I. INTRODUCTION

Hydrologists often deal with spatially distributed phenomena and the analysis of spatially distributed data. In hydrometeorology, the study of the spatial correlation of rainfall is required to derive good estimates of mean areal precipitation from point measurements, to replace missing data, and to design effective rainfall networks [see, e.g., Gandin, 1965, 1970; Rodriguez-Iturbe and Mejia, 1974; Bras and Rodriguez-Iturbe, 1976; Chua and Bras, 1982]. In hydrogeology, the study of spatial structures (trends and correlation of the residuals) of hydraulic head, pollutant concentration, and parameters, such as coefficients of transmissivity and storativity, is required for the development of hydrogeologic maps, the evaluation of groundwater resources, and the solution of the inverse problem [see, e.g., Delhomme, 1979; Gambolati and Volpi, 1979; Olea, 1975a, b; Kitanidis and Vomvoris, 1983].

Much of the interest in the application of statistical methods in the study of spatially distributed phenomena in hydrology has been motivated by the works of Gandin [1965] ('objective analysis') and Matheron [1971] ('kriging'). These works mainly deal with the application of linear minimum variance unbiased estimators in the case of spatially distributed variables. Matheron's theory of regionalized variables and its application in the field known as geostatistics have received much attention in geology, mining, and hydrology. However, before estimation can be applied, the structure of the field must be identified. In the framework of linear minimum variance unbiased estimation, structure selection is equivalent to the estimation of the first two moments of the fields of interest. The functional forms of the mean and the covariance function must be selected and their parameters be statistically estimated from available data and, possibly, prior information. In the geostatistical approach, the model is described by a drift (or 'trend') and a variogram.

This work deals with the problem of estimation of parameters of multidimensional spatial structures from available data. The problems of trend and covariance fitting have received less attention in the geostatistical literature and are far less well understood than linear minimum variance unbiased estimation ('kriging'). The focus of this paper is on the estimation of the parameters of polynomial generalized covariance functions [Matheron, 1973]. These are useful models describing the spatial structure of stationary increment random fields and have found applications in stochastic interpolation [Delfiner, 1976], in rainfall analysis and network design [Chua and Bras, 1982; Hughes and Lettenmaier, 1981], and in the study of spatial variability in groundwater hydrology [Chirin and Dagan, 1980; Gutjahr and Gelhar, 1981]. However, the proposed methodologies could be extended to the identification of other covariance functions.

Three parameter estimation procedures are developed and presented in this work: (1) maximum likelihood, (2) minimum variance unbiased quadratic, and (3) minimum norm.

The advantages and limitations of each method are evaluated, and their application is illustrated through an example. The first two estimators are derived on the assumption that the data are jointly Gaussian distributed. The third method employs the normality assumption only in the calculation of the variances of the parameter estimates. Procedures for non-Gaussian data must be developed, but it is suggested that the Gaussian case may be the most useful. To begin with, linear minimum variance estimation procedures are most appropriate when the data (possibly after a transformation) are nearly Gaussian. Furthermore, the advantages of relaxing the normality assumption may, in many real-world applications, be outweighed by the statistical error introduced in the estimation from effectively small samples of high moments (such as third and fourth moments). Finally, in our experience, the algebra and the computations are significantly more cumbersome in the non-Gaussian case, and this additional effort may be disproportionately large compared to the potential improvement. For these reasons, it will be assumed in this paper,
whenever needed, that the data follow a Gaussian distribution. This work focuses on the well-defined algorithmic problem of estimation of parameters given the model and the data and discusses problems of implementation of parameter estimation algorithms in the identification of spatial structures of hydrologic data.

2. A Brief Review of Available Methods

The most common approach is to plot the experimental covariance function or variogram and to select the model which appears to reproduce the behavior of the data in the best possible way. Sometimes, the model is selected based on theoretical considerations [e.g., Chiril and Dagan, 1980; Mizell et al., 1982]. Parameters are fitted, often in an ad hoc way, by preserving values of the covariance function or variogram at some characteristic distances. Another approach is to estimate the parameters which preserve moments of the data. While some authors [e.g., Journel and Huijbregts, 1978] are right in warning against the blind use of any automatic calibration procedure, the usefulness of procedures which utilize the capabilities of the modern computer to assist the modeler to fit a model to available data is also widely recognized.

For variogram estimation, a direct approach is through the moment estimator:

\[ \hat{\gamma}(h_j) = \frac{1}{2N_j} \sum_{i=1}^{N_j} [\gamma(x_i) - \gamma(x_i + h_j)]^2 \]

\[ h_j \leq |x_i - x_i'| < h_j' \]

(1)

To apply this approach, the axis of distance is divided into a number of intervals, each interval represented by one point \( h_j \). The \( j \)th interval is \( (h_j, h_j') \) and contains \( N_j \) pairs of measurements of the regionalized variable \( y \), function of the vector of spatial coordinates \( x \). Obviously, by increasing the length of the intervals one can trade off resolution for accuracy. Issues of robustness have been discussed by Cressie and Hawking [1980], who found that when the data include outliers (or follow heavy-tailed distributions) the \( M \) estimators [e.g., Hogg, 1974] seem to perform best.

In the case of parametric models, in particular those described by polynomial generalized covariance functions [Matheron, 1973], parameters can be estimated through weighted least squares [Delfiner, 1976, p. 61]. In this approach, first authorized combinations (or generalized increments) of the measurements are formed from the original data \( y \):

\[ z_i = \sum_j \hat{\lambda}_i y_j \]

such that the variance of the authorized combinations \( z_i \) be estimated from the generalized covariance \( \hat{\gamma} \):

\[ E(z_i^2 | \theta) = \sum_{i,j} \hat{\lambda}_i \hat{\lambda}_j \hat{\gamma}(x_i - x_j | \theta) \]

(3)

where \( \theta \) is the vector of the parameters. The weights \( \hat{\lambda}_i \) can be generated following one of the methods described by Starks and Fang [1982]. The parameters are estimated by minimizing a weighted sum of squares of the differences between the observed squares of the authorized combinations.

\[ z_i^2 = \sum_{i,j} \hat{\lambda}_i \hat{\lambda}_j y_i y_j \]

(4)

and their expected values \( E(z_i^2 | \theta) \). That is, the criterion of performance is

\[ \text{min} \sum_{i,j} W_i [z_i^2 - E(z_i^2 | \theta)]^2 \]

(5)

Note that if the field belongs to the class of intrinsic random functions of order \( k \) (IRF-\( k \)), \( E(z_i^2 | \theta) \) is linear in \( \theta \). As a consequence, the problem becomes equivalent to estimating the parameters of a linear regression with the well-known computational advantages. This approach is simple and computationally efficient but neglects correlations among the `data points' \( z_i^2 \) and accounts for heteroscedasticity only approximately through empirical selection of the weights.

Kajfissas and Bros [1981] describe an iterative regression approach based on Delfiner's [1976] work. The coefficients of the regression are calculated by minimizing the expression of (5) with unit weights. First, the coefficients are determined by creating generalized increments from \( \hat{\gamma}(h) = \| h \| \). Then these coefficients are used to create new generalized increments, and the procedure is repeated until the coefficients converge. A variant of the same iterative approach utilizes the cross-validation concept to fit the parameters so that the sum of the squared prediction errors is minimized [Davis and David, 1978; Hughes and Lettenmaier, 1981]. It is, however, rather difficult to analyze and understand the properties of these procedures. Hughes and Lettenmaier [1981] have suggested that the estimates are biased. It might also be emphasized that none of the aforementioned techniques provide measures of accuracy of the estimated parameters.

