

UC Santa Barbara

UC Santa Barbara Previously Published Works

Title

Linear spatial interpolation: Analysis with an application to San Joaquin Valley

Permalink

<https://escholarship.org/uc/item/5bz1228j>

Journal

Stochastic Environmental Research and Risk Assessment, 2(2)

ISSN

1436-3240

Authors

Loaiciga, HA

Shumway, RH

Yeh, WW-G

Publication Date

1988-06-01

DOI

10.1007/bf01543455

Peer reviewed

Linear spatial interpolation: Analysis with an application to San Joaquin Valley

H. A. Loaiciga

Dept. of Geological Science, Wright State University, Dayton, OH 45435, USA

R. H. Shumway

Division of Statistics, University of California, Davis, CA 95616, USA

W. W.-G. Yeh

Dept. of Civil Engineering, University of California, Los Angeles, CA 90024, USA

Abstract: The properties of linear spatial interpolators of single realizations and trend components of regionalized variables are examined in this work. In the case of the single realization estimator explicit and exact expressions for the weighting vector and the variances of estimator and estimation error were obtained from a closed-form expression for the inverse of the Lagrangian matrix. The properties of the trend estimator followed directly from the Gauss-Markoff theorem. It was shown that the single realization estimator can be decomposed into two mutually orthogonal random functions of the data, one of which is the trend estimator. The implementation of linear spatial estimation was illustrated with three different methods, i.e., full information maximum likelihood (FIML), restricted maximum likelihood (RML), and Rao's minimum norm invariant quadratic unbiased estimation (MINQUE) for the single realization case and via generalized least squares (GLS) for the trend. The case study involved large correlation length-scale in the covariance of specific yield producing a nested covariance structure that was nearly positive semidefinite. The sensitivity of model parameters, i.e., drift and variance components (local and structured) to the correlation length-scale, choice of covariance model (i.e., exponential and spherical), and estimation method was examined. The same type of sensitivity analysis was conducted for the spatial interpolators. It is interesting that for this case study, characterized by a large correlation length-scale of about 50 mi (80 km), both parameter estimates and linear spatial interpolators were rather insensitive to the choice of covariance model and estimation method within the range of credible values obtained for the correlation length-scale, i.e., 40–60 mi (64–96 km), with alternative estimates falling within $\pm 5\%$ of each other.

Key words: spatial statistics, hydrogeologic mapping, trend analysis, groundwater hydrology.

1 Introduction

Trend-surface analysis is a method for separating map or spatial data into two components, i.e., that of a regional nature and local variations. Such data decomposition is reflected in the geological terminology by reference to terms such as “regional dip” or “regional trends” as opposed to “local structures” or “local anomalies”. The classical statistical analysis of spatial data by the trend-surface method consists of fitting a parametric function to the trend or regional component by the least squares method. The fitted trend, i.e., usually a polynomial on the geographic coordinates describing location, is subsequently used in prediction of the trend at locations where there are no available data and for statistical inference on the predictions or estimated parameters. A closely related approach to trend-

surface analysis is a kriging method where we do not differentiate as of yet among its different variations. In its simplest expression kriging is a form of trend-surface analysis in the sense that it is commonly based on the decomposition of spatial data into a trend and a local component. However, the derivation of linear spatial interpolators such as kriging and related statistical properties are conceptually different to those characterizing the classical trend-surface analysis. The kriging estimator interpolates the regionalized or spatial realization itself (i.e., trend plus local fluctuation) as opposed to its mean and has properties distinct to the trend estimator. A review of trend-surface and kriging spatial interpolation is presented in this work, where a theoretical analysis of the two methods yields analytical expressions that clearly specify the similarities and differences between them. In order to contrast the two approaches we have mapped specific yields in the San Joaquin Valley, California (United States), and in the process, the issues of parameter estimation, linear interpolation and statistical inference are addressed in detail. The Full Information Maximum Likelihood (FIML), Restricted Maximum Likelihood (RML), and Minimum Norm Invariant Quadratic Unbiased Estimator (MINQUE) methods are used in parameter estimation for the purpose of comparison. Important implementation issues such as sensitivities to the choice of spatial covariance model and correlation length-scale and advantages/disadvantages of competing estimation methods are discussed in this study.

The approach to linear interpolation is developed using the linear statistical multivariate model throughout this work. Within this framework, important results such as the Gauss-Markoff theorem, the Cramer-Rao lower bound theorem and orthogonality considerations are readily available for the analysis. The reader is referred to the work by Rao (1973) as a foundational exposition of the general linear model theory, and to Agterberg (1974), Journel and Huijbregts (1978), Davis (1986), and de Marsily (1986), among others, for discussions on the theory and practice of spatial statistics.

2 The linear model and background theory

2.1 General definitions

Suppose that observations $z(\mathbf{x}_1), z(\mathbf{x}_2), \dots, z(\mathbf{x}_n)$ of a regionalized variable $z(\mathbf{x})$ are available at location described by Cartesian vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ where $\mathbf{x}_i$ denotes position in the space of coordinates x, y , and z . In general, $z(\mathbf{x}_i)$ could represent an averaged quantity, i.e.,

$$z(\mathbf{x}_i) = \frac{1}{V} \int_V z(\mathbf{u}) d\mathbf{u} \quad (1)$$

where V is an averaging domain (i.e., volume, area or line) centered at $\mathbf{x}_i$, or $z(\mathbf{x}_i)$ may be a point value of the regionalized variable that could represent aquifer properties such as hydraulic conductivity, specific storativity, etc.

The regionalized variable $z(\mathbf{x})$ is modeled in this study as the sum of a deterministic trend or drift, $m(\mathbf{x})$, and a random component $\xi(\mathbf{x})$, i.e.,

$$z(\mathbf{x}) = m(\mathbf{x}) + \xi(\mathbf{x}) \quad (2)$$

in which $m(\mathbf{x})$ is represented by the model

$$m(\mathbf{x}) = \sum_{r=0}^k b_r f_r(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}. \quad (3)$$

The coefficients b_r are constant but unknown and the functions $f_r(\mathbf{x})$ are in terms

of the coordinates of the location vector $\mathbf{x}$. The coefficients b_r and functions $f_r(\mathbf{x})$ are collected in vector form in the vectors $\boldsymbol{\beta}$ and $\mathbf{f}(\mathbf{x})$ respectively (the superscript T represents the transpose). The random component $\xi(\mathbf{x})$ is assumed second order (read also weakly) stationary, i.e., it has zero expected value and its covariance is assumed to be a function of the separation distance between any two points regardless of location and orientation in space. In particular, we adopt the nested covariance model of the form,

$$C(r_{ij}) = E[\xi(\mathbf{x}_i)\xi(\mathbf{x}_j)] \\ = E[(z(\mathbf{x}_i) - m(\mathbf{x}_i))(z(\mathbf{x}_j) - m(\mathbf{x}_j))] = \sum_{r=1}^p \sigma_r^2 g_r(r_{ij}) \quad (4)$$

in which the functions $g_r(r_{ij})$ are known (see Section 5) and depend on the separation distance $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$. The unknown parameters σ_r^2 represent variabilities at different scales (see, e.g., Journel and Huijbregts, 1978, p. 149). As an example, covariance models similar to Eq. (4) have been treated in depth by Matheron (1973), and Delfiner (1976).

In the event that $z(\mathbf{x}_i)$ represents an averaged quantity (see Eq. (1)), the covariance of such an averaged variable is readily expressed by integrating the point covariance Eq. (4) i.e.,

$$C(r_{ij}) = \frac{1}{V_i V_j} \int_{V_i} d\mathbf{v} \int_{V_j} d\mathbf{u} E[\xi(\mathbf{v})\xi(\mathbf{u})] \quad (5)$$

in which V_i and V_j are averaging domains centered at $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively. For the sake of clarity no distinction is made between point and averaged variables in the sequel unless it becomes indispensable for expositional purposes.

In the model described by Eqs. (2)-(4) the parameter set is given by $b_r, r = 1, 2, \dots, k$ and $\sigma_r^2, r = 1, 2, \dots, p$. Writing the n observations $z(\mathbf{x}_1), z(\mathbf{x}_2), \dots, z(\mathbf{x}_n)$ as the $n \times 1$ vector $\mathbf{z}^T = (z(\mathbf{x}_1), \dots, z(\mathbf{x}_n))$, the model is expressed in compact notation as

$$\mathbf{z} = F\boldsymbol{\beta} + \boldsymbol{\xi} \quad (6)$$

in which F is an $n \times k$ matrix whose j th row is given by $\mathbf{f}^T(\mathbf{x}_j)$ (see Eq. (3)), and $\boldsymbol{\xi}$ is an $n \times 1$ random vector. The covariance of $\boldsymbol{\xi}$ is written as follows,

$$C(\boldsymbol{\xi}) = C(\mathbf{z}) = \sum_{r=1}^p \sigma_r^2 Q_r \quad (7)$$

where Q_r is an $n \times n$ matrix whose ij th element is equal to $g_r(r_{ij})$ (see Eq. (4)). Equations (6)-(7) represent a model linear on the parameters $\boldsymbol{\beta}$ with a nested covariance. For the developments to follow (see subsection 4.3), it is convenient to rewrite Eqs. (6)-(7) in an equivalent form, i.e.,

$$\mathbf{z} = F\boldsymbol{\beta} + T_1 \boldsymbol{\xi}_1 + \dots + T_p \boldsymbol{\xi}_p \quad (8)$$

in which the $\boldsymbol{\xi}_i$'s are zero mean, independent random components with covariances equal to $\sigma_r^2 I_{nn}$ where I_{nn} is the $n \times n$ identity matrix and T_r represents the factor of Q_r , i.e., $Q_r = T_r T_r^T$. Since it is assumed that the Q_r 's are positive definite, T_r is readily available, say from the Choleski decomposition (see, e.g., Golub and van Loan 1983).

