

Generalized Covariance Functions in Estimation¹

Peter K. Kitanidis²

I discuss the role of generalized covariance functions in best linear unbiased estimation and methods for their selection. It is shown that the experimental variogram (or covariance function) of the detrended data can be used to obtain a preliminary estimate of the generalized covariance function without iterations and I discuss the advantages of other parameter estimation methods.

KEY WORDS: geostatistics, linear model, best linear unbiased estimation, experimental variogram, restricted maximum likelihood.

INTRODUCTION

The generalized covariance function or GCF (Matheron, 1973) is considered an esoteric concept (Journel, 1986) and its practical significance in geostatistical model development and parameter estimation is not recognized everywhere. One often hears from practitioners that GCF's are hard to grasp or work with; it has been suggested that one should not use models with variable mean because of difficulties in estimating generalized covariance functions; and the question of whether one should use "universal kriging" or "kriging with generalized covariances" is still debated. This confusion is somewhat hard to explain considering that the generalized covariance function is a natural extension of the amiable and familiar to geostatisticians variogram.

The significance of the generalized covariance function idea in stochastic processes, particularly in the role of describing stationary-increment or high-order intrinsic functions, has been discussed in other works. In contrast, I am concerned with fitting empirical models to data for estimation purposes. In this work, I will present my views on GCFs such as my belief that a more descriptive name for the *generalized* covariance function would have been *simplified* covariance function. I will show that having a variable mean facilitates the job of finding the covariance that is needed in kriging, contrary to a widely held belief.

¹Received 30 June 1992; accepted 3 February 1993.

²Civil Engineering, Stanford University, Stanford, California 94305-4020.

Finally, I will discuss parameter estimation methods including variogram analysis of detrended data.

THE MODEL

In linear geostatistics, the regionalized variable $z(\mathbf{x})$ is modeled as the sum of a deterministic part, called the *drift* or the *mean function*, and a *zero-mean stochastic process*:

$$z(\mathbf{x}) = \sum_{k=1}^p \beta_k f_k(\mathbf{x}) + \epsilon(\mathbf{x}) \quad (1)$$

where $z(\mathbf{x})$ is the *regionalized* or *field variable* at *location index* (i.e., array of coordinates) \mathbf{x} ; β_k , $k = 1, \dots, p$, are the *drift coefficients*; $f_k(\mathbf{x})$, $k = 1, \dots, p$, are *base functions* of the location index; and $\epsilon(\mathbf{x})$ is a zero-mean stochastic process.

Starting with this basic model, the common approach in geostatistics, as in other branches of applied statistics, is to view the drift coefficients as constant but unknown values and the base functions as preassigned. (In practice, the base functions are selected as part of the model). The stochastic process $\epsilon(\mathbf{x})$ is characterized through the two-point covariance function

$$E[\epsilon(\mathbf{x})\epsilon(\mathbf{x}')] = R(\mathbf{x}, \mathbf{x}' | \theta) \quad (2)$$

where $E[\]$ stands for expected value and R is a known expression for the covariance function with parameters θ . In practice, the covariance function is usually assumed to depend only on the separation distance $|\mathbf{x} - \mathbf{x}'|$. It is emphasized, however, that stationarity of one form or another is not presumed in our discussion of the generalized covariance function idea.

We will facilitate the analysis by using a compact vector notation that is customary in statistics (Kitanidis, 1987). Assuming that we have a batch of measurements of variable $z(\mathbf{x})$ at locations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, let:

$$\mathbf{z} = n \times 1 \text{ vector of the data, i.e., } \mathbf{z} = \begin{bmatrix} z(\mathbf{x}_1) \\ z(\mathbf{x}_2) \\ \dots \\ z(\mathbf{x}_n) \end{bmatrix}$$

$\mathbf{m} = n \times 1$ mean of \mathbf{z} , i.e., the i th element of \mathbf{m} is the expected value of the process at location \mathbf{x}_i , and $\mathbf{Q} = n \times n$ covariance matrix of \mathbf{z} , i.e., the ij th element of \mathbf{Q} is the covariance between $z(\mathbf{x}_i)$ and $z(\mathbf{x}_j)$.