Model selection and parameter estimation are particularly important when the model is going to be employed to simulate the response of the system under conditions different from the conditions under which the data were collected, in calculating the accuracy of estimates, and in network design. For example, in the case of rainfall network design, Rodriguez-Iiturbe and Mejea [1974] have found that '... it makes a great difference from the point of view of network design how the actual fitting of the parameters was performed.' In such cases it is important to select the model which is most consistent with data and the objectives of modeling, to fit its parameters in the best possible way, to detect data problems, and to evaluate goodness-of-fit criteria. In particular, the estimation error of the parameters should be evaluated, so that the sensitivity of decisions to parameter variability may be evaluated.

3. The Polynomial Generalized Covariance Function

An important and widely applicable class of nonstationary random functions (or random fields) is the class of random functions with stationary increments. Consider an \( d \) dimensional field \( y \). A linear combination of \( n \) sample values of a realization of the random function \( y \):

\[ z_i = \sum_{i=1}^{n} \hat{\lambda}_i y_i \]

(6)

defines an increment \( z_i \) with properties which depend on the properties of the original function \( y \) and the coefficients \( \hat{\lambda}_i \). The basic idea is that the coefficients \( \hat{\lambda}_i \) may be chosen so that the \( z_i \) increments form a stationary process even though the \( y \) function is nonstationary.

Of central interest are cases where the nonstationary function \( y \) includes polynomials of order \( k \) (\( k = 0 \) for a constant). Matheron [1973] examined such random functions which he called intrinsic random functions of order \( k \) (IRF-\( k \)). For IRF-\( k \), \( z \) is called a generalized increment of order \( k \) when the coefficients \( \hat{\lambda} \) satisfy the conditions:

\[ \sum_{i=1}^{n} \hat{\lambda}_i x_i^p_1 x_i^p_2 \cdots x_i^p_r = 0 \]

(7)
for all nonnegative integers \( p_1, \ldots, p_d \), such that \( p_1 + \cdots + p_d \leq k \). The reader is reminded that \( d \) is the dimension of the space and \( k \) is the order of the polynomials which must be filtered out in forming the increments in (6).

Thus for the one-dimensional space (\( d = 1 \)) the conditions are as follows for \( k = 0, 1, 2 \):

For \( k = 0 \)
\[
\sum_{i=1}^{n} \lambda_i = 0
\]

For \( k = 1 \)
\[
\sum_{i=1}^{n} \lambda_i = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1} = 0 \tag{8a}
\]

For \( k = 2 \)
\[
\sum_{i=1}^{n} \lambda_i = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1} = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1}^2 = 0 \tag{8b}
\]

For the more common case of random functions on the plane (\( d = 2 \)) the conditions for \( k = 0, 1, 2 \) are

For \( k = 0 \)
\[
\sum_{i=1}^{n} \lambda_i = 0
\]

For \( k = 1 \)
\[
\sum_{i=1}^{n} \lambda_i = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1} = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i2} = 0
\]

For \( k = 2 \)
\[
\sum_{i=1}^{n} \lambda_i = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1} = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i2} = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1}^2 = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i2}^2 = 0, \quad \sum_{i=1}^{n} \lambda_i x_{i1} x_{i2} = 0
\]

Matheron [1973] has shown that a valid isotropic generalized covariance function is the polynomial. According to Delfiner [1976] the polynomial generalized covariance can be given by

\[
g(h) = \sum_{p=0}^{k} (-1)^{p+1} \frac{a_p}{(2p + 1)!} \Gamma(d/2)p! \sqrt{\pi} \Gamma((2p + d + 1)/2) |h|^{2p+1} \tag{9}
\]

with coefficients which satisfy

\[
\sum_{p=0}^{k} a_p k^{-p} \geq 0 \tag{10}
\]

for any arbitrary nonnegative number \( t \). For example, for \( k = 0 \) the generalized covariance can be represented as

\[
g(h) = -\theta |h|
\]

where \( \theta \) is positive number. This generalized covariance function is appropriate for fields which have random walk realizations. Usually, the term \( c \delta(|h|) \), where \( \delta \) denotes Kronecker's delta and \( c \) is a nonnegative parameter, is added in the expression of the polynomial generalized covariance. This term represents small-scale variability referred to as the 'nugget effect' [see Journel and Huijbregts, 1978].

The reader is reminded that the generalized covariance function is not positive definite but leads to legitimate second moments of generalized increments or authorized combinations. In this sense, the polynomial covariance function is conditionally positive definite, provided that (10) holds. An important practical advantage of the polynomial generalized covariance function given by (9) is that it is linear in the parameters. Thus statistical inference from available data is somewhat facilitated. In particular, if the criterion is least squares, the parameter estimation problem can be solved in computationally very efficient ways which are similar to linear regression.

We will now proceed to the estimation of the parameters of the polynomial generalized covariance function (PGCF) of order \( k \) (assumed, for the time being, known). The PGCF (including the nugget effect term) can be represented as

\[
g(h) = \sum_{i=1}^{m} \theta_i g(|h|) \tag{11}
\]

where \( m \) is the number of unknown parameters, \( \theta_i \) are the parameters to be inferred, and \( g(h) \) are known functions. It is assumed that available observations form the vector

\[
y = (y_1, y_2, \cdots, y_n)^T
\]

However, instead of working with the original data, it is more convenient and appropriate to work with generalized increments or authorized combinations of the data. A linear transformation of the original data vector \( y \) into a vector of stationary data increments \( z \):

\[
z = Ty
\]

is sought, where \( T \) is a transformation matrix. \( T \) must be selected so that the generalized increments satisfy the conditions given by (7). In other words, \( T \) must be selected so that polynomial trends of order \( k \) are filtered out of the data. There are apparently many methods available to generate generalized data increments. Of particular interest is a method proposed by Delfiner [Starks and Fang, 1982]. Assume, for example, that the order of the intrinsic function describing a two-dimensional field is \( k = 1 \). Consider the regression equation which fits a linear function (a plane) to the data:

\[
y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \epsilon_i \quad i = 1, \cdots, n \tag{14}
\]

where \( b_0, b_1, \) and \( b_2 \), are the regression coefficients, \( x_{i1} \) and \( x_{i2} \) are the spatial coordinates of the \( i \)th data point, and \( \epsilon_i \) is the regression residual. In standard vector notation of the regression problem [see Draper and Smith, 1966],

\[
y = Xb + \epsilon \tag{15}
\]

The least squares estimate of the parameters is

\[
b = (X^T X)^{-1} X^T y \tag{16}
\]

and the fitted values are

\[
y = (X^T X)^{-1} X^T y \tag{17}
\]

The fitting residuals (or errors) are

\[
z = y - \hat{y} = (I - X(X^T X)^{-1} X^T)y \tag{18a}
\]

or

\[
z = Ty \tag{18b}
\]

A property of the fitting residuals vector \( z \) is that it is orthogonal with each of the columns of the matrix \( X \), that is, \( X^T z = 0 \)
0. In this example,

\[ \sum_{i=1}^{n} z_i = 0 \]
\[ \sum_{i=1}^{n} z_i x_{it} = 0 \]
\[ \sum_{i=1}^{n} z_i x_{it} = 0 \]

