors, should be acceptable. Usually, prior information about drift coefficients does not include details other than mean values, variances, and correlations. One case in which the normality assumption would be limiting is when there is significant prior knowledge about a drift coefficient which cannot be approximated by a Gaussian distribution, such as one that is uniform within a rather short interval. Fortunately, this would not be a case in which parameter uncertainty would be an important consideration, and approximations would be acceptable.

Finally, it has been assumed that the deterministic effects (the drift) are linear in the unknown parameters and that both the measurements and the unknown quantities are linear functions of the spatial function. Both assumptions simplify the analysis considerably. In many cases there is considerable latitude in selecting the form of the drift coefficients. Also, most of the measurements of quantities to be predicted are linear functions of the spatial function, such as point values or spatial averages. Even in cases where the linearity assumption is not strictly true, such as in the joint analysis of head and log transmissivity data, using the flow equations, the simulation studies of *Hoeksema and Kitanidis* [1985b] indicate that linearization may not introduce significant error in either parameter estimation or prediction. Nevertheless, indiscriminant use of linearization, no matter how large the variation or how nonlinear the problem, is certainly not encouraged.

First, the case of known covariance parameters and partially unknown drift coefficients was examined. The Bayesian distribution is Gaussian with mean and covariance matrix of estimation error which can be computed easily from formulas developed in this paper. The mean is a linear function of the observations and the prior mean of the drift coefficients while the covariance matrix does not depend on the outcome of the observations. Independent of distributional assumptions, the estimator is shown to be the linear minimum mean-squared-error estimator which uses both prior and sample information. The Bayesian linear estimator is at least as general as any

other linear estimator now available. For a priori known drift coefficients the estimator is shown to reduce to conventional linear minimum variance or Gaussian conditional mean estimator. However, if no prior information is available about drift coefficients, the Bayesian estimator is shown to reduce to the technique widely known as kriging (or cokriging, if more than one spatial function is involved). It is probably the first time that kriging is explicitly recognized as a special case of Bayesian estimation. According to this analysis, if negligible prior information is available about the drift coefficients, the correct procedure to use is kriging. If conventional linear estimation is used instead, the actual estimation error is, on the average, higher than the estimation error calculated by the estimation procedure. Kriging, on the other hand, is suboptimal when prior information about drift coefficients is available. For example, in the analysis of orographic precipitation data, some information about the effect of elevation or exposure is often available from the analysis of previous storms in the same basin or from regionalized studies. The Bayesian estimation procedure developed in this paper combines prior and sample information in an optimal way and provides measures of estimation accuracy which take into account parameter uncertainty.

The case of both drift and covariance function parameters being (at least partially) unknown also makes use of the analytical results of the previous case. The case of covariance function known except for a multiplicative constant, $1/\theta$, is first examined, and analytical results are obtained. As one might have expected, the Bayesian distribution of predictions given prior and sample information is multivariate Student t with the same mean as the Bayesian linear estimator. The covariance matrix is proportional to that of the Bayesian linear estimator. The unknown covariance parameter is estimated as a weighted sum of prior estimates and a quadratic function of the data and prior estimates of the drift coefficients. (For no prior information, the coefficient of proportionality is given by a minimum variance unbiased quadratic estimator.) Furthermore, the covariance matrix is shown to be magnified by a factor $v''/v'' - 2$, where v'' is the number of degrees of freedom available for the estimation of the covariance parameter. This magnification is the result of uncertainty in the covariance parameter. In the worst case, that of no prior information about either drift coefficients or the covariance parameter, v'' is equal to the number of measurements minus the number of unknown drift coefficients which affect the data. It increases as the accuracy of prior estimates of the parameters improves.

For the most commonly used covariance function models, such as the exponential with unknown variance and integral scale, analytical results cannot be obtained. The method of analysis suggested in this paper consists of analytical derivation of the Bayesian mean and variance of predictions and of the posterior of the drift coefficients, if appropriate, given the covariance parameters θ, and then numerical integration using the posterior marginal probability function of θ. The posterior marginal of θ, which summarizes all prior and sample information about the covariance parameters, is proportional to a marginal likelihood times the prior marginal of θ. The marginal likelihood does not involve the unknown drift coefficients but only their appropriately weighted prior estimates and the data.