From Eq. (1),

$$\mathbf{m} = E[\mathbf{z}] = \mathbf{X}\boldsymbol{\beta} \quad (3)$$

where \mathbf{X} is an $n \times p$ matrix of known coefficients (which are usually tied to the locations of the measurements) and $\boldsymbol{\beta}$ is the $p \times 1$ vector of the unknown coefficients of the mean, also known as drift coefficients. The covariance matrix is a function of some parameters θ ,

$$\mathbf{Q} = \mathbf{Q}(\theta) \quad (4)$$

For illustration, consider the stationary isotropic case (i.e., the expected value is constant and the two-point covariance depends only on the separation distance between these two points) with exponential covariance function. Then, $p = 1$, $\boldsymbol{\beta} =$ the constant mean of the intrinsic function, and

$$\mathbf{X} = \left[\begin{array}{c} 1 \\ 1 \\ \dots \\ 1 \end{array} \right] \Bigg\} n$$

The ij th element of \mathbf{Q} is $\theta_1 \exp(-|x_i - x_j|/\theta_2)$, where θ_1 is the variance parameter and θ_2 is the length parameter. Another example is given in Appendix B.

GCF AS SIMPLIFIED COVARIANCE

It is well known (Matheron, 1971, p. 194) that in ordinary kriging with constant mean one only needs to know the covariance function R within a constant. For example, kriging will generate the same result whether one uses:

$$R(h) = \sigma^2 \exp\left(-\frac{h}{l}\right) \quad (5)$$

where h is the separation distance, or

$$R(h) = \sigma^2 \exp\left(-\frac{h}{l}\right) + a \quad (6)$$

where a is an arbitrary constant. Formally, the reason the constant does not matter is the unbiasedness constraint; that is, the requirement that the expected value of estimation error should be zero for any numerical value of the unknown mean. Intuitively, adding constant a to the covariance function is the same as adding to the process a zero-mean random scalar with variance a ; since by design the estimation error in kriging is unaffected by the addition to z of a constant, a has no effect. Thus, a whole family of functions are indistinguishable

from the true covariance function for purposes of ordinary kriging (for example, Eq. (5) and (6) are the same for purposes of kriging). It is useful to picture that each of the many possible functions consists of a common part (or kernel) and a constant, C , which differentiates one function from another but which is inconsequential:

$$R(h) = K(h) + C \quad (7)$$

We call $K(h)$ the generalized covariance function. In practice, one can use for $K(h)$ any of the many functions which differ by a constant. That is, it does not really matter whether one calls the expression of Eq. (5) or (6) a generalized covariance function.

The important point is that for purposes of ordinary kriging one does not need to worry about finding the whole covariance function but only its essential part, the GCF. From an estimation viewpoint, this is good news since one has to infer less.

In ordinary kriging, we limit our attention to intrinsic functions and work with the variogram. However, as shown in Appendix A, the variogram is minus a GCF,

$$\gamma(h) = -K(h) \quad (8)$$

Universal kriging is a generalization of ordinary kriging. The regionalized variable is now represented as the sum of a deterministic function and a stochastic process with zero mean. The deterministic part is the summation of preassigned functions with unknown coefficients, as given in Eq. (1). The number of unbiasedness constraints is equal to the number of unknown coefficients. The more the unbiasedness constraints, the more the variability among the covariance functions that behave exactly the same. For example, for linear drift, the covariance function

$$R(h) = \sigma^2 \exp\left(-\frac{h}{l}\right) + a_1 + a_2 h^2 \quad (9)$$

where a_1 and a_2 are arbitrary constants, is the same for purposes of kriging with

$$R(h) = \sigma^2 \exp\left(-\frac{h}{l}\right) \quad (10)$$

This result is presented, for the sake of completeness, in Appendix B.

Thus, by adding more unbiasedness constraints, we end up with more freedom in choosing a covariance function compared to the constant-mean case. We now need to know only some of the terms that form the covariance function. We can write:

$$R(h) = K(h) + C(h) \quad (11)$$

where $K(h)$ is the essential part, or GCF, and $C(h)$ is the redundant part. Equation (11), which expresses that the GCF is the essential part of the covariance function, is useful in understanding kriging with unknown drift coefficients and related covariance estimation methods.

Generally, by introducing more terms in the deterministic part of the model, we move more terms from the essential part to the redundant part. The reason is plain: By describing more of the variability through a deterministic function, we reduce the importance of the covariance function (which describes the stochastic part). At the limit, when all structured variability is described through a deterministic function, the GCF is a simple pure nugget effect.

I emphasize that I am not advocating that one should use a variable-mean model when there are no good reasons for doing so. Use of the modeling assumption to describe some of the variability through a deterministic function (variable mean) leads to sharper predictions (i.e., with smaller mean square error) than if all variability is described through the stochastic term. However, being a stronger assumption, it requires a more convincing justification and should be used only when appropriate. The point I am making is that one should not avoid variable-mean models only because of difficulties in estimating the covariance function. If anything, the opposite is true. Our work in model development and fitting is generally made easier when we realize that we seek the generalized rather than the ordinary covariance function.