One of the main objectives of this study in the interpolation of the regionalized

variable $z(\mathbf{x})$ at specified points. Due to theoretical reasons to be indicated below it is assumed in the remainder of this section and in Section 3 that the drift parameter $\boldsymbol{\beta}$ and the variance components σ_r^2 , $r = 1, 2, \dots, p$, are known. This assumption is dropped in Section 4 where the statistical estimation of those parameters is addressed in at length.

2.2 Mean vs. single realization estimation

For the purpose of spatial interpolation, there are two fundamentally different approaches. First, suppose that the value of $z(\mathbf{x})$ at a location $\mathbf{x}_0$, $z(\mathbf{x}_0)$ is to be estimated. $z(\mathbf{x}_0)$ is viewed as a single realization of the ensemble at $\mathbf{x}_0$ and then the linear interpolator.

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i z(\mathbf{x}_i) = \boldsymbol{\lambda}^T \mathbf{z} \quad (9)$$

is proposed as an estimator of the single realization $z(\mathbf{x}_0)$. Imposing the unbiasedness condition

$$E[\hat{z}(\mathbf{x}_0)] = m(\mathbf{x}_0) = \mathbf{f}^T(\mathbf{x}_0)\boldsymbol{\beta} \quad (10)$$

implies that

$$\mathbf{F}^T \boldsymbol{\lambda} = \mathbf{f}(\mathbf{x}_0). \quad (11)$$

Requiring that $\hat{z}(\mathbf{x}_0)$ have minimum variance of estimation error $z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)$ amongst all linear unbiased interpolators of $z(\mathbf{x}_0)$ (i.e., satisfying Eq. (11), is tantamount to minimizing,

$$\text{Var}[z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)] = \sigma^2 - 2\boldsymbol{\lambda}^T \boldsymbol{\sigma}_0 + \boldsymbol{\lambda}^T \mathcal{Q} \boldsymbol{\lambda} \quad (12)$$

in which

$$\sigma^2 = \text{Var}[z(\mathbf{x}_0)], \quad (13)$$

$$\boldsymbol{\sigma}_0 = \text{Cov}[z(\mathbf{x}_0), \mathbf{z}], \quad (14)$$

$$\mathcal{Q} = \sum_{r=1}^p \sigma_r^2 \mathcal{Q}_r. \quad (15)$$

Equations (9), (11) and (12) define the so-called universal kriging estimator. The weighting vector $\boldsymbol{\lambda}$ is obtained from minimizing Eq. (12) subject to Eq. (11) leading to the Lagrangian linear system

$$\begin{matrix} n & k \\ \begin{bmatrix} \mathcal{Q} & \mathbf{F} \\ \mathbf{F}^T & 0 \end{bmatrix} \end{matrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma}_0 \\ \mathbf{f}(\mathbf{x}_0) \end{bmatrix} \quad (16)$$

in which $\boldsymbol{\mu}$ is a $k \times 1$ vector of Lagrange multipliers. Upon solution of Eq. (16) for $\boldsymbol{\lambda}$, the estimator $\hat{z}(\mathbf{x}_0)$ is obtained from Eq. (9) and the variance of $\hat{z}(\mathbf{x}_0)$ is

$$\text{Var}[\hat{z}(\mathbf{x}_0)] = \boldsymbol{\lambda}^T \mathcal{Q} \boldsymbol{\lambda}. \quad (17)$$

A second approach in linear interpolation is to estimate the mean $m(\mathbf{x}_0)$ of $z(\mathbf{x}_0)$ by the expression

$$\hat{m}(\mathbf{x}_0) = \mathbf{f}^T(\mathbf{x}_0)\hat{\boldsymbol{\beta}} \quad (18)$$

where $\hat{\boldsymbol{\beta}}$ is the estimate of the drift parameter vector (see Section 3). The variance of $\hat{m}(\mathbf{x}_0)$ is

$$\text{Var}[\hat{m}(\mathbf{x}_0)] = \mathbf{f}^T(\mathbf{x}_0)\boldsymbol{\Sigma}_{\beta}\mathbf{f}(\mathbf{x}_0) \quad (19)$$

where the covariance of $\hat{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\beta}$, will be given in Section 3.

A discussion of the properties of the estimators $\hat{z}(\mathbf{x}_0)$ and $\hat{m}(\mathbf{x}_0)$ follows in the next section.

3 Properties of $\hat{z}(\mathbf{x}_0)$ and $\hat{m}(\mathbf{x}_0)$

3.1 Mean and variance

It is convenient to find an analytical expression for the inverse of the Lagrangian matrix on the left-hand side of Eq. (16) in order to express $\boldsymbol{\lambda}$ as follows,

$$\boldsymbol{\lambda} = A_{11}\boldsymbol{\sigma}_0 + A_{12}\mathbf{f}(\mathbf{x}_0) \quad (20)$$

and

$$\boldsymbol{\mu} = -A_{22}\mathbf{f}(\mathbf{x}_0) - A_{12}^T\boldsymbol{\sigma}_0 \quad (21)$$

in which

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} = \begin{bmatrix} Q & F \\ F^T & 0 \end{bmatrix}^{-1} \quad (22)$$

where

$$A_{11} = Q^{-1} - Q^{-1}F(F^TQ^{-1}F)^{-1}F^TQ^{-1} \quad (23)$$

$$A_{12} = Q^{-1}F(F^TQ^{-1}F)^{-1} \quad (24)$$

$$A_{22} = -(F^TQ^{-1}F)^{-1} \quad (25)$$

and Eq. (22) is easily verified by straight matrix multiplication. Since F has full column rank and Q is positive definite A_{11} is also positive definite. From Eq. (20) it follows that

$$\text{Var}[\hat{z}(\mathbf{x}_0)] = \text{Var}[\boldsymbol{\lambda}^T\mathbf{z}] = \boldsymbol{\lambda}^TQ\boldsymbol{\lambda} = \boldsymbol{\sigma}_0^TA_{11}\boldsymbol{\sigma}_0 + \mathbf{f}^T(\mathbf{x}_0)(F^TQ^{-1}F)^{-1}\mathbf{f}(\mathbf{x}_0) \quad (26)$$

and the variance of the estimation error $z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)$ is obtained by substitution of Eq. (20) into Eq. (12) to yield

$$\text{Var}[z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)] = \mathbf{f}^T(\mathbf{x}_0)(F^TQ^{-1}F)^{-1}\mathbf{f}(\mathbf{x}_0) + \sigma^2 - 2\mathbf{p}^T\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_0^TA_{11}\boldsymbol{\sigma}_0 \quad (27)$$

in which

$$\mathbf{p} = A_{12}\mathbf{f}(\mathbf{x}_0) = Q^{-1}F(F^TQ^{-1}F)^{-1}\mathbf{f}(\mathbf{x}_0). \quad (28)$$

Notice that Eq. (27) represents the minimum variance of estimation error, a feature of the linear interpolator (9). For example, suppose that one wishes to estimate $z(\mathbf{x}_0)$ by the mean estimator $\hat{m}(\mathbf{x}_0)$ as defined in Eq. (18). Thus, the variance of estimation error $z(\mathbf{x}_0) - \hat{m}(\mathbf{x}_0)$ necessarily exceeds the variance in Eq. (27).

Table 1. Properties of estimators

Property	Estimator	
	$\hat{z}(\mathbf{x}_0)$	$\hat{m}(\mathbf{x}_0)$
Equation	$\lambda^T \mathbf{z}$	$\mathbf{p}^T \mathbf{z}$
Estimated variable	$z(\mathbf{x}_0)$	$m(\mathbf{x}_0)$
Expected value	$m(\mathbf{x}_0)$	$m(\mathbf{x}_0)$
Variance	$\mathbf{f}(\mathbf{x}_0)^T (F^T Q^{-1} F)^{-1} \mathbf{f}(\mathbf{x}_0) + \sigma_0^T A_{11} \sigma_0$	$\mathbf{f}(\mathbf{x}_0)^T (F^T Q^{-1} F)^{-1} \mathbf{f}(\mathbf{x}_0)$
Variance of estimation error	$\mathbf{f}(\mathbf{x}_0)^T (F^T Q^{-1} F)^{-1} \mathbf{f}(\mathbf{x}_0) + \sigma^2 - 2\mathbf{p}^T \sigma_0 - \sigma_0^T A_{11} \sigma_0$	$\mathbf{f}(\mathbf{x}_0)^T (F^T Q^{-1} F)^{-1} \mathbf{f}(\mathbf{x}_0) + \sigma^2 - 2\mathbf{p}^T \sigma_0$

To show this, first notice that the generalized least squares (GLS) estimator of the drift parameter vector β is

$$\hat{\beta} = (F^T Q^{-1} F)^{-1} F^T Q^{-1} \mathbf{z} \quad (29)$$

and has covariance

$$\Sigma_{\beta} = (F^T Q^{-1} F)^{-1} \quad (30)$$

also, from Eqs. (18) and (29), it follows that

$$\hat{m}(\mathbf{x}_0) = \mathbf{f}^T(\mathbf{x}_0) \hat{\beta} = \mathbf{p}^T \mathbf{z} \quad (31)$$

where $\mathbf{p}$ has been defined in Eq. (28). Equation (31) shows that $\hat{m}(\mathbf{x}_0)$ is linear on $\mathbf{z}$, and incidentally,

$$E[\hat{m}(\mathbf{x}_0)] = \mathbf{p}^T E\mathbf{z} = \mathbf{f}^T(\mathbf{x}_0) \beta = m(\mathbf{x}_0) \quad (32)$$

i.e., $\hat{m}(\mathbf{x}_0)$ is an unbiased estimator of the mean. Using Eqs. (19), (30), and (31) yields

$$\text{Var}[z(\mathbf{x}_0) - \hat{m}(\mathbf{x}_0)] = \mathbf{f}(\mathbf{x}_0)^T (F^T Q^{-1} F)^{-1} \mathbf{f}(\mathbf{x}_0) + \sigma^2 - 2\mathbf{p}^T \sigma_0 \quad (33)$$

that exceeds the variance in Eq. (27) by the amount $\sigma_0^T A_{11} \sigma_0$, as expected, since the estimator in Eq. (9) has the minimum variance of estimation error amongst all linear (and unbiased) estimators of $z(\mathbf{x}_0)$.