Consequently, \( z \) is one set of generalized increments. The matrix \( T \) which transforms the original data \( y \) to data increments \( z \) has some interesting properties. First, it can be easily verified that

\[ TX = 0 \]

which verifies that the transformation of (18) filters out polynomial trends of the assumed order. Note also that \( T \) is symmetric and idempotent, i.e., \( T = T^2 = T^3 = \cdots \). Such matrices are called projection matrices (see Halmos [1957] or Lawson and Hanson [1974]) and have some ostensible similarities to the unit matrix. In fact, for the incremental data

\[ z = T \]

It can also be shown that if the rank of the matrix \( X \) is \( p \), the number of conditions imposed by the intrinsic function (in our example \( p = 3 \)), then the eigenvalues of \( X (X^T X)^{-1} X^T \) are either 0 or 1 and the number of nonzero eigenvalues (which is equal to the rank of the matrix) is \( p \). Consequently, for \( n \) original data points, \( T \) has \( N = n - p \) eigenvalues equal to 1 and \( p \) eigenvalues equal to zero. Consequently, \( p \) of the increments can be dropped as linearly dependent on the rest. The estimates of the parameters and the calculated estimation error covariance matrix will not be affected by the increments dropped.

The data increments which will be employed in the estimation can thus be generated from

\[ z = \Lambda y \]

where \( \Lambda \) is an \( N \times n \) matrix of rank \( N \) which is derived from matrix \( T \) by dropping \( p \) rows. The moments of \( z \) are

\[ E(z) = 0 \]
\[ E(zz^T) = \Lambda E(y y^T) \Lambda^T \]

By replacing \( E(y y^T) \) by the polynomial generalized covariance

\[ E(y y^T) = K = \sum_{i=1}^{m} K_i \theta_i \]

then

\[ E(z z^T) = \sum_{i=1}^{m} (\Lambda K_i \Lambda^T) \theta_i = \sum_{i=1}^{m} Q_i \theta_i \]

where \( \theta_1, \theta_2, \cdots, \theta_n \) are the parameters to be estimated and \( K_1, K_2, \cdots, K_m \) are symmetric matrices which can be calculated before the estimation of the parameters. Note that \( K_i \) has dimensions \( n \times n \) and its \( jk \) element is equal to \( g_i(h_{jk}) \), where \( h_{jk} \) is the distance between points where measurements \( y_j \) and \( y_k \) were taken. \( Q \) has dimensions \( N \times N \).

4. **Maximum Likelihood Estimation**

The joint probability density function of \( N \) measurements from a Gaussian field with zero mean is given by

\[ p(x \mid \theta) = (2\pi)^{-N/2} |Q|^{-1/2} \exp \left( -\frac{1}{2} x^T Q^{-1} x \right) \]

where \( z \) is the vector with elements the \( N \) available measurements, verticals denote determinant, and

\[ Q = E[zz^T \mid \theta] \]

The vector of parameters will be determined through maximum likelihood estimation or the minimization of the negative log likelihood:

\[ L(z \mid \theta) = -\ln p(z \mid \theta) = \frac{N}{2} \ln (2\pi) + \frac{1}{2} \ln |Q| + \frac{1}{2} z^T Q^{-1} z \]

The derivative of the negative log likelihood with respect to a scalar parameter \( \theta_j \) is

\[ \frac{\partial L}{\partial \theta_j} = \frac{1}{2} \text{Tr} (Q^{-1} Q_j) - \frac{1}{2} z^T Q^{-1} Q_j Q^{-1} z \]

where the following relations [Schweppes, 1973] were used:

\[ \frac{\partial}{\partial \theta_j} \ln |Q| = \text{Tr} \left[ Q^{-1} Q_j \right] \]
\[ \frac{\partial}{\partial \theta_j} Q^{-1} = -Q^{-1} Q_j Q^{-1} \]

For the solution, a gradient-based iterative method for the minimization of the negative log likelihood function can be employed. The basic iteration is

\[ \theta_{i+1} = \theta_i - \rho_i R_i s_i \]

where \( \theta_i \) is the vector of parameters in the \( i \)th iteration and \( s_i \) is the vector gradient of the negative log likelihood function, i.e.,

\[ s_i = \frac{\partial L}{\partial \theta} \bigg|_{\theta = \theta_i} \]

\( \rho_i \) is a scalar step size parameter which can be chosen to ensure that

\[ L(\theta_{i+1}) < L(\theta_i) \]

In the method of steepest descent, \( R_i \) is the unit matrix, while in Newton's method, \( R_i \) is the inverse of the second derivative (Hessian) of \( L \) estimated at the \( i \)th iteration. The rate of convergence of the method of steepest descent is usually agonizingly slow near the optimum, while the costs associated with the computation of second derivatives required for Newton's method are usually too high. In quasi-Newton methods [see Luenberger, 1973], \( R_i \) is an approximation to the inverse of the second derivative matrix

\[ R_i \simeq (\partial^2 L / \partial \theta^2)^{-1} \bigg|_{\theta = \theta_i} \]

In the method of scoring (Gauss-Newton), \( R_i \) is approximated by the inverse of the Fisher information matrix at the \( i \)th iteration. i.e., \( R_i = M_i^{-1} \). The Fisher information matrix is

\[ M_i = E[\partial^2 L / \partial \theta^2] \bigg|_{\theta = \theta_i} = \text{Tr} \left( \frac{\partial L}{\partial \theta} \left( \frac{\partial L}{\partial \theta} \right)^T \right) \bigg|_{\theta = \theta_i} \]

The \((j, k)\) element of the Fisher information matrix is

\[ M(j, k) = \text{Tr} \left( Q^{-1} Q_j Q^{-1} Q_k \right) \]

which can be computed (see Appendix A) by making use of Gaussian moment factoring (for the estimation of higher moments, to be

\[ M(j, k) = \frac{1}{2} \text{Tr} \left( Q^{-1} Q_j Q^{-1} Q_k \right) \]
The inverse of the Fisher information matrix at the optimum (i.e., at the last iteration) is, for unbiased estimates, the Cramer-Rao lower bound to the estimation error covariance matrix of the parameter vector:

\[ M_i^{-1} \leq E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \]  

(40)

where \( \theta \) is the vector of the actual parameters and \( \hat{\theta} \) is the vector of the parameter estimates. In this work, the inverse of the Fisher information matrix is used as an approximation to the covariance matrix of the error of the maximum likelihood estimates.

Maximum likelihood estimates are asymptotically (in practice, for reasonably large samples) unbiased and of minimum variance. They are also asymptotically Gaussian distributed. Tests on the estimates of the parameters can thus be readily performed. Convergence is usually rapidly achieved. Experience has shown that in many cases when the assumption of normality is not strictly true, the described methodology still gives near-optimal (in a quadratic sense) results.

Note that the numerical values of the estimates are not affected by the addition to the original data of any arbitrary polynomial trend of order up to \( k \). That is, if \( y \) were replaced by \( y + Xb \), where \( X \) is given by (15) and \( b \) is any arbitrary set of coefficients, the estimates would remain the same.