DATA DETRENDING

We turn our attention now to the problem of determining the covariance. In ordinary kriging (with constant mean) most practitioners graphically fit an equation to the experimental variogram. The same approach has been extended to universal kriging by using the detrended data, i.e., the original data from which the fitted drift has been subtracted. This approach has been criticized (Armstrong, 1984) because the variogram of the detrended data is different from the variogram of the original stochastic process and dependent on the method of detrending. It is well known (Starks and Fang, 1982) that the presence of a drift distorts the experimental variogram.

We will confirm here that indeed the detrended data has a different covariance function (or variogram) from the original process. However, we will show rigorously that the original and the detrended data have the same GCF.

In matrix notation, we will consider the general least squares criterion of agreement between the data and the trend:

$$(\mathbf{z} - \mathbf{X}\hat{\beta})^T \mathbf{A}^{-1} (\mathbf{z} - \mathbf{X}\hat{\beta}) \quad (12)$$

where \mathbf{A} is a symmetric matrix (satisfying appropriate positive definiteness re-

quirements so that the criterion has a unique minimum). The values of the drift coefficients that minimize this fitting criterion are:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1} \mathbf{z} \quad (13)$$

Then, we form the detrended data, i.e., the original data minus the drift using the fitted coefficients

$$\mathbf{z}_d = \mathbf{z} - \mathbf{X}\hat{\beta} = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1}) \mathbf{z} \quad (14)$$

We will compute the mean and covariance of the detrended data or residuals. Taking expectations:

$$\begin{aligned} E[\mathbf{z}_d] &= (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1}) E[\mathbf{z}] \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1}) \mathbf{X}\beta = \mathbf{0} \end{aligned} \quad (15)$$

after using Eq. (3). Thus, the expected value of the residuals is zero.

The covariance of the residuals is:

$$\begin{aligned} \mathbf{Q}_d &= E[\mathbf{z}_d(\mathbf{z}_d)^T] = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1}) \\ &E[(\mathbf{z} - \mathbf{X}\beta)(\mathbf{z} - \mathbf{X}\beta)^T] (\mathbf{I} - \mathbf{A}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T) \\ &= (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1}) \mathbf{Q} (\mathbf{I} - \mathbf{A}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T) \\ &= \mathbf{Q} - \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1} \mathbf{Q} - \mathbf{Q} \mathbf{A}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \\ &\quad + \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^{-1} \mathbf{Q} \mathbf{A}^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{A}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \end{aligned} \quad (16)$$

Thus, the covariance matrix of the detrended residuals is not \mathbf{Q} , the covariance matrix of the original data (for example, the rank of matrix \mathbf{Q} is generally n whereas the rank of \mathbf{Q}_d is not higher than $n - p$). This holds true even if the covariance were somehow known so that we could choose $\mathbf{A} = \mathbf{Q}$ in which case the expression would simplify:

$$E[\mathbf{z}_d(\mathbf{z}_d)^T] = \mathbf{Q} - \mathbf{X}(\mathbf{X}^T \mathbf{Q}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \neq \mathbf{Q} \quad (17)$$

Thus, as previously noticed by many other investigators, the covariance matrix of the detrended data is not the same with the covariance matrix of the original data.

However, the difference is in the nonessential part. To remove the nonessential part, form authorized (or generalized) increments by multiplying the detrended data by any matrix \mathbf{G} such that

$$\mathbf{G}\mathbf{X} = \mathbf{0} \quad (18)$$

The covariance matrix of the $\mathbf{G}\mathbf{z}_d$ is $\mathbf{G}\mathbf{Q}_d\mathbf{G}^T$; using Eq. (16),

$$\mathbf{G}\mathbf{Q}_d\mathbf{G}^T = \mathbf{G}\mathbf{Q}\mathbf{G}^T \quad (19)$$

Thus, the original and the detrended data have the same generalized covariance function. The practical significance of this result is that: The experimental covariance (or variogram) of the detrended data can be used to estimate the generalized (not the ordinary) covariance function.

Since the original and the detrended data share the same generalized covariance, one may ask what are the advantages of using the detrended data instead of the original data in experimental variogram analysis. The answer is that, when the original data is used, the trend swamps the experimental variogram making the job of inferring the generalized covariance function even more difficult. Detrending is helpful because it removes much of the redundant (drift related) variability in the experimental variogram revealing the generalized covariance function.

The simplest and most computationally efficient approach is to detrend using ordinary least squares. That is, take \mathbf{A} equal to the identity matrix, \mathbf{I} , in which case the detrended data can be computed from:

$$\mathbf{z}_d = \mathbf{z} - \mathbf{X}\hat{\beta} = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{z} \quad (20)$$

In other words, the familiar method of fitting an equation to the experimental variogram is applicable in the variable-mean case as it is in the constant-mean case. However, as well known, this procedure is subjective and there are good chances that different analysts will fit different models to the same data. Nevertheless, one can use this approach to obtain a preliminary estimate of the generalized covariance function without concern about the distortion which is limited to the redundant part of the covariance function.