Interestingly, the linear, unbiased estimator $\hat{m}(\mathbf{x}_0)$ has variance

$$\text{Var}[\hat{m}(\mathbf{x}_0)] = \mathbf{f}^T(\mathbf{x}_0) [F^T Q^{-1} F]^{-1} \mathbf{f}(\mathbf{x}_0) \quad (34)$$

a direct result from Eqs. (19) and (30). From the Gauss-Markoff theorem (see, e.g., Rao 1973, p. 221), $\hat{m}(\mathbf{x}_0)$ has the minimum variance amongst all linear unbiased estimators of the mean $m(\mathbf{x}_0)$. A direct comparison of Eqs. (26) and (34) indicates that the variance of $\hat{z}(\mathbf{x}_0)$ exceeds the variance of $\hat{m}(\mathbf{x}_0)$ by the amount $\sigma_0^T A_{11} \sigma_0$ as expected from the Gauss-Markoff result. Table 1 summarizes the properties of $\hat{m}(\mathbf{x}_0)$ and $\hat{z}(\mathbf{x}_0)$.

Under normality of the regionalized variable $z(\mathbf{x}_0)$, $100(1 - \alpha)\%$ confidence intervals for the single realization $z(\mathbf{x}_0)$ and the mean $m(\mathbf{x}_0)$ are given respectively by

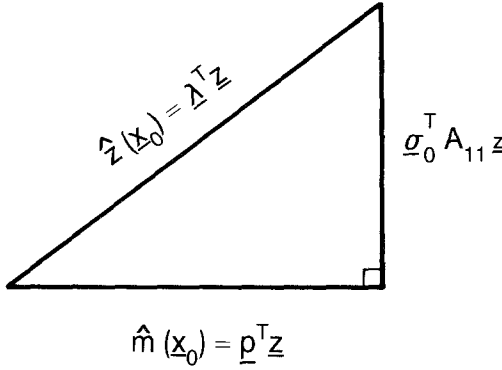


Figure 1

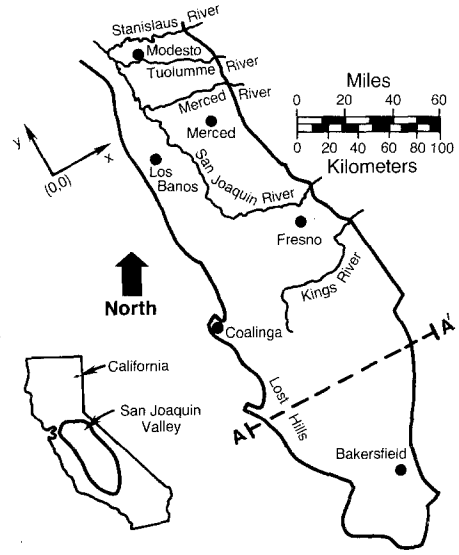


Figure 2

Figure 1. Orthogonal decomposition of the single realization estimator**Figure 2.** The San Joaquin Valley and study area

$$\hat{z}(\mathbf{x}_0) \pm z_{\alpha/2} \text{Var}[z(\mathbf{x}_0) - \hat{z}(\mathbf{x}_0)]^{1/2} \quad (35)$$

and

$$\hat{m}(\mathbf{x}_0) \pm z_{\alpha/2} \text{Var}[\hat{m}(\mathbf{x}_0)]^{1/2} \quad (36)$$

in which $z_{\alpha/2}$ is such that $P(Z \geq z_{\alpha/2}) = \alpha/2$ where Z is a standard normal variable with zero mean and unit variance; the variances in Eqs. (35)-(36) are given by Eqs. (27) and (34) respectively.

3.2 An orthogonal decomposition

There is a geometric interpretation of the relationship between the estimators $\hat{z}(\mathbf{x}_0)$ and $\hat{m}(\mathbf{x}_0)$. From Eqs. (20) and (28),

$$\boldsymbol{\lambda} = A_{11} \boldsymbol{\sigma}_0 + \mathbf{p} \quad (37)$$

from which

$$\boldsymbol{\lambda}^T \mathbf{z} = \hat{z}(\mathbf{x}_0) = \boldsymbol{\sigma}_0^T A_{11} \mathbf{z} + \mathbf{p}^T \mathbf{z} = \boldsymbol{\sigma}_0^T A_{11} \mathbf{z} + \hat{m}(\mathbf{x}_0) \quad (38)$$

where Eq. (31) was used in establishing Eq. (38). Thus $\hat{z}(\mathbf{x}_0)$ may be decomposed into two linear combinations of $\mathbf{z}$ that are orthogonal to each other, i.e.,

$$E(\boldsymbol{\sigma}_0^T A_{11} \mathbf{z} \cdot \mathbf{p}^T \mathbf{z}) = 0 \quad (39)$$

which follows from the fact that $A_{11} Q \mathbf{p} = \mathbf{0}$ and $A_{11} F = 0$. Figure 1 illustrates

graphically the relationship between the estimators $\hat{z}(\mathbf{x}_0)$ and $\hat{m}(\mathbf{x}_0)$. Notice that $E[\boldsymbol{\sigma}_0^T A_{11} \mathbf{z}] = 0$ so that the estimator $\hat{z}(\mathbf{x}_0)$ may be decomposed into the sum of $\hat{m}(\mathbf{x}_0)$ and a variable orthogonal to it, namely $\boldsymbol{\sigma}_0^T A_{11} \mathbf{z}$, that has zero mean and covariance $\boldsymbol{\sigma}_0^T A_{11} \boldsymbol{\sigma}_0$ (this covariance follows from the fact that $A_{11} Q A_{11} = A_{11}$).

3.3 The kriging estimator is a perfect interpolator

We conclude this section by proving an interesting property of the kriging estimator $\hat{z}(\mathbf{x}_0)$, i.e., that if $\mathbf{x}_0 = \mathbf{x}_i$, where i is any one of the locations of the control points or data, then $\hat{z}(\mathbf{x}_i) = z(\mathbf{x}_i)$, so that the interpolator at $\mathbf{x}_0 = \mathbf{x}_i$ equals the data value at $\mathbf{x}_i$ (i.e., the kriging estimator is a perfect interpolator). Let $\mathbf{x}_0 = \mathbf{x}_i$, then, the weighting vector $\boldsymbol{\lambda}_i$ from Eq. (20) may be written

$$\boldsymbol{\lambda}_i = A_{11} \boldsymbol{\sigma}_i + A_{12} \mathbf{f}(\mathbf{x}_i). \quad (40)$$

Since $\mathbf{x}_0 = \mathbf{x}_i$ then $\boldsymbol{\sigma}_i = \text{Cov}[z(\mathbf{x}_i), \mathbf{z}]$ which is the i th column of the covariance matrix Q given in Eq. (15), i.e., $Q = (\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_i, \dots, \boldsymbol{\sigma}_n)$. From Eq. (22)

$$A_{11} Q + A_{12} F^T = I_{nn} \quad (41)$$

where I_{nn} denotes the $n \times n$ identity matrix. Equation (41) can be decomposed columnwise so that the equation corresponding to the i th column is

$$A_{11} \boldsymbol{\sigma}_i + A_{12} \mathbf{f}(\mathbf{x}_i) = \mathbf{e}_i \quad (42)$$

in which $\mathbf{e}_i$ is the i th unit vector with all of its n elements equal to zero except for the i th component which is equal to one. Comparing Eqs. (40) and (42) it follows that $\boldsymbol{\lambda}_i = \mathbf{e}_i$ so that $\hat{z}(\mathbf{x}_i) = \boldsymbol{\lambda}_i^T \mathbf{z} = \mathbf{e}_i^T \mathbf{z} = z(\mathbf{x}_i)$ and the proof that the kriging estimator is a perfect interpolator is complete. The reader can readily verify that the solution $\boldsymbol{\lambda}_i = \mathbf{e}_i$ satisfies the unbiasedness condition (11) and the minimum variance of estimation error property (12).

The implementation of the estimators $\hat{z}(\mathbf{x}_0)$ and $\hat{m}(\mathbf{x}_0)$ requires that the variance components σ_r^2 , $r = 1, 2, \dots, p$ be specified or estimated. In the next section three methods are discussed that yield the variance components: full information maximum likelihood (FIML) for the simultaneous estimation of $\boldsymbol{\beta}$ and σ_r^2 , $r = 1, 2, \dots, p$, restricted maximum likelihood (RML) and minimum invariant quadratic unbiased estimation (MINQUE).