An interesting property of maximum likelihood estimates is the following. Consider the statistic

\[ \chi^2 = z^T \hat{Q}^{-1} z \]  

(41)

which follows a chi-squared distribution with \( N \) degrees of freedom. Consequently, it has mean \( N \) and variance \( 2N \). The maximum likelihood solution satisfies (30):

\[ \text{Tr}(\hat{Q}^{-1}Q_j - Q_j\hat{Q}^{-1}Q_j\hat{Q}^{-1}zz^T) = 0 \quad j = 1, \ldots, m \]  

(42)

where the circumflex denotes estimated value and

\[ \hat{Q} = \sum_{j=1}^{m} \theta_jQ_j \]

Multiplying the \( j \)th equation by \( \theta_j \) and summing up,

\[ \text{Tr}\left(\hat{Q}^{-1}\sum_{j=1}^{m} \theta_jQ_j - \hat{Q}^{-1}\sum_{j=1}^{m} \theta_jQ_j\hat{Q}^{-1}zz^T\right) = 0 \]  

(43a)

or

\[ \chi^2 = z^T\hat{Q}^{-1}z = N \]  

(43b)

Since \( \text{Tr}(I) = N \), where \( I \) is an \( N \times N \) unit matrix. Consequently, when the estimates of the parameters of the polynomial generalized covariance are maximum likelihood, the \( \chi^2 \) statistic is fitted to its mean value.

It is of interest to examine the particular case where only one parameter is estimated, all other parameters being fixed to zero. In this case, \( Q = Q_j\theta_j \), where \( \theta_j \) is the parameter to be estimated, and (30) can be solved without iterations to provide the maximum likelihood estimate

\[ \hat{\theta}_j = z^TQ_j^{-1}z/N \]  

(44)

Taking expected values

\[ E(\hat{\theta}_j) = E[\text{Tr}(Q_j^{-1}zz^T)]/N = \theta_j \]  

(45)

which means that the estimate given by (44) is unbiased for any size of the sample and, in fact, independently of the Gaussian assumption. The variance of the estimate is

\[ E[(\hat{\theta}_j - \theta_j)^2] = \frac{1}{N^2} \text{Tr}(Q_j^{-1}zz^T) \text{Tr}(zz^TQ_j^{-1}) - \theta_j^2 \]  

(46)

Making use of Gaussian moment factoring for the calculation of the fourth moment,

\[ E[(\hat{\theta}_j - \theta_j)^2] = \frac{2\theta_j^2}{N} \]  

(47)

This variance is equal to the inverse of the Fisher information, given by (39), indicating that the maximum likelihood estimate is, in this particular case, efficient or minimum variance for any value of \( N > 0 \). The estimate is consistent; that is, as \( N \to \infty \), the estimation error tends to zero. Note that the variance of the estimation error depends only on the value of the parameter and the number of generalized increments in the sample. Finally, under the Gaussian assumption, \( \theta_j/\hat{\theta}_j \) is chisquared distributed with \( N \) degrees of freedom. Confidence intervals can thus be readily computed.

5. Minimum Variance Unbiased Quadratic Estimation

Consider that the parameters must be estimated as quadratic functions of the authorized increments of the data:

\[ \hat{\theta}_j = z^TF_jx \]  

(48)

where \( F_j \) are \( N \times N \) matrices to be selected according to the following specifications.

**Unbiasedness.** The following condition must be satisfied:

\[ \theta_j = E(\hat{\theta}_j) = E[\text{Tr}(F_jzz^T)] = \sum_{i=1}^{m} \text{Tr}(F_jQ_i) \theta_i \]  

(49)

\[ j = 1, \ldots, m \]

where \( \theta_j \) denotes the actual parameter and \( \hat{\theta}_j \) denotes its estimate. For these equations to have unique solutions,

\[ \text{Tr}(F_jQ_i) = 1 \quad j = i \]  

(50)

\[ \text{Tr}(F_jQ_i) = 0 \quad j \neq i \]

**Minimum variance.** The estimation-error variance for parameter \( \theta_j \) is

\[ E((\hat{\theta}_j - \theta_j)^2) = E(z^T(F_jx - \theta_j)^2) = E[\text{Tr}(F_jzz^TF_jxzz^T)] - \theta_j^2 \]  

(51)

The matrix \( F_j \) must be selected so that this variance is minimum subject to the unbiasedness conditions. The useful case of Gaussian distributed data will be discussed.

Applying Gaussian moment factoring and taking into account the unbiasedness conditions,

\[ E((\hat{\theta}_j - \theta_j)^2) = 2 \text{Tr}(F_jQ_FQ_j) \]  

(52)

This variance must be minimized subject to the unbiasedness conditions, which are equality constraints. The Lagrangian is given by

\[ 2 \text{Tr}(F_jQ_FQ_j) - \sum_{i=1}^{m} \lambda_i \text{Tr}(F_jQ_i) - \lambda_j (\text{Tr}(F_jQ_i) - 1) \]  

(53)

where \( \lambda \) is the vector of Lagrange multipliers. Necessary conditions for optimality are

\[ 4Q_FQ_j - \sum_{i=1}^{m} \lambda_i Q_i = 0 \]

(54a)

\[ \text{Tr}(F_jQ_i) = 1 \]

(54b)

\[ \text{Tr}(F_jQ_i) = 0 \quad i = 1, \ldots, m \quad i \neq j \]

(54c)
where in the derivation, the following relations were taken into account [Athans and Schewepe, 1965]:

\[
\frac{\partial}{\partial X} \text{Tr}(AXBX) = A'X'B + B'TX'AT
\]

\[
\frac{\partial}{\partial X} (AX) = A'
\]

where \(A\), \(B\), and \(X\) are square matrices of appropriate dimensions.

It is obvious from (52) that the optimum \(F_j\) will be a function of \(Q\) and consequently of the yet unknown parameters \(\theta\). In such cases, a common procedure [see Rao, 1973] is to approximate \(Q\) by its best prior estimate \(\hat{Q}_0\) multiplied by an arbitrary constant. The solution can then be outlined as follows. Solving for \(F_j\)

\[
F_j = \frac{1}{4} \sum_{i=1}^{m} \lambda_i \hat{Q}_0^{-1}Q_0^{-1}J_i
\]

then

\[
\sum_{i=1}^{m} \text{Tr} \left( Q^{-1}_0 \hat{Q}_0^{-1}Q^{-1}J_i \right) \lambda_i = 0 \quad i = 1, \ldots, m \quad i \neq j \quad (56a)
\]

\[
\sum_{i=1}^{m} \text{Tr} \left( Q^{-1}_0 \hat{Q}_0^{-1}Q^{-1}J_i \right) \lambda_i = 4 \quad (56b)
\]

From this system of \(m\) linear equations with \(m\) unknowns (equations (56)) the Lagrange multipliers \(\lambda_i\) are determined. Then \(F_j\) is calculated from (55), and the parameters are calculated from (48). The same procedure is performed for every \(j\).

Summarizing the properties of this estimator:

1. The parameter estimates are unbiased. One can easily verify that

\[
E(\hat{\theta}_j) = E(\text{Tr}[F_jx'x]) = \sum_{i=1}^{m} \text{Tr} (F_jQ)\theta_i = \theta_j
\]

This result is independent of the validity of the Gaussian assumption or the assumed \(Q_0\).

2. The parameter estimates are minimum variance. (In practice, this property may be somewhat approximate since the form of the estimator was assumed quadratic and the variance which was minimized was calculated assuming Gaussian data and prior estimates of the parameters.)