Another, and in my opinion more objective and promising, approach for estimating generalized covariance functions is to use *Restricted maximum Likelihood* and related techniques (see review in Kitanidis, 1987). The idea is to adjust the parameters so that the square difference between observations and model predictions is as small as possible (Kitanidis, 1991).

AN APPLICATION

We will analyze hydraulic head data from the Jordan aquifer, in Iowa (Hocksema and Kitanidis, 1984). The location of the measurements is shown on Fig. 1. As is often the case with hydraulic head in deep aquifers at a regional scale, the data indicate an approximate linear drift. The nodes of the experimental variogram of the data are shown on Fig. 2. Based on statistical arguments as well as hydrogeologic information additional to the data set, it was deemed appropriate to try a variable mean model with:

$$m(x_1, x_2) = \beta_1 + \beta_2 x_1 + \beta_3 x_2 \quad (21)$$

where x_1, x_2 are spatial coordinates and $\beta_1, \beta_2, \beta_3$ are constant but unknown coefficients.

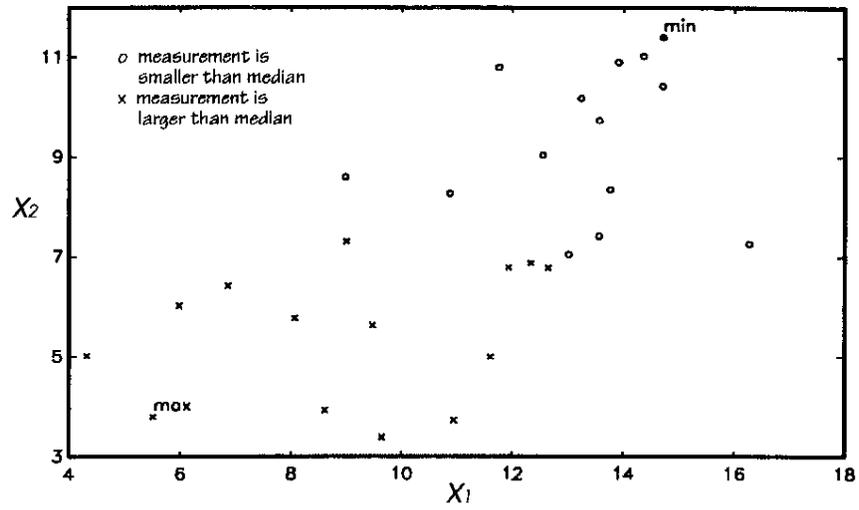


Fig. 1. Location of hydraulic head data.

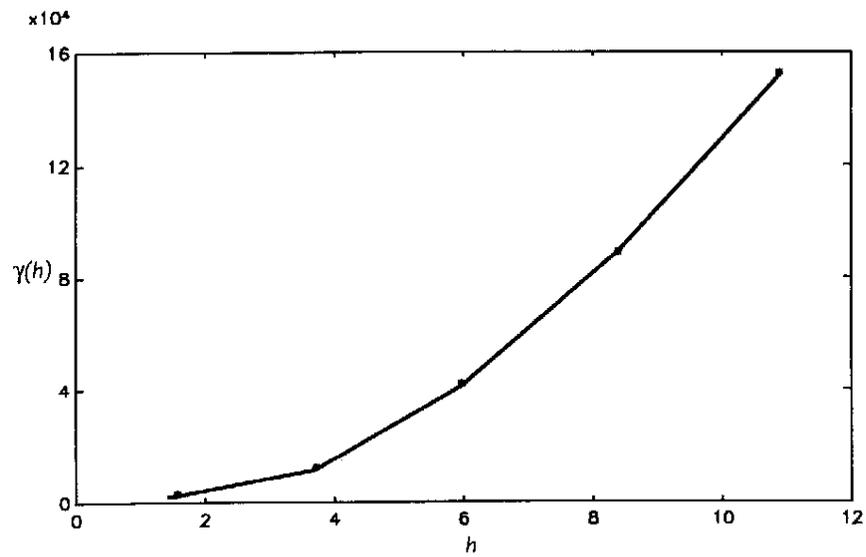


Fig. 2. Experimental (semi)variogram of original data.