4 Parameter estimation in the linear model

4.1 Simultaneous estimation of drift parameters and variance components by FIML

Assuming normality of the data $\mathbf{z}$ the log-likelihood function corresponding to the linear model given by Eqs. (6)-(7) is as follows

$$f \doteq -\frac{1}{2} \ln \left| \sum_{r=1}^p \sigma_r^2 Q_r \right| - \frac{1}{2} (\mathbf{z} - F\boldsymbol{\beta})^T \left(\sum_{r=1}^p \sigma_r^2 Q_r \right)^{-1} (\mathbf{z} - F\boldsymbol{\beta}). \quad (43)$$

The FIML estimators of $\boldsymbol{\beta}$ and σ_r^2 , $r = 1, 2, \dots, p$ are obtained by the Newton-Raphson method. Let $\boldsymbol{\Theta}^T = (\boldsymbol{\beta}^T, \boldsymbol{\sigma}^T)$, where $\boldsymbol{\sigma}^T = (\sigma_1^2, \dots, \sigma_p^2)$, be the param-

eter set. In the Newton-Raphson method the $(r+1)$ st iteration is described by the expression

$$\Theta^{(r+1)} = \Theta^{(r)} - \omega \Phi^{-1} \nabla f \quad (44)$$

in which Φ is the Hessian or matrix of second derivatives evaluated at $\Theta^{(r)}$, i.e.,

$$\Phi = \left[\frac{\partial^2 f}{\partial \Theta \partial \Theta^T} \right] \quad (45)$$

and ∇f is the gradient of the function f evaluated at $\Theta^{(r)}$,

$$\nabla f = \left[\frac{\partial f}{\partial \Theta} \right] \quad (46)$$

ω is a steplength factor that ensures that $f[\Theta^{(r+1)}]$ is a minimum with respect to ω ($0 < \omega \leq 1$). The elements of the gradient vector are given by the following expressions (recall $Q = \sum_{r=1}^p \sigma_r^2 Q_r$)

$$\frac{\partial f}{\partial \beta} = F^T Q^{-1} \mathbf{z} - F^T Q^{-1} F \beta \quad (47)$$

(let $\tilde{\mathbf{z}} = \mathbf{z} - F\beta$, and $\text{tr} \equiv \text{trace}$)

$$\frac{\partial f}{\partial \sigma_r^2} = -\frac{1}{2} \text{tr}(Q^{-1} Q_r) + \frac{1}{2} \tilde{\mathbf{z}}^T (Q^{-1} Q_r Q^{-1}) \tilde{\mathbf{z}}, \quad r = 1, 2, \dots, p \quad (48)$$

whereas the components of the Hessian matrix are,

$$\frac{\partial^2 f}{\partial \beta \partial \beta^T} = -F^T Q^{-1} F \quad (49)$$

$$\frac{\partial^2 f}{\partial \beta \partial \sigma_r^2} = F^T Q^{-1} Q_r Q^{-1} F \beta - F^T Q^{-1} Q_r Q^{-1} \mathbf{z}, \quad r = 1, 2, \dots, p \quad (50)$$

$$\frac{\partial^2 f}{\partial \sigma_r^2 \partial \sigma_m^2} = \frac{1}{2} \text{tr}[Q^{-1} Q_r Q^{-1} Q_m] - \tilde{\mathbf{z}}^T [Q^{-1} Q_r Q^{-1} Q_m Q^{-1}] \tilde{\mathbf{z}}, \quad r, m = 1, 2, \dots, p. \quad (51)$$

The covariance of the FIML estimators is approximated by the negative of the inverse of the Hessian evaluated at the solution Θ^* , i.e.

$$\text{Cov}(\Theta^*) \simeq [-\Phi(\Theta^*)]^{-1}. \quad (52)$$

A variation of the Newton-Raphson method is the method of scoring (Rao 1973) in which the iterations are given by Eq. (44) with $-\Phi$ replaced by the information matrix H ,

$$H \equiv -E \left[\frac{\partial^2 f}{\partial \Theta \partial \Theta^T} \right] \quad (53)$$

so that Eq. (44) becomes

$$\Theta^{(r+1)} = \Theta^{(r)} + \omega H^{-1} \nabla f. \quad (54)$$

The elements of the information matrix H are derived from Eqs. (49)-(51) after multiplying those equations by -1 and taking expectations to yield

$$-E\left[\frac{\partial^2 f}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right] = F^T Q F \quad (55)$$

$$-E\left[\frac{\partial^2 f}{\partial \boldsymbol{\beta} \partial \sigma_r^2}\right] = \mathbf{0} \quad (56)$$

$$-E\left[\frac{\partial^2 f}{\partial \sigma_m^2 \partial \sigma_r^2}\right] = \frac{1}{2} \text{tr}[Q^{-1} Q_r Q^{-1} Q_m], \quad r, m = 1, 2, \dots, p. \quad (57)$$

From the Cramer-Rao inequality (Theil 1971) the inverse of the information matrix constitutes a lower bound to the covariance of the FIML estimator.

The FIML and scoring methods require a consistent initial estimator to begin the iterative search for a local maximum. Such initial estimator may be provided by the MINQUE method as explained below. It should be pointed out the FIML (or for that matter any estimation algorithm based on the log-likelihood function) method does not necessarily converge to a local optimum, and in fact the log-likelihood function could be unbounded in which case the method fails. It will be shown in the application Section 5 that such a situation is bound to occur when the covariance matrix (15) is nearly positive semidefinite, a likely scenario in hydrogeologic applications where the regionalized variable has a large correlation length-scale. Section 5 presents such an example.

As a caveat, if the linear model (6)-(7) were modified to be nonlinear on the drift parameter (see, e.g., Gambolati and Volpi 1979) e.g.,

$$\mathbf{z} = F(\boldsymbol{\beta}) + \boldsymbol{\xi} \quad (58)$$

in which $F(\boldsymbol{\beta})$ indicates that the elements of the matrix F are nonlinear functions of $\boldsymbol{\beta}$, then it may be convenient to recourse to quasi-Newton methods as presented in hydrogeological applications by, say, Carrera and Neuman (1986).

4.2 Restricted maximum likelihood (RML) estimation of variance components

Patterson and Thompson (1971) introduced a modification of the FIML method: the RML approach in which the data vector $\mathbf{z}$ is projected into a subspace where the mean $F\boldsymbol{\beta}$ is equal to zero. Specifically, Eq. (6) is premultiplied by the projection matrix T^* such that

$$T^* \mathbf{z} = T^* F \boldsymbol{\beta} + T^* \boldsymbol{\xi} = T^* \boldsymbol{\xi} \quad (59)$$

so that (letting $\mathbf{y}^* = T^* \mathbf{z}$)

$$E(\mathbf{y}^*) = \mathbf{0} \quad (60)$$

and

$$\text{Cov}(\mathbf{y}^*) = T^* Q T^{*T} = \sum_{r=1}^p \sigma_r^2 T^* Q_r T^{*T}. \quad (61)$$

It is easily verified that the choice $T^* = I_{nn} - F(F^T F)^{-1} F^T$ is such that $T^* F = \mathbf{0}$. Clearly, T^* is symmetric and idempotent (i.e., $T^* = T^* T^*$) with rank equal to its trace which is equal to $n - k$, where k is the column dimension of the matrix F . Because of the rank deficiency of T , k of its rows are linear combination of the remaining $n - k$ rows and are redundant. In order to avoid singularity of the covariance of the transformed vector $\mathbf{y}^*$, one must use a projection matrix T that results from deleting (any) subset of k rows from the matrix T^* so that the transformed data vector used in estimation in

$$\mathbf{y} = T\mathbf{z} \quad (62)$$

with expected value $E(\mathbf{y}) = \mathbf{0}$ and (nonsingular) covariance

$$\text{Cov}(\mathbf{y}) = \sum_{r=1}^p \sigma_r^2 V_r \quad (63)$$

where

$$V_r = TQ_rT^T. \quad (64)$$

The dimensions of the vector $\mathbf{y}$ and matrix T are $(n - k) \times 1$ and $(n - k) \times n$ respectively.

From Eqs. (62)-(63) and under normality of $\mathbf{z}$ the log-likelihood function becomes

$$f_R \doteq -\frac{1}{2} \ln |V| - \frac{1}{2} \mathbf{y}^T V^{-1} \mathbf{y} \quad (65)$$

in which

$$V = \sum_{r=1}^p \sigma_r^2 V_r. \quad (66)$$

It is apparent that Eq. (66) is a function of the variance components σ_r^2 alone and the Newton-Raphson method (see Eq. (44)) may be applied to Eq. (65) to yield a local optimum. The elements of the gradient of Eq. (65) are

$$\frac{\partial^2 f_R}{\partial \sigma_r^2} = -(1/2) \text{tr}[V^{-1} V_r] + \frac{1}{2} \mathbf{y}^T (V^{-1} V_r V^{-1}) \mathbf{y}, \quad r = 1, 2, \dots, p \quad (67)$$

whereas the elements of the Hessian are given by

$$\frac{\partial^2 f_R}{\partial \sigma_r^2 \partial \sigma_m^2} = \frac{1}{2} \text{tr}[V^{-1} V_r V_m] - \mathbf{y}^T [V^{-1} V_r V^{-1} V_m V^{-1}] \mathbf{y}, \quad r, m = 1, 2, \dots, p. \quad (68)$$

The Newton-Raphson solution based on Eq. (44) using the gradient and Hessian whose elements are defined by Eqs. (67) and (68), respectively, constitutes the RML method. The covariance of the variance components can be approximated by the negative of the inverse of the Hessian matrix. The method of scoring is also applicable to estimate the σ_r^2 in which case the information matrix has elements

$$-E \frac{\partial^2 f_R}{\partial \sigma_r^2 \partial \sigma_m^2} = \frac{1}{2} \text{tr}[V^{-1} V_r V^{-1} V_m], \quad r, m = 1, 2, \dots, p. \quad (69)$$