The estimation error covariance matrix can be easily derived by making use of the unbiasedness conditions and Gaussian moment factoring:

\[
E[(\hat{\theta}_j - \theta_j)(\hat{\theta}_j - \theta_j)] = 2 \text{Tr}(F_jQF_jQ)
\]

3. The parameter estimates are invariant to the addition to the original data of any arbitrary polynomial trend of order up to \(k\).

In applications, prior estimates of the ratios of the parameters, \(\theta_2/\theta_1, \ldots, \theta_m/\theta_1\), must be available. The parameters given by the minimum variance unbiased quadratic estimation procedure have been found to be not very sensitive to the exact values of these ratios, and the procedure performs well. An alternate procedure is, after applying the minimum variance unbiased quadratic estimation procedure, to calculate a new \(Q_0\) and to repeat the calculation in an iterative fashion until convergence is achieved. This procedure converges quite rapidly and gives results identical to maximum likelihood estimates. (A mathematical proof is given in Appendix C.) However, additional studies will be required to evaluate the advantages and disadvantages of each procedure. In particular, the small-sample properties of each estimation procedure must be studied further.

Note that in the particular case where only one parameter is estimated, all other parameters being fixed to zero, the method of minimum variance quadratic estimation does not require prior estimates of the parameters and gives exactly the same results as the method of maximum likelihood.

The minimum variance unbiased quadratic estimation procedure can, in principle, be followed in the more general case of non-Gaussian data. The main practical difficulty, however, is that the estimation of the fourth joint moment of data involves large sampling errors. In fact, in many practical situations the advantages of relaxing the assumption of Gaussian-distributed data may be outweighed by the sampling error introduced in the statistical estimation from limited data of the fourth moment properties of the field. Furthermore, the computational algorithms are, in general, much more complicated than in the Gaussian case. In many cases it may not appear worthwhile to apply complicated and computationally expensive minimization algorithms when the objective function has very large sampling variability. In many real-world applications (and since the application of linear minimum variance unbiased linear estimation theory is predicated on the tacit assumption that the data are near Gaussian), it may be more expedient to seek a transformation which will make the data more Gaussianlike. Of course, development of estimation procedures for non-Gaussian data is important and is certainly encouraged.

6. MINIMUM NORM ESTIMATES

The proposed maximum likelihood and minimum variance unbiased quadratic estimators have, as already discussed, good accuracy properties. However, they also involve a relatively high computational cost, especially when many measurements are available. Much of this cost is associated with the inversion of the covariance matrix of the data and the need for iterations. In this section, a least squares estimator, appropriate for preliminary data analysis, will be presented.

Consider the covariance matrix of

\[
z = Ty
\]

where \(y\) is the vector of the original measurements and \(T\) is the projection matrix defined by (18). According to the model, the covariance of \(z\) is given by

\[
E(zz') = TE(yy')T = TKTK
\]

where \(K\) is an \(n \times n\) matrix whose \(ij\)th element is the generalized covariance between point \(i\) and point \(j\).

According to the data, the covariance of \(z\) is given by

\[
Ty'y'T
\]

A reasonable fitting criterion is to select the parameters which minimize the difference between the matrices \(TKTK\) and \(Ty'y'T\) in a quadratic sense. Using matrix notation and, in particular, the notion of the trace (see Appendix B), the criterion is

\[
\min \left\{ \text{Tr} \left[ (TKTK - Ty'y'T)TKTK - Ty'y'T) \right] \right\}
\]

This trace may be interpreted as the (Euclidean) length or norm of a matrix. Consequently, the criterion may be called 'minimum norm.' Note that this criterion is not the same with
the least squares criterion given in (5). Using the properties of the projection matrix, this criterion may be written as

$$\min_{\theta} \left[ \text{Tr} \left[ T(K - yy^T)T(K - yy^T) \right] \right]$$

(61b)

Note that $K = \Sigma K_i \beta_i$, where $\theta_i$ are the parameters to be estimated and $K_i$ are known $n \times n$ matrices. The necessary conditions for optimality are that

$$2 \text{Tr} \left[ TK_i \left( \sum_{i=1}^{m} K_i \beta_i - yy^T \right) \right] = 0 \quad i = 1, \ldots, m$$

(62)

or

$$\sum_{i=1}^{m} \text{Tr} \left[ TK_i TK_i \beta_i \right] = y^T TK_i Ty \quad i = 1, \ldots, m$$

(63)

This is a linear system of $m$ equations with $m$ unknowns which determines the unknown parameters. Let $A$ be the symmetric $m \times m$ matrix with $ij$th element

$$A_{ij} = \text{Tr} \left[ TK_i TK_j \right]$$

(64)

and $e$ the $m$ vector with ith element the quadratic function of the data:

$$e_i = y^T TK_i Ty$$

(65)

The vector of the parameter estimates is then given by

$$\hat{\theta} = A^{-1} e$$

(66)

This estimator has the following interesting properties:

1. The parameter estimates are quadratic functions of the data and are invariant to the addition to the data of any arbitrary polynomial trend of order up to $k$. That is, if the measurements vector were $y + Xb$ instead of $y$, where $b$ is an arbitrary vector of coefficients and $X$ is the matrix of monomials (as already defined, for example, through (15)), then the parameter estimates would not be changed because

$$e_i = (y + Xb)^T TK_i (y + Xb) = y^T TK_i Ty$$

(67)

since $TX = 0$ and $T = T^T$.

2. The estimates are unbiased. To prove this claim, take the expected values of the parameter estimates:

$$E(\hat{\theta}) = A^{-1} E(e)$$

But

$$E(e_i) = E \left[ \text{Tr} \left( TK_i Ty^T \right) \right] = \sum_{i=1}^{m} \text{Tr} \left( TK_i TK_i \right) \beta_i$$

Thus

$$E(e) = A \theta \quad E(\hat{\theta}) = \theta$$

3. The development of the proposed estimator did not involve the assumption of normality or, in fact, probability theory but is based on the minimization of a reasonable quadratic fitting criterion suggested by engineering judgment. The methodology does not involve the inversion of large-dimensional matrices, is noniterative, and has many similarities with ordinary linear regression.

While it has been established that the proposed estimator gives unbiased estimates, an important practical question relates to the accuracy of these estimates. Consider the estimation error covariance matrix of $\hat{\theta}$, $E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T]$. Since $\hat{\theta}$ is a quadratic function of the data, the estimation error covariance matrix involves the fourth moment of the data. For simplicity, it will be assumed in the estimation of the covariance matrix that the fourth moments of the data can be calculated through Gaussian moment factoring from the second moment:

$$\text{Cov} (\hat{\theta}) = 2 \hat{\Sigma}^{-1} C \hat{\Sigma}^{-1}$$

(68)

where $C$ is a symmetric $m \times m$ matrix with $ij$th element:

$$C_{ij} = \text{Tr} \left( TK_i TK_j TK_i TK_j \right)$$

(69)

$$K = \sum_{i=1}^{m} K_i \beta_i$$

Experience with minimum norm estimation indicates that it gives adequate results in many cases, such as when the structure in the data is rather simple. Minimum norm estimation is particularly useful in preliminary or exploratory data analysis. It provides reasonable initial estimates of the parameters which can be further improved through maximum likelihood or minimum variance unbiased quadratic estimation.

7. Problems of Implementation and an Application

The proposed maximum likelihood, minimum variance unbiased quadratic, and minimum norm (least squares) estimators have been applied and tested in a number of cases. Here, some general problems associated with the implementation of the presented algorithms in real-world problems will be presented, illustrated with the analysis of a set of orographic precipitation data.