The experimental variogram of the original data, Fig. 2, is not very helpful because it suggests a quadratic power model. However, an h^2 term is redundant in universal kriging with a linear drift, Eq. (21). The problem is that the drift has swamped the variability of the stochastic part in the experimental variogram. For this reason, the data were detrended and the variogram was plotted in Fig. 3. It appears that an exponential variogram might be appropriate. This corresponds to the following generalized covariance function:

$$K(h) = \theta_1 \exp(-h/\theta_2) \quad (22)$$

where h is separation distance and θ_1, θ_2 are positive parameters. The parameters can be fitted graphically from the experimental variogram but the procedure is somewhat subjective.

Another approach is to select θ_2 which minimizes the prediction error and θ_1 from the mean square error of estimation in the method of orthonormal residuals (see Kitanidis, 1991, p. 752). This method leads to best estimates $\hat{\theta}_1 = 4228, \hat{\theta}_2 = 6$. Values of θ_2 in the range 5–7 provide essentially equally good fits. On Fig. 4, the continuous line represents the fitted model and is shown to be in agreement with the experimental variogram. Note that the apparent discrepancy between the model and the experimental variogram at large lags is of no concern for two reasons: First, the sampling error associated with the experimental variogram at large lags is so large that it would be inappropriate to modify the model to obtain better reproduction of the experimental variogram at large distances. Second, variograms that differ by a quadratic function are

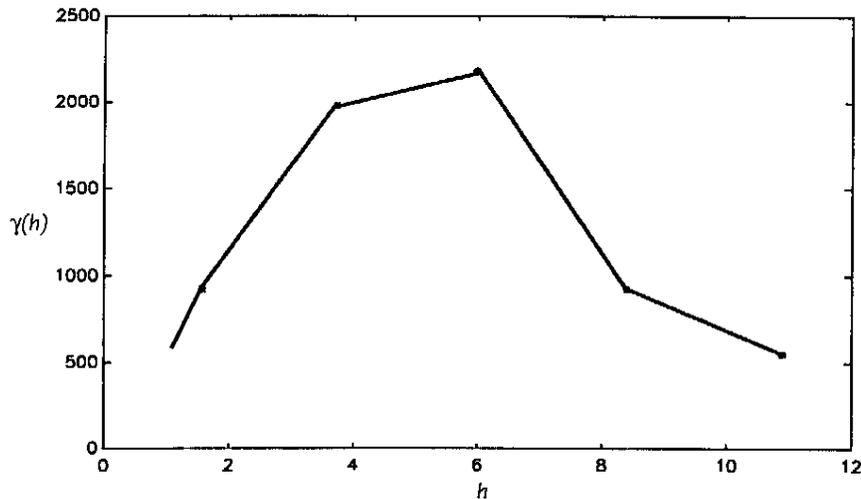


Fig. 3. Experimental variogram of detrended data.

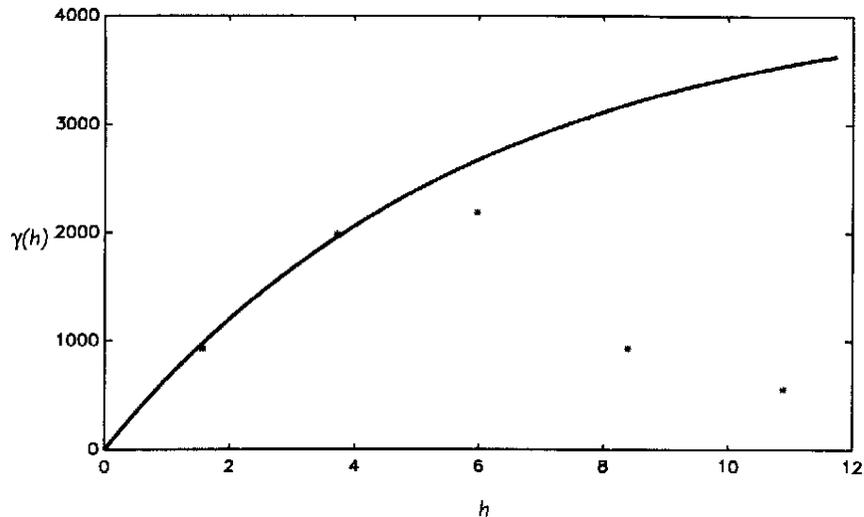


Fig. 4. Experimental variogram and fitted equation (exponential GCF).

practically the same for purposes of estimation when a linear drift is included in the model (see Eq. (9)).

DISCUSSION

The concept of generalized covariances is not limited to a particular model, such as intrinsic functions of order k , or form, such as polynomial generalized covariance function. It is an idea underlying any best linear unbiased estimator based on the linear model with unknown coefficients (see Eq. (1)). This involves all common cases of linear geostatistics including ordinary and universal kriging and co-kriging.