Upon solution for the estimates σ_r^2 , $r = 1, 2, \dots, p$, the covariance Q is approximated as follows

$$\hat{Q} = \sum_{r=1}^p \hat{\sigma}_r^2 Q_r. \quad (70)$$

A natural estimator of the drift vector is then

$$\hat{\beta} = (F^T \hat{Q}^{-1} F)^{-1} F^T \hat{Q}^{-1} \mathbf{z} \quad (71)$$

with approximate covariance

$$\text{Cov}(\hat{\beta}) = (F^T \hat{Q}^{-1} F)^{-1}. \quad (72)$$

4.3 Minimum invariant quadratic unbiased estimation (MINQUE) of variance components

The MINQUE approach is due to Rao (1971, 1972). Consider the estimation of the linear function $\mathbf{c}^T \boldsymbol{\sigma}$ (recall $\boldsymbol{\sigma}^T = (\sigma_1^2, \dots, \sigma_p^2)$). In particular, when the vector $\mathbf{c}$ contains all zeroes except for its r th element being equal to one then the estimated quantity is σ_r^2 . In the MINQUE method the linear combination $\mathbf{c}^T \boldsymbol{\sigma}$ is estimated by the quadratic function $\mathbf{z}^T A \mathbf{z}$ where the matrix A must be determined after imposing certain conditions on the quadratic estimator $\mathbf{z}^T A \mathbf{z}$. The first condition is that of invariance, i.e., if the drift parameter β is replaced by $\beta_d = \beta - \beta_0$ then the model (6) becomes

$$\mathbf{z}_d = F\beta_d + \xi \quad (73)$$

where $\mathbf{z}_d = \mathbf{z} - F\beta_0$. Invariance under translation of the β parameter requires that the estimator $\mathbf{z}_d^T A \mathbf{z}_d$ be equal to $\mathbf{z}^T A \mathbf{z}$ for all β_0 , i.e.,

$$(\mathbf{z} - F\beta_0)^T A (\mathbf{z} - F\beta_0) = \mathbf{z}^T A \mathbf{z} \quad (74)$$

implying that

$$AF = 0. \quad (75)$$

The second condition imposed upon the estimator $\mathbf{z}^T A \mathbf{z}$ is that it be unbiased,

$$E(\mathbf{z}^T A \mathbf{z}) = \mathbf{c}^T \boldsymbol{\sigma}. \quad (76)$$

After taking the expected value on the left-hand side of Eq. (76) one obtains

$$\text{tr}[A \sum_{r=1}^p \sigma_r^2 Q_r] = \mathbf{c}^T \boldsymbol{\sigma} \quad (77)$$

from which

$$\text{tr}[AQ_r] = c_r, \quad r = 1, 2, \dots, p \quad (78)$$

which is the unbiasedness condition. The last property of the quadratic estimator is that it has minimum norm. Referring to the equivalent representation (8) of model (6) it is evident that if the error vectors ξ_i were known, then a natural estimator of σ_r^2 , $\forall r$, would be $\xi_r^T \xi_r / n$, and hence for $\mathbf{c}^T \boldsymbol{\sigma}$ would be

$$\sum_{r=1}^p c_r [\xi_r^T \xi_r / n] = \mathbf{e}^T D \mathbf{e} \quad (79)$$

in which

$$\mathbf{e}^T = [\xi_1^T, \xi_2^T, \dots, \xi_p^T] \quad (80)$$

and

$$D = \text{Block diagonal } [(c_1/n)I_{nn}, \dots, (c_p/n)I_{nn}]. \quad (81)$$

Note that the proposed estimator is

$$\mathbf{z}^T A \mathbf{z} = \xi^T A \xi \quad (82)$$

where the invariance condition (75) was used to establish Eq. (82). The right-hand side of Eq. (82) can be rewritten, by virtue of the representation (8) as follows,

$$\xi^T A \xi = e^T U^T A U e \quad (83)$$

in which

$$U = [T_1, T_2, \dots, T_p]. \quad (84)$$

The difference between the natural estimator (79) and the proposed estimator (83), i.e.,

$$e^T D e - \xi^T A \xi = e^T [D - U^T A U] e \quad (85)$$

may be made small in the sense of minimizing the Euclidean norm (i.e., $B^2 = \sum_{i,j} b_{ij}^2$) of $U^T A U - D$, i.e.,

$$\underset{A}{\text{minimize}} \quad U^T A U - D^2. \quad (86)$$

It may be easily shown that the minimization of Eq. (86) is equivalent to

$$\underset{A}{\text{minimize}} \quad \text{tr}[A W A W] \quad (87)$$

where

$$W = U U^T = \sum_{r=1}^p Q_r. \quad (88)$$

In addition, Eq. (87) is subject to the invariance and unbiasedness conditions, Eqs. (75) and (78) respectively. From multivariate normal theory, it follows that if the data vector $\mathbf{z}$ is normal then the variance of the invariant estimator $\mathbf{z}^T A \mathbf{z}$ is:

$$\text{Var}(\mathbf{z}^T A \mathbf{z}) = 2\text{tr}[A Q A Q]. \quad (89)$$

Thus, under normality, the variance of the estimator $\mathbf{z}^T A \mathbf{z}$ is available in exact form from the expression in Eq. (89). Notice, however, that none of the developments leading to Eq. (87) requires that $\mathbf{z}$ be normal.

Rao (1973, pp. 65-66) shows that the matrix A that minimizes Eq. (87) subject to Eqs. (75) and (78) is given by

$$A = \sum_{r=1}^p b_r Q_r Q_r \quad (90)$$

in which

$$R = W^{-1} - W^{-1} F (F^T W^{-1} F)^{-1} F^T W^{-1} \quad (91)$$

and the b_r 's are obtained from the following system of linear equations,

$$\sum_{r=1}^p b_r \text{tr}[Q_r Q_r Q_j] = c_j, \quad j = 1, 2, \dots, p. \quad (92)$$

Upon MINQUE estimation of σ , the drift parameter and its covariance are given by Eqs. (71)-(72) respectively.

The MINQUE approach is attractive for several reasons. It has desirable statistical properties and the solution of the minimization problem (see Eq. (87)) is reduced to solving a $p \times p$ linear system of equations as indicated by Eq. (92).

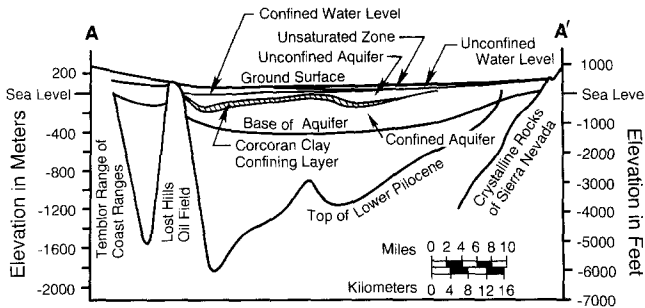


Figure 3. Typical stratigraphic cross-section of the San Joaquin Valley

There is no need for initial estimates and, under normality the variance of the estimators are available in exact form (see Eq. (89)), although the method does not impose any distributional assumptions.

4.4 Parameter estimation and linear interpolation of the regionalized variable

Sections 2 and 3 were developed under the assumption that model parameters were known. In practice one must estimate such parameters as indicated in subsections 4.1 - 4.3 and subsequently use such estimates in the interpolation of $z(\mathbf{x})$ or its mean $m(\mathbf{x})$. The introduction of parameter estimates represent an approximation to the theoretical results presented in Sections 2 and 3. In general, as the data $\mathbf{z}$ becomes larger there will be a greater correspondence to the theoretical properties discussed in Sections 2 and 3. Section 5 illustrates the results and methods presented heretofore with an actual case study.

5 Application to mapping of specific yield in the San Joaquin Valley, California

5.1 Geologic setting

The San Joaquin Valley (see Fig. 2) is about 400 km long by 55 to 65 km wide. The 3.4 million hectares (ha) of the valley floor lie within the Sierra Nevada to the west, the Coast Ranges to the east, the Tehachapi Mountains to the south, and the delta of the Sacramento and San Joaquin rivers to the north. The San Joaquin Valley is California's most productive block of farmland with over 95% of the water use in the valley allocated to agriculture (Loaiciga and Mariño 1987). The valley's groundwater is drawn from aquifers having an estimated storage capacity of 100 million dam³ (1 dam³ = 10³ m³). The geologic structure beneath the surface of the valley consists of sedimentary layers sloping downward from the east side of the valley to a low point near the valley's west side. From the trough or low point, the layers slope steeply upward into the Coast Range mountains to the west. On the east edge of the fold, water bearing sediments are cutoff by rocks of the Sierra Nevada. The southern boundary of the groundwater basin is the Tehachapi and San Emigdio mountains. The hard rocks found in the east, west and south sides of the valley act as impermeable barriers cutting off flow into or out of the valley. The loose sediments are continuous across the northern boundary of the groundwater basin and present no barrier to subsurface flow. The geology of the San Joaquin Valley is complex when considered in detail but if viewed as a single large unit, it is composed basically of four layers as seen in Fig. 3: (1) an unsaturated layer lying between the ground surface and the water table; (2) an

unconfined saturated layer; (3) a saturated leaky confining clay layer of low hydraulic conductivity; and (4) a confined saturated layer underlying the leaky confining layer. For all practical purposes, groundwater flow can be assumed to be horizontal in the confined and unconfined aquifers and vertical through the leaky confining layer. The presence of the leaky confining layer and localized clay lenses is due to the deposition of fine soil particles over geologic time when the valley trough was occupied by large bodies of water.