Conditional Positive Definiteness Conditions

The polynomial generalized covariance function must be conditionally positive definite, and as a consequence, its parameters must satisfy certain conditions (arising from (10)). For example, for the two-dimensional first-order field with white noise,

$$K(h) = \delta(0)\theta_1 + |h|\theta_2 + |h|^2\theta_3$$

(70)

where $\delta$ represents Kronecker’s delta, the three parameters must satisfy the following relations:

$$\theta_1 \geq 0 \quad \theta_2 \leq 0 \quad \theta_3 \geq 0$$

(71)

There is no assurance, however, that the parameters which are estimated from the data, following one of the proposed procedures, will necessarily satisfy these conditions. Of course, these procedures could be modified to enforce these constraints, but the increase in computational cost is likely to be large. More importantly, failure of the fitted parameters to satisfy the positive definiteness conditions in a given application should not be viewed as an isolated ‘algorithmic’ problem. In a practical situation, this failure usually is a telltale sign that the assumed model is inconsistent with the data or overparameterized (i.e., includes effectively unidentifiable parameters), or the data include ‘outliers.’ Practical experience indicates that if such causes are not present, parameter estimates satisfy the conditional positive definiteness conditions even though no special provisions have been made.

An Example

This subsection contains a comparison among the results of various parameter estimation methods applied to the precipitation data (29 measurements) given by Kafritas and Bras [1981]. In Figure 1 the squares of the differences of all pairs of data values have been plotted against their separation distances. The generalized covariance function for these data has
been calculated according to the following methods: (1) raw variogram (equation (1)), (2) iterative regression [Kafritsas and Bras, 1981], (3) weighted regression [Delfiner, 1976], (4) minimum norm, (5) minimum variance unbiased quadratic (MVUQ), and (6) maximum likelihood (ML).

The raw variogram is, essentially, the best quadratic fit to the plotted points (see Figure 1). Kafritsas and Bras [1981] determined, following Delfiner's [1976] approach, that the order of the intrinsic function is one (with the third parameter fixed at zero) with parameters

\[ \theta_1 = 0.00286 \]
\[ \theta_2 = -0.00174 \]

A weighted regression did not give satisfactory results when the weights were taken equal to \((T')^{-1}\) [see Delfiner, 1976, p. 62]. For equal weight, the result was

\[ \hat{\theta}_1 = 0.0128 \]
\[ \hat{\theta}_2 = 0 \]

Estimation errors are not calculated in these methods. The results of the methods proposed in this paper are as follows.

Minimum norm estimation has given

\[ \hat{\theta}_1 = 0.00928 \quad \text{s.e.} (\hat{\theta}_1) = 0.00497 \]
\[ \hat{\theta}_2 = -0.000446 \quad \text{s.e.} (\hat{\theta}_2) = 0.000692 \]

where s.e. denotes standard error of estimation. The correlation between the estimation errors is 0.76.

Minimum variance unbiased quadratic estimation (with \(Q_0\) calculated using the minimum norm estimates) has given

\[ \hat{\theta}_1 = 0.00597 \quad \text{s.e.} (\hat{\theta}_1) = 0.00398 \]
\[ \hat{\theta}_2 = -0.000933 \quad \text{s.e.} (\hat{\theta}_2) = 0.000804 \]

with correlation between errors 0.72.

Minimum variance unbiased quadratic estimation (with \(Q_0\) calculated iteratively until the parameter estimates converged) has given

\[ \hat{\theta}_1 = 0.00236 \quad \text{s.e.} (\hat{\theta}_1) = 0.00293 \]
\[ \hat{\theta}_2 = -0.00171 \quad \text{s.e.} (\hat{\theta}_2) = 0.000930 \]

with correlation between errors 0.72.

Finally, maximum likelihood estimation has given

\[ \hat{\theta}_1 = 0.00237 \quad \text{s.e.} (\hat{\theta}_1) = 0.00293 \]
\[ \hat{\theta}_2 = -0.00171 \quad \text{s.e.} (\hat{\theta}_2) = 0.000930 \]

with correlation 0.72 between standard errors. Note that the ML and the iterative MVUQ estimators, although ostensibly different, have given practically identical results. In all these cases the order of the function was taken equal to 1 with the third parameter fixed to zero.

Experience with kriging indicates that if the purpose of esti-
mation is linear interpolation, then the differences between the results of the iterative regression and ML or MVUQ might have little practical significance. However, it is reasonable to expect that the calculated variances of kriging errors as well as the results of network analyses would be affected. This example also has illustrated that the estimation of the parameters of the generalized covariance function involves sizeable errors. The 90% confidence region (given approximately by an ellipse under the assumption of normally distributed parameter estimates) is shown in Figure 2. It is obvious that the parameters can take a wide range of values. Research must be undertaken to evaluate the effects of parameter uncertainty in problems such as the optimization of rainfall data collection networks or the calculation of the estimation error of mean areal precipitation.

Parameter Identifiability

Starks and Fang [1982] have suggested that this estimation problem is affected by collinearity [see Belsley et al., 1980], resulting in highly unstable and inflated estimates of the parameters which are highly correlated among themselves. This problem will be examined by using the precipitation data of Kaftinsas and Bras [1981]. A straightforward way to evaluate the effects of collinearity is by examining the condition number of the covariance matrix. A simpler and more intuitive way to make this point is geometric, by examining the results of the minimum norm criterion. The matrix of the second derivatives (Hessian) of the surface whose minimum is sought is easily calculated to have elements

$$A_{ij} = \text{Tr} (TK_i TK_j)$$  

(72)

The Hessian is thus not a function of the actual parameters or the values of the measurements but depends only on the geometric position of the measurements, the assumed order of the field, and which of the parameters will be estimated. The parameters can be rescaled so that

$$\text{Tr} (TK_i TK_j) = 1 \quad i = 1, \ldots, m$$

That is, the Hessian has the form of a correlation matrix (rescaling of the parameters enhances numerical accuracy, significantly for high-order models, and has been employed in the computer programs developed for these studies). The Hessian for intrinsic function of order zero is

$$\begin{pmatrix}
1 & -0.375 & 1 \\
-0.375 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix}$$

for intrinsic function of order 1,

$$\begin{pmatrix}
1 & 0.061 & 1 \\
0.061 & 0.325 & -0.874 & 1 \\
1 & 1 & 1
\end{pmatrix}$$

and for intrinsic function of order 2,

$$\begin{pmatrix}
1 & 0.710 & 1 \\
0.710 & 0.81 & 1 \\
0.381 & 0.803 & 1 \\
1 & 1 & 1
\end{pmatrix}$$

Consider now the inverse of the eigenvalues of the Hessian in the three cases:

Zero order 0.73 1.6
First order 0.45 1.43 1.41
Second order 0.42 0.88 1.84 21.5

Geometrically, these values give the radii of curvature of the quadratic fitting criterion in the orthogonal directions. It is obvious that as the order increases, the quadratic surface flattens in some directions, although not as quickly as suggested by Starks and Fang [1982]. (It might be noted that Starks and Fang's [1982] results refer to the weighted regression method, which, in our experience, is affected by collinearity more than the methods which were proposed in this paper.) Estimation of the parameters in these directions is difficult since noise in the data may displace the 'minimum' at great distances along these directions. Numerically, the inverse of $A$ (in (65)) tends to blow up magnifying perturbations in the data. The severity of collinearity, as emphasized by Belsley et al. [1980] and Lawson and Hanson [1974], depends on the intensity of the noise or errors in the data. In our example, estimation of all three parameters for the first-order model has given

$$\hat{\theta}_1 = 0.00495$$

$$\hat{\theta}_2 = -0.00155$$

$$\hat{\theta}_3 = -0.0000116$$

These parameters have been calculated following the minimum norm method, but the other methods did not give much better results. Note that these parameter estimates are unacceptable since they fail to satisfy the constraints associated with the conditional positive definiteness of the generalized covariance function. However, this failure could reasonably be attributed to the large variance associated with the estimation of the parameters. Collinearity, when present, leads to unstable and unreliable parameter estimates. From a practical viewpoint, collinearity can be avoided by reducing the number of parameters to be estimated from the data or incorporating prior information. The most practical approach is to drop some of the parameters. In this example, dropping the third parameter gave acceptable results. A general result, which was confirmed in all examined cases, is that the number of parameters which can be estimated from available data in a statistically meaningful way is rather small. (The work of Yakowitz and Szidarovszky...
[1982] is relevant at this point.) However, simple models with few parameters may be adequate for most real-world applications.

**Effect of Influential Data**

In all applied methods the parameters of the covariance function are essentially estimated as quadratic functions of the data. Consequently, the estimates may be influenced at an inordinately high degree by a few data points which behave differently from the bulk of the data and at the same time, because of their position, have high leverage. Such 'outliers' may result from gross observational or recording errors, and their detection and correction or removal may significantly improve the accuracy of estimation. However, legitimate data points may also be highly influential. In any case, it is important to develop procedures which call the attention of the modeller to the presence of influential data. The modeller can then use judgment and probably seek additional information to decide whether such points should be dropped from or given smaller weight in the analysis. It is felt that in applications, this approach may be more appropriate than formal robust estimation procedures [see Belsley et al., 1980].

In this work, outliers were sought by examining graphic plots of the original data and the standardized (independent standard normal variates) incremental data [see Kitandis and Vamvoudis, 1983, appendix]. The influence of suspected outliers was determined by dropping suspicious measurement one at a time from the data and repeating the estimation of the parameters. In the particular example, outliers did not appear to be a serious problem.

**Model Selection**

The focus of this paper has been on the well-defined algorithmic problem of parameter estimation from available data assuming that the model is known. Selection of the appropriate model is at least as important in determining the success or failure of the modeling effort but cannot be easily reduced to a straightforward algorithmic problem. Considerations which must be taken into account in selecting the appropriate model include (1) the objectives of the study, (2) prior experience with similar problems, and (3) parsimony (the simplest model consistent with available data).

If attention is limited to the class of intrinsic functions of order $k$, then the problem becomes equivalent to selecting the order $k$ and the nonzero parameters. Delfiner [1976] and Davis and David [1978] suggest automated procedures to determine the order $k$ based on a 'validation criterion.' Chua and Bras [1982] found that the estimate of the kriging error depends on the assumed order of the field. The optimization of rainfall data collection networks may also be affected [see Hughes and Lettenmeier, 1982].

To select the order of the function, an ad hoc trial-and-error procedure has been followed in this work. In the absence of prior information, the lowest-order model which adequately describes the spatial structure or interdependence of the data is chosen. In preliminary data analysis, one may fit, through linear regression, polynomial trends of order $k$. In practice, one needs to check only $k = 0, 1, \ldots, 2$. If a polynomial of order $k$ seems to fit (in a statistically convincing way) the data, this is an indication that the order of the intrinsic function may be $k$. The experimental variogram provides useful information about the structure of the data. However, it is rather difficult to read too many details from this plot, especially if a high-order model is appropriate, because these points are correlated among themselves, have unequal variances, and follow a skewed distribution. Cressie and Hawkins [1980] suggest plotting the squared root of the absolute value (rather than the square) of the difference versus the separation distance (see Figure 3). These points are approximately symmetrically distributed.

To find the effect of the assumed order on the estimates of the parameters, the procedures which were developed for this work were also applied for the zero-order model. Using the minimum norm estimation procedure with the zero-order model,

$\hat{\theta}_1 = 0.00603$ s.e. $(\hat{\theta}_1) = 0.0092$

$\hat{\theta}_2 = -0.000716$ s.e. $(\hat{\theta}_2) = 0.000926$

and correlation between estimation errors equal to 0.90. Minimum variance unbiased quadratic and maximum likelihood estimation gave essentially identical results:

$\hat{\theta}_1 = 0.00427$ s.e. $(\hat{\theta}_1) = 0.00295$

$\hat{\theta}_2 = -0.00116$ s.e. $(\hat{\theta}_2) = 0.000658$

with correlation 0.63 between estimation errors. Note that the estimates of the parameters depend on the assumed order of the model.

**8. Conclusions**

The spatial structure or interdependence of hydrologic data can often be described stochastically through polynomial generalized covariance functions [Matheron, 1973]. This class of generalized covariances is analytically attractive because it is linear in the parameters and can be used in a wide variety of applications. In this work, the problem of statistical estimation of the parameters of the polynomial generalized covariance functions has been examined. Three new methods have been presented and applied in an example:

1. Maximum likelihood estimation. The likelihood function of the data is maximized in an iterative fashion. Estimates are asymptotically minimum variance, unbiased, and normally distributed.

2. Minimum variance unbiased quadratic estimation. The parameters are given as quadratic functions of the data determined so that they are minimum variance and unbiased.

3. Minimum norm estimation. The parameters are determined by minimizing a reasonable quadratic fitting criterion. The estimates are unbiased.

The second moment of the estimation error of the parameters is calculated with any of these methods. The first two estimators seem to be, in some respects, the most accurate and well founded methods available for the estimation of the parameters of the polynomial generalized covariance function. In their present form they employ the assumption that the data follow (possibly after a transformation) a Gaussian distribution. Although this assumption can be relaxed, it is believed that in practice, the Gaussian case if of most interest. The computational cost associated with the application of these methods is higher than the (noniterative) weighted regression method, particularly if there are many data points (say, more than 100). However, the computational cost involved in the estimation of the spatial structure of the data remains a very small percentage of the total cost associated with a complete hydrologic or geostatistical study. As a consequence, the applicability of the proposed approaches should not be limited by computational considerations. The maximum likelihood and the minimum variance unbiased quadratic estimators
have very good properties and, in our experience, are clearly more accurate than the (noniterative) weighted regression method. The minimum norm estimator is very simple and efficient and is most appropriate for preliminary data analysis. However, the performance of each procedure must be evaluated further through simulation studies.

The proposed approaches can be extended and applied with other covariance functions with or without drift. Prior information about the parameters can also be incorporated in the estimation procedure with any of the proposed methodologies. The proposed parameter estimation methods can be very useful in evaluating and accounting for the effect of parameter uncertainty in the design of hydrologic data collection networks. Future research must also address problems of hypothesis testing and model selection and validation in the context of the analysis of spatial data, particularly in hydrology. In real-world applications, neither the model nor the data can be considered 'given.' The modeler must proceed cautiously, resolving conflicts between modeling assumptions and data in an iterative way. The development of systematic and efficient methods for the detection of influential data ('outliers') would be particularly useful in practice.