In linear geostatistics, spatial variability is described through the drift (mean function) and the covariance function. The drift is the sum of known functions multiplied by constant but unknown coefficients. The best linear unbiased estimator is a linear function of the observations with weights selected so that: (a) the mean estimation error is zero for any values of the drift coefficients, and (b) the mean square error is as small as possible. Restrictions imposed to eliminate the unknown coefficients due to the first requirement make part of the covariance function redundant. That is, the covariance can be written as the sum of an essential part, called generalized covariance function, and a redundant part which can be neglected for purposes of best linear unbiased estimation.

Thus, the generalized covariance function is a simplified version of the customary covariance function and is obtained by neglecting redundant terms.

The larger the part of the spatial variability that is described through the deterministic term, the simpler the generalized covariance that is needed and the less sensitive the predictions on the covariance that is used. In this sense, the job of selecting a covariance function for predictive purposes is easier when the mean is variable. Therefore, one should not shy away from using variable-mean models only out of concern for difficulties associated with generalized covariance functions.

A preliminary estimate of the generalized covariance can be obtained through the familiar method of fitting an equation to the experimental variogram of the detrended data. It has been shown here that the original data and the properly detrended data share the same generalized covariance, even though their ordinary covariances differ. Thus, one can detrend the data and then glean the generalized covariance function from the experimental variogram of the detrended data. This estimate may be sufficient for some applications or may be improved by using other parameter estimation techniques that seek to optimize the fit of the model to the data.

Consequently, I see no reason to perpetuate the distinction between “universal kriging” and “kriging with generalized covariances.” It is known (for example, see Chistensen, 1990) although perhaps not widely understood that as best linear unbiased estimators, the two methods are the same. Some practitioners differentiate between the two methods on the basis of how parameters are estimated. Universal kriging is associated with estimation using the exponential variogram of the detrended data whereas kriging with generalized covariances is associated with other estimation algorithms, such as that of Delfiner (1976) and Kitanidis (1983). However, we have seen here that even this distinction is not important because the parameter estimation approach associated with “universal kriging” is appropriate (in the sense that it is appropriate in the constant mean case) only if interpreted as estimation of generalized covariance!

APPENDIX A

As well known, a process $z(\mathbf{x})$ is intrinsic if

$$E[z(\mathbf{x}) - z(\mathbf{x}')] = 0 \quad (23)$$

$$\frac{1}{2} E[(z(\mathbf{x}) - z(\mathbf{x}'))^2] = \gamma(\mathbf{x} - \mathbf{x}') \quad (24)$$

for any two points \mathbf{x} and \mathbf{x}' , where γ is known as the semivariogram (or just plain variogram).

A process is stationary if it has constant mean and the covariance function depends only on the separation vector:

$$E[z(\mathbf{x})] = E[z(\mathbf{x}')] = m \quad (25)$$

$$E[(z(\mathbf{x}) - m)(z(\mathbf{x}') - m)] = R(\mathbf{x} - \mathbf{x}') \quad (26)$$

for any two points with location indexes \mathbf{x} and \mathbf{x}' , where m is the mean of the process (the expectation at any point) and R is the covariance function.

Let us briefly discuss the relation between intrinsic and stationary functions while at the same time showing the relation between the covariance, the generalized covariance, and the variogram. It is instructive to start with the class of stationary functions and show how it can be extended to the class of intrinsic functions. If $z(\mathbf{x})$ is a stationary function, we can rewrite Eq. (25) and (26) to eliminate m :

$$E[z(\mathbf{x}) - z(\mathbf{x}')] = 0 \quad (27)$$

$$\begin{aligned} E[(z(\mathbf{x}) - m)(z(\mathbf{x}') - m)] &= E[z(\mathbf{x})z(\mathbf{x}') - z(\mathbf{x})m - z(\mathbf{x}')m + m^2] \\ &= E[-\frac{1}{2}(z(\mathbf{x}) - z(\mathbf{x}'))^2 \\ &\quad + \frac{1}{2}z(\mathbf{x})^2 + \frac{1}{2}z(\mathbf{x}')^2 - z(\mathbf{x})m - z(\mathbf{x}')m + m^2] \\ &= -\frac{1}{2}E[(z(\mathbf{x}) - z(\mathbf{x}'))^2] + R(\mathbf{0}) \end{aligned} \quad (28)$$

However, from Eqs. (24), (26), and (28), we obtain:

$$R(\mathbf{x} - \mathbf{x}') = -\gamma(\mathbf{x} - \mathbf{x}') + R(\mathbf{0}) \quad (29)$$

From this equation, because $R(\mathbf{0})$ (variance or sill) is a constant, one can see that $-\gamma(\mathbf{x} - \mathbf{x}')$ serves as a generalized covariance function and can be used in kriging instead of the covariance function even if the process is stationary.