5.2 Selection of the drift $F\beta$ and covariance structure

The variable of interest in this study is the specific yield, S_y , that is defined as the volume of water drained per unit horizontal area of unconfined (or phreatic) aquifer and per unit drop of the ground water saturated surface. The California Department of Water Resources (1982) reported 217 measurements of specific yields over the San Joaquin Valley. The interpolation and estimation methodologies presented in Sections 2-4 are applied to the northernmost part of the valley (i.e., above coordinate $Y = -5$ mi in Fig. 2) where the specific yield shows a trend characterized by the lowest values at the eastern edge of the valley, gently sloping upward in a westward direction until reaching a maximum around the valley trough and then decreasing towards the western side of the valley. In the south-north direction the distribution of specific yields is relatively uniform. Figure 4 displays a histogram of measured specific yields in the northernmost part of the study area showing a remarkable symmetric empirical frequency for the chosen intervals (S_y values in percentages). The observed trend suggests that a quadratic polynomial may provide an adequate fit to the spatial distribution of specific yield. The quadratic polynomial is given by the following expression

$$m(\mathbf{x}) = b_0 + b_1x + b_2y + b_3x^2 + b_4y^2 + b_5xy \quad (93)$$

where the b_i 's are the drift parameters and x and y represent the coordinates of the location vector $\mathbf{x}$. With the notation set forth in Eq. (3) it is seen that $\mathbf{f}^T(\mathbf{x}) = (1, x, y, x^2, y^2, xy)$ and $\boldsymbol{\beta}^T = (b_0, b_1, b_2, b_3, b_4, b_5)$. It is emphasized that the selection of the drift model was based on the observed trend of the regionalized variable which clearly rules out linear or higher than second order polynomial fit. It is plausible in cases where model discrimination is not obvious from the observed data to resource to statistical identification techniques, such as for example, the Akaike Information Criterion (Akaike 1974). Due to the gently and clear curvilinear trend observed in this study the use of statistical discrimination techniques was not warranted.

Two of the most widely used covariance models in hydrogeologic studies, i.e., the isotropic exponential and the spherical were considered in this work. The truncated exponential model is given by

$$C(r_{ij}) = \begin{cases} \sigma_1^2\delta_{ij} + \sigma_2^2\exp[-r_{ij}/L], & r_{ij} \leq L \\ 0, & r_{ij} > L \end{cases} \quad (94)$$

and the spherical covariance by

$$C(r_{ij}) = \begin{cases} \sigma_1^2\delta_{ij} + \sigma_2^2[1 - 3r_{ij}/2L + r_{ij}^3/2L^3], & r_{ij} \leq L \\ 0, & r_{ij} > L \end{cases} \quad (95)$$

where $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ represents the magnitude of the separation vector between any two locations $\mathbf{x}_i$ and $\mathbf{x}_j$; δ_{ij} is the Kronecker delta, and L is the correlation

length-scale beyond which spatial correlation becomes negligible. The genesis of the nugget effect (σ_1^2) in hydrogeologic phenomena is related to concentrated small scale variability, to measurement errors or to the fact that the data have not been collected with a sufficiently detailed resolution to show the underlying continuous behavior of the regionalized variable (see, e.g., de Marsily 1986). The variance component σ_2^2 accounts for structured variability of the regionalized variable.

In this study we followed an iterative approach to determine the correlation length-scale L based on the entire 217 measurements of specific yields in the San Joaquin Valley. The iterative determination of L was done by using the empirical covariance $\hat{C}(r_{ij})$,

$$\hat{C}(r_{ij}) = \sum [\hat{\xi}(\mathbf{x}_i) \hat{\xi}(\mathbf{x}_j)] / n_r \quad (96)$$

in which $\hat{\xi}(\mathbf{x}_i)$ and $\hat{\xi}(\mathbf{x}_j)$ denote the estimated residuals,

$$\hat{\xi}(\mathbf{x}_k) = z(\mathbf{x}_k) - \hat{m}(\mathbf{x}_k), \quad k = i, j \quad (97)$$

and n_r is the number of pairs of control points $z(\mathbf{x}_i)$ and $z(\mathbf{x}_j)$ separated by a distance r_{ij} , and $\hat{m}(\mathbf{x}_k)$ is the fitted trend. Equation (96) was used iteratively, i.e., first, the trend was fitted by ordinary least squares (OLS), assuming no correlation structure, and the estimated residuals (97) were obtained to calculate $\hat{C}(r_{ij})$. The correlation length-scale L was estimated from the empirical covariance as the distance over which the covariance becomes negligible (see Fig. 5). Based on that L the variance components σ_r^2 , $\forall r$, were computed with the MINQUE method. The drift or trend was re-estimated using the generalized least squares (GLS) approach (see Eqs. (29) and (31)) and a new set of residuals (see Eq. (97)) were derived and used to revise the empirical covariance. The procedure was repeated until a stable empirical covariance was obtained with a range of approximately 50 mi (80 km) as shown in Fig. 5.

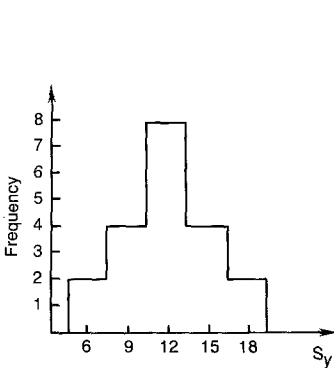


Figure 4

Figure 4. Histogram of specific yields within the study area

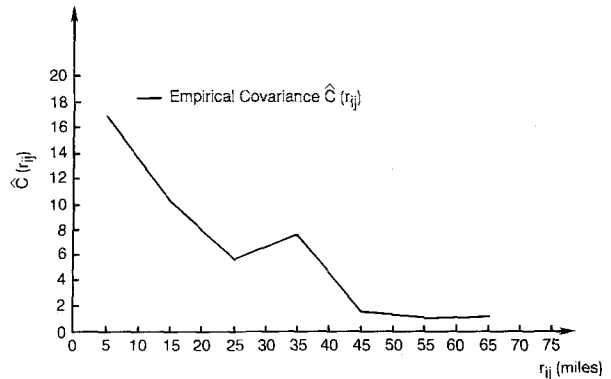


Figure 5

Figure 5. Iterative estimate of the covariance

Table 2. Estimated variance components and drift parameters by MINQUE method^a

Parameters	Range ^b <i>L</i>		
	40	50	60
σ_1^2	0.6834 (0.7832)	0.7424 (0.9766)	0.9573 (1.1040)
σ_2^2	13.45 (6.9904)	14.01 (7.5782)	14.61 (7.8832)
b_0	9.0859 (15.524)	9.1686 (15.874)	9.6026 (16.523)
b_1	0.5381 (0.5630)	0.5352 (0.5762)	0.5197 (0.5992)
b_2	-0.5453 (0.4899)	-0.5466 (0.5020)	-0.5532 (0.5244)
b_3	-0.0084 (0.0001)	-0.0084 (0.0001)	-0.0083 (0.0001)
b_4	0.0091 (0.0001)	0.0091 (0.0001)	0.0091 (0.0001)
b_5	0.0029 (0.0001)	0.0030 (0.0001)	0.0032 (0.0001)

^a Figures within parentheses are the standard deviation of estimators.

^b The range is expressed in miles (1 mile = 1.61 km).

For the sake of comparison, two other correlation length-scales equal to 40 mi (64 km) and 60 mi (96 km) were considered in the application to assess the impact of the range r on parameter and linear spatial estimation. Ripley (1981) and Neuman and Jacobson (1984) reported an analogous iterative scheme to establish the empirical semivariogram. Gambolati and Galeati (1987) presented evidence indicating that kriged estimate (and their variances of estimation error) based on the iterative empirical semivariogram are roughly comparable to those obtained from the initial estimate of the empirical semivariogram based on ordinary least squares. It will be shown below that parameter estimates of σ_1^2 and σ_2^2 and the drift β were only modestly sensitive to variations of the correlation length-scale within the range 40-60 mi (64-96 km), whereas the kriged variables $\hat{z}(\mathbf{x}_0)$ and interpolated means $\hat{m}(\mathbf{x}_0)$ (and their estimation errors) were quite insensitive to the choice of L , supporting at least partly arguments set forth by Gambolati and Galeati (1987).

In principle, the correlation length-scale could be estimated directly by FIML or RML. However, this alternative introduces considerable nonlinearities and, in addition, for large values of L the exponential and spherical models are approximated by $\sigma_1^2 \delta_{ij} + \sigma_2^2$ introducing serious problems in the FIML estimation to be addressed below. In view of the complexities introduced by the parameter L it is advantageous to approximate the location of the sill and thus the magnitude of correlation length-scale by means of the empirical semivariogram and covariance as done in this work, and then explore the sensitivity of the parameter and interpolated variables to change in the correlation length-scale.

5.3 Parameter estimation

5.3.1 Minimum invariant quadratic unbiased estimation (MINQUE)

The simplest method for the estimation of the variance components σ_1^2 and σ_2^2 is the MINQUE method as explained in Section 4. The MINQUE approach is valid regardless of the data distribution. In particular, normality is not assumed, although in this study the data showed a remarkable symmetry (see Fig. 4). In order to assess the impact of the correlation length-scale on the estimates of σ_1^2 and σ_2^2 , L was set equal to 40, 50 and 60 mi (64, 80 and 96 km). Upon estimation of σ_1^2 and σ_2^2 , the drift parameter $\hat{\beta}$ and its covariance were calculated according to Eqs. (71) and (72) respectively. Table 2 shows the estimates of σ_1^2 , σ_2^2 and $\hat{\beta}$ for different values of L with an exponential covariance model. The standard deviations of the estimators are shown within parentheses. It is seen in Table 2 that the variance components $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ increase (decrease) as L increases (decreases) for the range of correlation length-scale considered, although such variations are quite moderate. The drift parameters are even less sensitive to change in correlation length-scale with the estimator $\hat{b}_0$ being the most responsive to the changes in L and also showing the largest standard deviation of estimation. Interestingly, the structured variability $\hat{\sigma}_2^2$ was an order of magnitude larger than the local variability $\hat{\sigma}_1^2$ meaning that structural spatial correlation is the dominant component of covariance. As will be shown in the discussion of the FIML method below, when $\hat{\sigma}_2^2 \gg \hat{\sigma}_1^2$ and there is a large correlation length-scale the log-likelihood function becomes unbounded. Table 3 displays the parameter estimates using the MINQUE method when the isotropic covariance model was used. From Tables 2 and 3 it is apparent that the choice of covariance model had negligible influence on the parameter estimates since their values were essentially equal under the two covariance model assumptions. Since the initial estimates for the RML and FIML methods were the MINQUE values, all results presented in the sequel will correspond to the exponential covariance model.