**APPENDIX A: ESTIMATION OF THE FISHER INFORMATION MATRIX FOR GAUSSIAN DATA**

The $k$th element of the Fisher information matrix is

$$
M_k = E\left\{ \frac{\partial L}{\partial \theta_j} \frac{\partial L}{\partial \theta_k} \right\}
$$

$$
= E\left[ \frac{1}{2} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} \right) - \frac{1}{2} z^T Q^{-1} \frac{\partial Q}{\partial \theta_j} Q^{-1} z \right]
$$

$$
- \frac{1}{2} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_k} \right) \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} \right)
$$

$$
+ E \left[ \frac{1}{2} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} z z^T \right) \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_k} Q^{-1} z z^T \right) \right]
$$

(\text{A1})

The estimation thus involves fourth moments of the data. For Gaussian-distributed data these moments can be readily computed. The second term can be calculated as follows. Using the properties of the trace (see Appendix B)

$$
E\left[ \frac{1}{2} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} z z^T \right) \text{Tr} \left( z z^T Q^{-1} \frac{\partial Q}{\partial \theta_k} Q^{-1} \right) \right]
$$

$$
= E \left[ \frac{1}{2} \sum_a \sum_b \sum_c \sum_d \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} Q^{-1} \right)_{ab} z_a z_b z_c z_d \left( Q^{-1} \frac{\partial Q}{\partial \theta_k} Q^{-1} \right)_{cd} \right]
$$

(\text{A2})

But for Gaussian data with zero mean,

$$
E[z_a z_b z_c z_d] = E[z_a z_b] E[z_c z_d] + E[z_a z_c] E[z_b z_d] + E[z_a z_d] E[z_b z_c]
$$

(\text{A3})
Consequently, (A2) becomes
\[ \frac{1}{2} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} \right) \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_k} \right) + \frac{1}{4} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} - Q^{-1} \frac{\partial Q}{\partial \theta_i} \right) \]

Thus
\[ M_{jk} = \frac{1}{2} \text{Tr} \left( Q^{-1} \frac{\partial Q}{\partial \theta_j} - Q^{-1} \frac{\partial Q}{\partial \theta_i} \right) \]  \hspace{1cm} (A4)

**APPENDIX B: TRACES AND INNER PRODUCTS OF MATRICES**

The trace of an \( N \times N \) matrix is defined by
\[ \text{Tr} \left( A \right) = \sum_{i=1}^{N} A_{ii} \]

Two important properties of the trace are
\[ \text{Tr} \left( A + B \right) = \text{Tr} \left( A \right) + \text{Tr} \left( B \right) \quad \text{Tr} \left( AB \right) = \text{Tr} \left( BA \right) \]

The inner product of two \( N \times N \) matrices can be defined in terms of the trace as follows:
\[ \text{Tr} \left( AB^T \right) = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} B_{ij} \]

This is an appropriate inner product because
\[ \text{Tr} \left( AB^T \right) = \text{Tr} \left( BA^T \right) \quad (a: \text{real scalar}) \]

\[ \text{Tr} \left( \left( A + B \right) C^T \right) = \text{Tr} \left( AC^T \right) + \text{Tr} \left( BC^T \right) \]

The inner product of a matrix with itself can be interpreted as the squared euclidean length (or Frobenious norm) of the matrix
\[ \text{Tr} \left( AA^T \right) = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij}^2 \]

In this sense, a matrix least squares criterion is as follows:
\[ \min \text{Tr} \left( \left[ Y - X \beta \right] \left[ Y - X \beta \right]^T \right) \]

Two matrices \( A \) and \( B \) are called orthogonal when \( \text{Tr} \left( AB^T \right) = 0 \). An important relation is that
\[ \text{Tr} \left( ABC \right) = \text{Tr} \left( BCA \right) = \text{Tr} \left( CAB \right) \]

**APPENDIX C: ML AND ITERATIVE MVUQ ESTIMATORS**

First, it will be shown that the covariance matrix of the MVUQ estimation procedure, whose \( j \)th element is given by (57), can be expressed as the inverse of a matrix \( M \) whose \( ij \)th element is
\[ M_{ij} = \frac{1}{2} \text{Tr} \left( Q^{-1} Q_i Q^{-1} Q_j \right) \]

From (56) it can be easily shown that the Lagrange multiplier in the calculation of \( F_j \) is
\[ \lambda_{ij} = 2 \left( M^{-1} \right)_{ij} \]

where \( \left( M^{-1} \right)_{ij} \) is the \( ij \)th element of \( M^{-1} \). Then, from (55),
\[ F_j = \frac{1}{2} \sum_{i=1}^{m} \left( M^{-1} \right)_{ij} Q_i Q^{-1} Q_j Q^{-1} \]

Similarly,
\[ F_i = \frac{1}{2} \sum_{j=1}^{m} \left( M^{-1} \right)_{ij} Q_i Q^{-1} Q_i Q^{-1} \]

Thus
\[ F_j F_i Q = \frac{1}{4} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( M^{-1} \right)_{ij} \left( M^{-1} \right)_{ji} Q_i Q^{-1} Q_i Q^{-1} Q_i Q^{-1} Q_i Q^{-1} \]

Then
\[ F_j F_i Q = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( M^{-1} \right)_{ij} \left( M^{-1} \right)_{ji} Q_i Q^{-1} Q_i Q^{-1} \]

Taking traces and noting that \( M \) and \( M^{-1} \) are symmetric,
\[ \text{Tr} \left( F_j F_i Q \right) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left( M^{-1} \right)_{ij} \left( M^{-1} \right)_{ji} = \frac{1}{2} \left( M^{-1} \right)_{ij} \]

Thus the expression for the covariance matrix of the MVUQ estimator coincides with the inverse of the Fisher information matrix for the ML estimator.

Second, it will be shown that the MVUQ estimates, obtained through iterative application of (56), (55), and (48) satisfies (30), the necessary condition for maximum likelihood. Assuming that the iterative MVUQ has converged, from (C3) and (48), the MVUQ estimates are
\[ \hat{\theta}_j = \text{Tr} \left( F_j z z^T \right) = \frac{1}{2} \sum_{i=1}^{m} \left( M^{-1} \right)_{ji} \text{Tr} \left( Q_i Q_i z z^T \right) \]

Solving in terms of \( \text{Tr} \left( Q_i Q_i z z^T \right) \),
\[ \text{Tr} \left( Q_i Q_i z z^T \right) = 2 \sum_{j=1}^{m} M_{ij} \hat{\theta}_j = \sum_{j=1}^{m} \text{Tr} \left( Q_i Q_i z z^T \right) \]

Hence
\[ \text{Tr} \left( Q_i Q_i z z^T \right) = \text{Tr} \left( Q_i Q_i \right) \]

Consequently, the iterative MVUQ estimates satisfy the necessary conditions for maximum likelihood estimation.

**Acknowledgments.** The assistance of Oscar Lara, Mike Burkart, Bob Lane, Lynn Gelhar, and anonymous reviewers is gratefully acknowledged. Bob Lane assisted with most of the computations and made useful suggestions in the automatization of the developed procedures. Kathy Holstein typed and Bob Hoeksema reviewed the manuscript. The material is based upon work supported by the National Science Foundation under grant CME-8106577.

**REFERENCES**


(Received December 6, 1982; revised April 13, 1983; accepted April 20, 1983.)