Furthermore, since the value of the constant does not matter, we can extend the class of stationary functions to include cases where this variance is arbitrarily large. Thus, we obtain the class of the intrinsic functions that includes stationary functions with finite sill as well as stationary-increment functions (such as Brownian motions) with unbounded variance.

APPENDIX B

We will consider a spatial function z defined on two dimensions (x, y) that is modeled as a sum of a linear trend and a zero-mean stochastic process:

$$z(x, y) = \beta_1 + \beta_2x + \beta_3y + \epsilon(x, y) \quad (30)$$

where x and y are spatial coordinates, β_1 , β_2 , and β_3 are unknown drift coefficients, and ϵ is a known zero-mean stochastic process with covariance function:

$$R(h) = \sigma^2 \exp\left(-\frac{h}{l}\right) \quad (31)$$

where $h = \sqrt{(x - x')^2 + (y - y')^2}$ is distance and σ^2 , l are parameters.

Note that if n observations are available, then

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \dots & \dots & \dots \\ 1 & x_n & y_n \end{bmatrix} \quad (32)$$

Given these observations (which we will denote by z_1, z_2, \dots, z_n) we want to estimate z_o , the value of z at location (x_o, y_o) . Limiting our attention to linear estimators,

$$\hat{z}_o = \sum_{i=1}^n \lambda_i z_i \quad (33)$$

our task is to select $\lambda_1, \lambda_2, \dots, \lambda_n$. We will enforce the unbiasedness condition, i.e., that the expected value of the estimation error should be zero,

$$E \left[\sum_{i=1}^n \lambda_i z_i - z_o \right] = 0 \quad (34)$$

Using Eq. (30),

$$E \left[\sum_{i=1}^n \lambda_i (\beta_1 + \beta_2 x_i + \beta_3 y_i + \epsilon(x_i, y_i)) - (\beta_1 + \beta_2 x_o + \beta_3 y_o + \epsilon(x_o, y_o)) \right] = 0 \quad (35)$$

or, using that ϵ is a zero-mean process and rearranging terms,

$$\left(\sum_{i=1}^n \lambda_i - 1 \right) \beta_1 + \left(\sum_{i=1}^n \lambda_i x_i - x_o \right) \beta_2 + \left(\sum_{i=1}^n \lambda_i y_i - y_o \right) \beta_3 = 0 \quad (36)$$

The only way to guarantee that this expression will be zero for any value of β_1 , β_2 , and β_3 is to enforce the unbiasedness constraints:

$$\sum_{i=1}^n \lambda_i = 1 \quad (37)$$

$$\sum_{i=1}^n \lambda_i x_i = x_o \quad (38)$$

$$\sum_{i=1}^n \lambda_i y_i = y_o \quad (39)$$

Next, we compute the mean square error:

$$MSE = E \left[\left(\sum_{i=1}^n \lambda_i z_i - z_o \right)^2 \right] \quad (40)$$

Making use of the unbiasedness constraints,

$$\begin{aligned} MSE &= E \left[\left(\sum_{i=1}^n \lambda_i (z_i - E[z_i]) - (z_o - E[z_o]) \right)^2 \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j R(h_{ij}) - 2 \sum_{i=1}^n \lambda_i R(h_{io}) + R(0) \end{aligned} \quad (41)$$

where h_{ij} is the distance between the locations of measurements i and j and h_{io} is the distance between the locations of measurement i and the unknown. Using the method of the Lagrange multipliers, the solution that minimizes the mean square error subject to unbiasedness constraints given by solving the system of $n + 3$ equations with $n + 3$ unknowns ($\lambda_1, \lambda_2, \dots, \lambda_n, \nu_1, \nu_2, \nu_3$):

$$\sum_{j=1}^n \lambda_j R(h_{ij}) + \nu_1 + \nu_2 x_i + \nu_3 y_i = R(h_{io}) \quad \text{for } i = 1, 2, \dots, n \quad (42)$$

combined with Eqs. (37)–(39).

We can now verify that, due to the unbiasedness constraint, the MSE is the same whether one uses $R(h)$ or $R(h) + a_1 + a_2 h^2$, where a_1 and a_2 are arbitrary coefficients.

$$\begin{aligned} MSE &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j [R(h_{ij}) + a_1 + a_2 h_{ij}^2] \\ &\quad - 2 \sum_{i=1}^n \lambda_i [R(h_{io}) + a_1 + a_2 h_{io}^2] + R(0) \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j R(h_{ij}) - 2 \sum_{i=1}^n \lambda_i R(h_{io}) + R(0) \end{aligned} \quad (43)$$