For a priori known drift coefficients the marginal likelihood has the same form as the ordinary likelihood function, except that the drift coefficients are not parameters to be estimated from data but known constants. However, at the other extreme, that of negligible prior information about the drift coef-

ficients, the marginal likelihood is proportional to the likelihood of any complete set of generalized increments (i.e., the maximum number of linearly independent increments which are not affected by unknown drift coefficients). It has thus been shown that for a priori unknown drift coefficients, only generalized increments of the data are relevant in the calculation of covariance parameters in the prediction problem.

APPENDIX A: DERIVATION OF KRIGING EQUATION IN MATRIX-VECTOR FORM

The problem is to determine an estimator which is a linear function of the data, i.e.,

$$\hat{y}_0 = \Lambda y \tag{A1}$$

where Λ is the $m \times n$ matrix of weights. The estimation error is

$$e = \hat{y}_0 - y_0 = \Lambda y - y_0 \tag{A2}$$

The matrix of the weights is selected according to the following specifications:

1. The estimator is unbiased in the sense that

$$E(e) = \Lambda X \beta - X_0 \beta = 0 \tag{A3}$$

The only way this equation can hold for any value of β is that

$$\Lambda X - X_0 = 0 \tag{A4}$$

2. The matrix of the mean squared error

$$E(ee^T) = \Lambda Q_{yy} \Lambda^T - \Lambda Q_{y0} - Q_{0y} \Lambda^T + Q_{00} \tag{A5}$$

is minimized in some well-defined sense.
The problem of determining Λ may then be defined as

$$\min \mathrm{Tr} \, [\Lambda Q_{yy} \Lambda^T - \Lambda Q_{y0} - Q_{0y} \Lambda^T + Q_{00}] \tag{A6}$$

subject to

$$\Lambda X - X_0 = 0 \tag{A7}$$

The Lagrangian corresponding to the constrained problem is

$$\mathrm{Tr} \, [\Lambda Q_{yy} \Lambda^T - \Lambda Q_{y0} - Q_{0y} \Lambda^T + Q_{00}] - \mathrm{Tr} \, [(\Lambda X - X_0)M] \tag{A8}$$

where M is a $p \times m$ matrix of Lagrange multipliers. Taking derivatives [see *Schweppe*, 1973, p. 509] with respect to Λ and M yields the following matrix equations:

$$2Q_{yy} \Lambda^T - 2Q_{y0} - XM = 0 \tag{A9}$$

$$\Lambda X - X_0 = 0 \tag{A10}$$

The solution to this linear system is

$$M = 2(X^T Q_{yy}^{-1} X)^{-1} X_0^T - 2(X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1} Q_{y0} \tag{A11}$$

and

$$\begin{aligned} \Lambda &= Q_{0y} Q_{yy}^{-1} + \tfrac{1}{2} M^T X^T Q_{yy}^{-1} \\ &= Q_{0y} Q_{yy}^{-1} + (X_0 - Q_{0y} Q_{yy}^{-1} X)(X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1} \end{aligned} \tag{A12}$$

and the corresponding covariance matrix

$$\begin{aligned} E(ee^T) &= Q_{00} - Q_{0y} Q_{yy}^{-1} Q_{y0} + \tfrac{1}{4} M^T (X^T Q_{yy}^{-1} X) M \\ &= Q_{00} - Q_{0y} Q_{yy}^{-1} Q_{y0} + (X_0 - Q_{0y} Q_{yy}^{-1} X) \\ &\quad \cdot (X^T Q_{yy}^{-1} X)^{-1} (X_0 - Q_{0y} Q_{yy}^{-1} X)^T \end{aligned} \tag{A13}$$

APPENDIX B: DISTRIBUTION-FREE DERIVATION OF LINEAR BAYESIAN ESTIMATOR

An estimator $\hat{y}_0$ is sought so that it is linear in the prior estimates of the drift parameters and the measurements

$$\hat{y}_0 = Ab' + By \tag{B1}$$

where A is $m \times p$ and B is $m \times n$. The prior unbiased estimate of the parameters is b' with estimation matrix V_b. Matrices A and B are to be determined according to the following specifications:

1. The estimation error

$$e = \hat{y}_0 - y_0 \tag{B2}$$

must have average of zero

$$E(e) = A\beta + BX\beta - X_0 \beta = 0 \tag{B3}$$

which yields the vector equation (corresponds to $m \times p$ scalar equations):

$$A + BX - X_0 = 0 \tag{B4}$$

2. The estimation error must have minimum variance

$$E(ee^T) = A V_b A^T + B Q_{yy} B^T - B Q_{y0} - Q_{0y} B^T + Q_{00} \tag{B5}$$

The values of A and B may thus be calculated by minimizing the trace of the covariance matrix subject to the unbiasedness constraint. The Lagrangian of this constrained optimization problem is

$$\begin{aligned} L &= \mathrm{Tr} \, (A V_b A^T + B Q_{yy} B^T - B Q_{y0} - Q_{0y} B^T + Q_{00}) \\ &\quad - \mathrm{Tr} \, [M(A + BX - X_0)] \end{aligned} \tag{B6}$$

where M is a $p \times m$ matrix of Lagrange multiples. Taking derivatives with respect to A, B, and M yields the system

$$2V_b A^T - M = 0 \tag{B7}$$

$$2Q_{yy} B^T - 2Q_{y0} - XM = 0 \tag{B8}$$

$$A + BX - X_0 = 0 \tag{B9}$$

The solution is

$$A = (X_0 - Q_{0y} Q_{yy}^{-1} X)(V_b^{-1} + X^T Q_{yy}^{-1} X)^{-1} V_b^{-1} \tag{B10}$$

$$\begin{aligned} B &= Q_{0y} Q_{yy}^{-1} + (X_0 - Q_{0y} Q_{yy}^{-1} X) \\ &\quad \cdot (V_b^{-1} + X^T Q_{yy}^{-1} X)^{-1} X^T Q_{yy}^{-1} \end{aligned} \tag{B11}$$

$$M = 2(V_b^{-1} + X^T Q_{yy}^{-1} X)^{-1}(X_0^T - X^T Q_{yy}^{-1} Q_{y0}) \tag{B12}$$

The covariance matrix is

$$E(ee^T) = Q_{00} - Q_{0y} Q_{yy}^{-1} Q_{y0} + \tfrac{1}{4} M^T (V_b^{-1} + X^T Q_{yy}^{-1} X) M \tag{B13}$$

REFERENCES

Berger, J. O., *Statistical Decision Theory: Foundations, Concepts and Methods*, Springer-Verlag, New York, 1980.

Chua, S. H., and R. L. Bras, Optimal estimators of mean areal precipitation in regions of orographic influence, *J. Hydrol.*, *57*(1/2), 23–48, 1982.

Dagan, G., Solute transport in heterogeneous porous formations, *J. Fluid Mech.*, *145*, 151–177, 1984.

Dagan, G., Stochastic modeling of groundwater flow by un-

conditional and conditional probabilities: The inverse problem, *Water Resour. Res.*, *21*(1), 65–72, 1985.

Delhomme, J. P., Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach, *Water Resour. Res.*, *15*(2), 269–280, 1979.

Geisser, S., Bayesian estimation in multivariate analysis, *Am. Math Soc.*, *36*, 150–159, 1985.

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

Halpern, E. F., Polynomial regression from a Bayesian approach, *J. Am. Stat. Assoc.*, *68*(341), 137–143, 1973.

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, 1985a.

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

Hughes, J. P., and D. P. Lettenmaier, Data requirements for kriging: Estimation and network design, *Water Resour. Res.*, *17*(6), 1641–1650, 1981.

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

Kitanidis, P. K., Minimum-variance unbiased quadratic estimation of

covariances of regionalized variables, *J. Int. Assoc. Math. Geol.*, *17*(2), 195–208, 1985.

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

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

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

Raiffa, H., and R. Schlaifer, *Applied Statistical Decision Theory*, MIT Press, Cambridge, Mass., 1961.

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

Smith, L., and R. A. Freeze, Stochastic analysis of steady state groundwater flow in a bounded domain, 2, Two-dimensional simulations, *Water Resour. Res.*, *15*(6), 1543–1559, 1979.

Valdés, J. B., I. Rodriguez-Iturbe, and G. J. Vicens, Bayesian generation of synthetic streamflows, 2, The multivariate case, *Water Resour. Res.*, *13*(2), 291–295, 1977.

P. K. Kitanidis, St. Anthony Falls Hydraulic Laboratory, Department of Civil and Mineral Engineering, University of Minnesota, Mississippi River at 3rd Avenue, S. E., Minneapolis, MN 55414.