In summary, it is seen that the parameter estimates are robust relative to the choice of L so that errors in estimating L as explained above did not have an appreciable effect in this study, something that may be reasonably attributed to the large correlation length-scale of specific yields.

5.3.2 Restricted maximum likelihood (RML) estimation

The implementation of the Newton-Raphson search was initiated by using the MINQUE estimators $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. Table 4 displays the RML estimates $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\beta}$ for $L = 40, 50$, and 60 mi (64, 80 and 96 km). $\hat{\beta}$ and its covariance were computed from the GLS Eqs. (71)-(72) respectively. The variance components estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ tend to increase (decrease) as the range L increases (decreases) in a similar fashion as observed with the MINQUE estimates for the values of L considered. An interesting feature of the RML estimates $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ is that their standard deviations were substantially larger than those obtained for the MINQUE estimates (compare with Tables 2 and 3), a result not at all unexpected due to the minimum norm property which translates into minimum variance under normality given additional conditions as shown by Rao (1972) (among quadratic, unbiased and invariant estimators). From Table 4 is apparent that in general the estimates

Table 3. Estimated variance components and drift parameters by MINQUE method and spherical covariance^a

Parameters	Range ^b <i>L</i>		
	40	50	60
σ_1^2	0.6971 (0.7989)	0.7647 (0.9989)	0.9665 (1.1150)
σ_2^2	13.85 (7.0603)	14.29 (7.8640)	14.76 (14.7561)
b_0	9.0676 (15.748)	9.1879 (16.062)	9.60141 (16.613)
b_1	0.5387 (0.5710)	0.5345 (0.5822)	0.5199 (0.6017)
b_2	-0.5449 (0.4970)	-0.5468 (0.5079)	-0.5532 (0.5263)
b_3	-0.0084 (0.0001)	-0.0084 (0.0001)	-0.0083 (0.0001)
b_4	0.0091 (0.0001)	0.0091 (0.0001)	0.0091 (0.0001)
b_5	0.0029 (0.0001)	0.0030 (0.0001)	0.0032 (0.0001)

^a Figures within parentheses are the standard deviation of estimators.^b The range is expressed in miles (1 mile = 1.61 km).**Table 4.** Parameter estimates by RML method^a

Parameters	Range ^b <i>L</i>		
	40	50	60
σ_1^2	0.7134 (2.067)	0.7480 (2.217)	0.9643 (2.321)
σ_2^2	13.163 (20.763)	13.832 (21.380)	14.520 (22.486)
b_0	9.2130 (15.421)	9.2086 (15.811)	9.6312 (16.492)
b_1	0.5336 (0.5595)	0.5337 (0.5736)	0.5189 (0.5975)
b_2	-0.5472 (0.4879)	-0.5472 (0.5000)	-0.5536 (0.5235)
b_3	-0.0084 (0.0001)	-0.0084 (0.0001)	-0.0082 (0.0001)
b_4	0.0091 (0.0001)	0.0091 (0.0001)	0.0091 (0.0001)
b_5	0.0030 (0.0001)	0.0030 (0.0001)	0.0032 (0.0001)

^a Figures within parentheses are the standard deviation of estimators.^b The range is expressed in miles (1 mile = 1.61 km).

$\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\beta}$ were only moderately sensitive to changes in L .

5.3.3 Full information maximum likelihood (FIML) estimation

It is convenient to write the linear model as

$$\mathbf{z} = F\boldsymbol{\beta} + T_1\boldsymbol{\xi}_1 + T_2\boldsymbol{\xi}_2 \quad (98)$$

according to Eq. (8) with $T_1 = I_{nm}$ so that $\text{Cov}(\boldsymbol{\xi}_1) = \sigma_1^2 I_{nm}$; and T_2 being such that

$$T_2 T_2^T = Q_2 \text{ and } \text{Cov}(\boldsymbol{\xi}_2) = \sigma_2^2 Q_2. \quad (99)$$

Suppose that T_2 is chosen to be

$$T_2 = EG^{1/2} \quad (100)$$

where the columns of the orthogonal matrix E are the eigenvectors of Q_2 and $G^{1/2}$ is a diagonal matrix whose elements are the square roots of the eigenvalues of Q_2 . It was found that the determinant of the matrix Q_2 was $|Q_2| = 3 \times 10^{-14}$ which is quite small, so that Q_2 is almost positive semidefinite. The near singularity of Q_2 is due to the large correlation length-scale L which implies that the columns of Q_2 show strong collinearity. It is easily shown that based on the representation (98) the log-likelihood function (43) is approximately equal to

$$f \simeq -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n \ln(\lambda_i \sigma_2^2 + \sigma_1^2) \quad (101)$$

in which λ_i is the i th eigenvalue of Q_2 . Equation (101) is a direct consequence of two facts: (i) the matrix Q_2 is near singular (i.e., $|Q_2| \rightarrow 0$), and (ii) the structural variance component σ_2^2 is one order of magnitude larger than σ_1^2 as shown by the MINQUE and RML results. It is clear from Eq. (101) that the log-likelihood function is maximized by $\sigma_2^2 \rightarrow 0$ and $\sigma_1^2 \rightarrow 0$. Notice that the RML method was successful but a set of constraints were applied to the data in the form of Eq. (62).

In summary, FIML failed to yield consistent estimates of σ_1^2 , σ_2^2 and $\boldsymbol{\beta}$ because the spatial covariance Q_2 was nearly rank deficient, a likely scenario when there is strong structural correlation of the regionalized variable. Evidently, the MINQUE and RML methods are easier to implement than FIML and in addition, are not as susceptible to ill-conditioning. The joint estimation of variance components and drift parameters is not recommended when the correlation length-scale is large and the local scale variability is minor relative to the structural variability.

5.4 Linear spatial interpolation

The spatial interpolation results to be discussed in this section are all based on the MINQUE parameter estimates with a correlation length-scale of 50 mi (80 km). The interpolation results were quite insensitive to the changes of correlation length-scale within the interval 40-60 mi (64-96 km) and to the differences between the variance components estimates from either the MINQUE or RML methods (see Tables 2, 3 and 4), with the spatial interpolation results being within $\pm 5\%$ from each other in all cases. Figures 6 and 7 show contour lines of kriged values $\hat{z}(\mathbf{x})$ and interpolated means $\hat{m}(\mathbf{x})$ respectively. It is apparent from Figs. 6 and 7

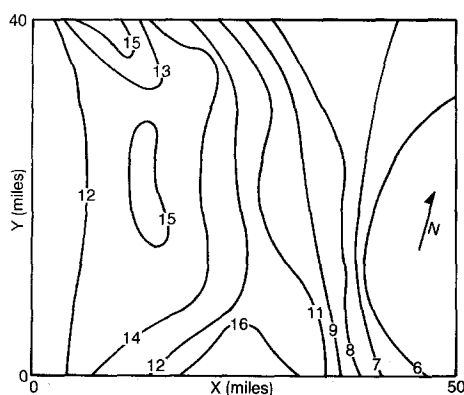


Figure 6

Figure 6. Kriged contours of the regionalized variable

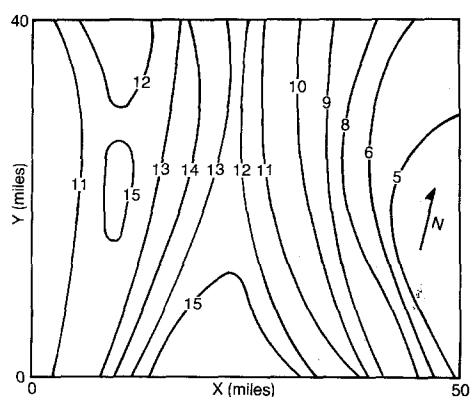


Figure 7

Figure 7. Contours of the estimated trend of the regionalized variable

Table 5. Illustrative results of single realization and mean estimation

Property	Coordinate ^b		
	(39.3, 17.1)	(26.5, 8.0)	(53.6, 24.5)
$\hat{z} = \lambda^T z$	13.6903	14.5571	9.2090
$\text{Var}(\hat{z})$	12.5269	19.2669	12.9380
$\text{Var}(\hat{z} - z)$	2.4546	7.6508	2.08378
$\hat{m} = f^T \beta$	12.5948	14.3242	9.7883
$\text{Var}(\hat{m})$	11.9981	19.0827	12.2798
$\text{Var}(\hat{m} - z)$	2.9834	7.83501	2.74278

^a Range $L = 50$ mi (80 km).