The proof follows:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j a_1 - 2 \sum_{i=1}^n \lambda_i a_1 = \left[\left(\sum_{i=1}^n \lambda_i \right) \left(\sum_{j=1}^n \lambda_j \right) - 2 \left(\sum_{i=1}^n \lambda_i \right) \right] a_1 = 0$$

and

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j a_2 h_{ij}^2 - 2 \sum_{i=1}^n \lambda_i a_2 h_{io}^2 \\
&= \left[\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j ((x_i - x_j)^2 + (y_i - y_j)^2) \right. \\
&\quad \left. - 2 \sum_{i=1}^n \lambda_i ((x_i - x_o)^2 + (y_i - y_o)^2) \right] a_2 \\
&= \left[\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (x_i^2 + x_j^2 - 2x_i x_j + y_i^2 + y_j^2 - 2y_i y_j) \right. \\
&\quad \left. - 2 \sum_{i=1}^n \lambda_i (x_i^2 + x_o^2 - 2x_i x_o + y_i^2 + y_o^2 - 2y_i y_o) \right] a_2 = 0 \quad (44)
\end{aligned}$$

We will show that the bracketed expression is zero. This expression can be simplified:

$$\begin{aligned}
& \sum_{i=1}^n \lambda_i (x_i^2 + y_i^2) \left(\sum_{j=1}^n \lambda_j \right) + \left(\sum_{i=1}^n \lambda_i \right) \sum_{j=1}^n \lambda_j (x_j^2 + y_j^2) \\
&\quad - 2 \left(\sum_{i=1}^n \lambda_i x_i \right) \left(\sum_{j=1}^n \lambda_j x_j \right) - 2 \left(\sum_{i=1}^n \lambda_i y_i \right) \left(\sum_{j=1}^n \lambda_j y_j \right) \\
&\quad - 2 \sum_{i=1}^n \lambda_i (x_i^2 + y_i^2) + 4 \left(\sum_{i=1}^n \lambda_i x_i \right) x_o \\
&\quad + 4 \left(\sum_{i=1}^n \lambda_i y_i \right) y_o - 2 \left(\sum_{i=1}^n \lambda_i \right) (x_o^2 + y_o^2) \\
&= 2 \sum_{i=1}^n \lambda_i (x_i^2 + y_i^2) - 2x_o^2 - 2y_o^2 - 2 \sum_{i=1}^n \lambda_i (x_i^2 + y_i^2) + 4x_o^2 + 4y_o^2 \\
&\quad - 2(x_o^2 + y_o^2) = 0 \quad (45)
\end{aligned}$$

Therefore, adding to the covariance function the term $a_1 + a_2 h^2$ has no effect on the mean square error or the associated kriging equations.

ACKNOWLEDGMENTS

Funding for this study was provided by the Office of Research and Development, U.S. Environmental Protection Agency, under agreement R-815738-01 through the WRHSRC. Funding for this work has also been provided by the

National Science Foundation under Grant BCS-8914812. The content of this article does not necessarily represent the views of these agencies.

REFERENCES

- Armstrong, M., 1984, Problems with Universal Kriging: *Math Geo.*, v. 16, p. 101-108.
- Christensen, R., 1990, The Equivalence of Predictions from Universal Kriging and Intrinsic Random Function Kriging: *Math. Geo.*, v. 22, p. 655-664.
- Delfiner, P., 1976, Linear Estimation of Nonstationary Spatial Phenomena, in M. Guarascio, M. David, and C. Huijbregts, Eds., *Advanced Geostatistics in the Mining Industry*: Reidel, Dordrecht.
- Hoeksema, R. J., and Kitanidis P. K., 1984, An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling: *Water Res. Res.*, v. 20, p. 1003-1020.
- Journel, A., 1986, Geostatistics: Models and Tools for the Earth Sciences: *Math. Geo.*, v. 18, p. 119-140.
- Kitanidis, P. K., 1983, Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrologic Applications: *Water Res. Res.*, v. 19, p. 909-921.
- Kitanidis, P. K., 1987, Parametric Estimation of Covariances of Regionalized Variables: *Wat. Res. Bull.*, v. 23, p. 557-567.
- Kitanidis, P. K., 1991, Orthonormal Residuals in Geostatistics; Model Criticism and Parameter Estimation: *Math. Geo.*, v. 23, p. 741-758.
- Matheron, G., 1971, *The Theory of Regionalized Variables and Its Applications*: Ecole des Mines, Fontainebleau, France.
- Matheron, G., 1973, The Intrinsic Random Function: *Adv. Appl. Prob.*, v. 5, p. 438-468.
- Starks, T. H., and Fang, J. H., 1982, The Effect of Drift on the Experimental Semivariogram: *Math. Geo.*, v. 14, p. 309-319.