^b Coordinates (X, Y) in miles measured from the origin as shown in Fig. 2; 1mi = 1.61 km.

that $\hat{z}(\mathbf{x})$ and $\hat{m}(\mathbf{x})$ follow a similar general spatial pattern with increasing specific yields in the east-west direction up to about $X \approx 25$ mi (40 km) and then decreasing rapidly in the same direction. In the south-north orientation there is a relative uniform distribution of specific yields except for the discontinuities represented by the contour lines $\hat{z}(\mathbf{x}) = 16$ and $\hat{z}(\mathbf{x}) = 15$ close to ordinates $Y = 0$ and $Y = 40$ in Fig. 6 respectively. The fitted mean $\hat{m}(\mathbf{x})$ in Fig. 7 shows clearly the smooth variability of the drift in the study area. It is emphasized that the kriged values $\hat{z}(\mathbf{x})$ estimate the drift plus local variability (i.e., single realization) whereas $\hat{m}(\mathbf{x})$ is a GLS approximation to the regional trend and this is the main reason behind the local differences between the values in Figs. 6 and 7.

Table 5 shows three (arbitrarily) selected locations with their corresponding kriged and drift estimates as well as their variances and variances of estimation errors. The reader may readily verify that the variances of estimators and of estimation errors conform to the theoretical results presented in Table 1, i.e., the kriged values have smaller variances of estimation errors but higher variance of estimators as compared to the estimated mean or trend. Figure 8 displays the contours of variance of estimation errors $\text{Var}(\hat{z}(\mathbf{x}) - z(\mathbf{x}))$ (continuous line) and

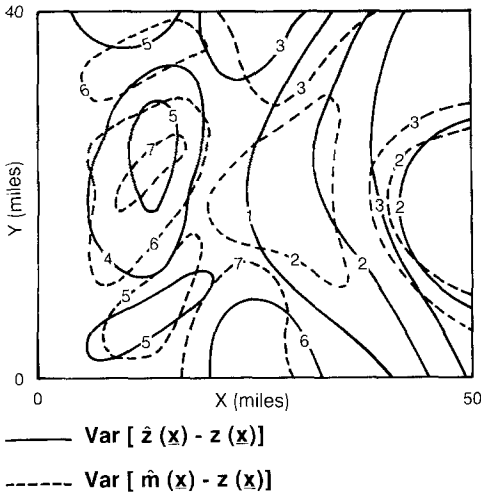


Figure 8. Contours of the variance of estimation errors of $\hat{z}(\mathbf{x})$ (continuous line) and $\hat{m}(\mathbf{x})$ (dotted line)

$\text{Var}(\hat{m}(\mathbf{x}) - z(\mathbf{x}))$ (dotted line) observed within the study area. Clearly, the estimation error $\hat{z}(\mathbf{x}) - z(\mathbf{x})$ has the smaller variance as a direct consequence of the kriging property (12).

6 Summary and conclusions

This work has reviewed the theory of linear spatial estimation of random functions. The emphasis of this paper was to contrast the properties of the single realization estimator vis a vis the trend estimator. These two estimators are linear functions of the control data vector $\mathbf{z}$, yet they estimate different entities and have fundamentally different properties which have been derived analytically. In the case of the single realization estimator $\hat{z}(\mathbf{x})$ explicit and exact expressions for the weighting vector $\boldsymbol{\lambda}$ and the variances of estimator and estimation error were obtained from a closed-form expression of the inverse of the Lagrangian matrix. The properties of the trend estimator $\hat{m}(\mathbf{x})$ follow directly from the Gauss-Markoff theorem. In particular it was shown that the single realization estimator $\hat{z}(\mathbf{x})$ can be decomposed into two orthogonal linear random functions of the data $\mathbf{z}$, one of which is the trend estimator $\hat{m}(\mathbf{x})$.

The implementation of linear spatial estimation was illustrated with three different methods, i.e., full information maximum likelihood (FIML), restricted maximum likelihood (RML), and Rao's minimum invariant quadratic unbiased estimation (MINQUE) for the single realization case and via generalized least squares (GLS) for the trend. The case presented a difficult feature characterized by a large correlation length-scale in the covariance producing a nearly positive semidefinite nested covariance structure. The sensitivity of model parameters, i.e., drift and variance components (local and structured) to the correlation length-scale, choice of covariance model (i.e., exponential and spherical), and estimation method was examined. The same type of sensitivity analysis was conducted for the spatial interpolators. Interestingly, for this case study characterized by a large correlation length-scale of about 50 mi (80 km), both parameter estimates and linear spatial interpolators were rather insensitive to the choice of covariance model and estimation method within the range of credible values obtained for the correlation length-

scale, i.e., 40-60 mi (64-96 km), with differences within alternative estimates falling with $\pm 5\%$ of each other.

Amongst the estimation methods, MINQUE had several attractive features. Besides being unbiased, invariant to linear transformations (see Eq. (74)) and minimum norm, it leads to the solution of a simple linear system of equations (see Eq. (92)) to yield the variance components. Among the three competing estimation methods, MINQUE is less susceptible to numerical problems caused by the inherent collinearity typical of covariance matrices associated with hydrogeological regionalized variables, and under normality, it provides an exact expression for the variance of the covariance components estimators (see Eq. (89)).

The joint estimation of drift and covariance parameters by the FIML methods showed that when the covariance matrix (see Eq. (15)) is nearly positive semidefinite and the nugget effect (i.e., local variability) is small relative to the structured variability, then the log-likelihood function becomes unbounded with its unique and global maximum at $\sigma_1^2 \rightarrow 0$ and $\sigma_2^2 \rightarrow 0$ which are statistically inconsistent estimators. The RML method circumvents this problem by imposing a set of linearly independent restrictions that project the data vector $\mathbf{z}$ into a space of lower dimensionality which is orthogonal to the trend reducing the parameter set to the variance components. Starting the search for their optimal estimates at the MINQUE values produced rapid convergence to a local maximum, although it was observed a rather flat behavior of the log-likelihood function in the neighborhood of the convergence point.

In summary, the application of linear model theory to parameter estimation and spatial interpolation of regionalized variables is a valuable tool and sound alternative to methods based on the formulation of the inverse problem through the numerically discretized equations governing physical processes (see, e.g., Yeh 1986, for a review). In particular, linear theory requires an understanding of the correlation structure of the function of interest, and if the estimation of the trend is relevant, then drift model identification becomes an issue also. This work shows through a case study that the MINQUE approach is suitable for parameter estimation in the event of a large correlation length-scale, and in addition had desirable statistical properties such as unbiasedness and invariance, as well as minimum norm, and exact variance of estimators expressions under normality, that are readily established regardless of the sample size, something which is not possible for nonlinear methods based on the likelihood principle. A theoretical analysis of single realization vs. trend estimation has furthered our understanding of the properties and nature of trend-surface analysis and kriging, and a case study illustrated the issues involved in the application of linear model theory to hydrogeological modeling.

References

- Agterberg, F.P. 1974: Geomathematics - Mathematical background and geoscience applications. Amsterdam: Elsevier Publish. Co.
- Akaike, H. 1974: A new look at the statistical model identification. IEEE Trans. Autom. Contr. AC-19, 716-732
- California Department of Water Resources. 1982: The hydrologic-economic model of the San Joaquin Valley, Bulletin 214, Sacramento, California, U.S.A.
- Carrera, J.; Neuman, S.P. 1986: Estimation of aquifer parameters under transient and steady state conditions, 2. Uniqueness, Stability and solution algorithms. Water Resour. Res. 22, 228-242
- Davis, J.C. 1986: Statistics and data analysis in geology, 2nd ed. New York: John Wiley
- De Marsily, G. 1986: Quantitative hydrogeology. Orlando: Academic Press, Florida, U.S.A.

- Delfiner, P. 1976: Linear estimation of nonstationary spatial phenomena. In: Guarascio, M.; David, M.; Huijbregts, C. (eds.) Advanced geostatistics in the mining industry, NATO-ASI, Ser. C24, Reidel, Dordrecht, The Netherlands
- Gambolati, G.; Volpi, G. 1979: A conceptual deterministic analysis of the kriging technique in hydrology. *Water Resour. Res.* 15, 625-629
- Gambolati, G.; Galeati, G. 1987: Comment on 'Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels.' by Neuman, S.P.; Jacobson, E.A. *Math. Geol.* 19, 249-257
- Golub, G.H.; van Loan, C.F. 1983: Matrix computations. Baltimore: John Hopkins University Press, Maryland, U.S.A.
- Journel, A.G.; Huijbregts, Ch.J. 1978: Mining geostatistics. London: Academic Press
- Loaiciga, H.A.; Mariño, M.A. 1987: Long-term feasibility analysis of irrigated agriculture in the San Joaquín Valley, California. *Procs. Intl. Symp. Water for the future: Water Resources Developments in Perspective*, Vol. 2, 275-283. In: Wunderlich, W.O.; Prins, J.E. (eds.) Rome, Italy, April 6-11, published for the Intl. Assoc. Hydraul. Res. by Balkema, A.A., Rotterdam and Boston
- Matheron, G. 1973: The intrinsic random functions and their applications. *Advanced in Applied Probability*. 5, 439-468
- Neuman, S.P.; Jacobson, E.A. 1984: Analysis of nonintrinsic spatial variability by residual kriging with applications to regional groundwater levels. *Math. Geol.* 16, 499-521
- Patterson, H.D.; Thompson, R. 1971: Recovery of interblock information when block sizes are unequal. *Biometrika*. 58, 545-554
- Rao, C.R. 1971: Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Analysis*. 1, 445-456
- Rao, C.R. 1972: Estimation of variance and covariance components in linear models. *J. Am. Statist. Assoc.* 67, 112-115
- Rao, C.R. 1973: Linear statistical inference and its applications, 2nd ed., New York: John Wiley
- Ripley, B.D. 1981: Spatial statistics. New York: John Wiley
- Theil, H. 1971: Principles of econometrics. New York: John Wiley
- Yeh, W.W-G. 1986: Review of parameter identification procedures in groundwater hydrology. The inverse problem. *Water Resour. Res.* 22, 95-108

Accepted April 25, 1988.