12

ORDINARY KRIGING

In the previous chapter we compared several point estimation methods and saw that different methods were "best" according to different estimation criteria. In this chapter we will look at ordinary kriging, a method that is often associated with the acronym B.L.U.E. for "best linear unbiased estimator." Ordinary kriging is "linear" because its estimates are weighted linear combinations of the available data; it is "unbiased" since it tries to have m_R, the mean residual or error, equal to 0; it is "best" because it aims at minimizing σ_R^2, the variance of the errors. All of the other estimation methods we have seen so far are also linear and, as we have already seen, are also theoretically unbiased. The distinguishing feature of ordinary kriging, therefore, is its aim of minimizing the error variance.

The goals of ordinary kriging are ambitious ones and, in a practical sense, unattainable since m_R and σ_R^2 are always unknown. In the previous chapter, our calculations of the mean error and the error variance were possible only because we had access to the exhaustive data set. In practical situations we never know the true answers or the actual errors before we attempt our estimation. The importance of this for ordinary kriging is that we never know m_R and therefore cannot guarantee that it is exactly 0. Nor do we know σ_R^2; therefore, we cannot minimize it. The best we can do is to build a model of the data we are studying and work with the average error and the error variance for the model. In ordinary kriging, we use a probability model in which the bias and the error variance can both be calculated

and then choose weights for the nearby samples that ensure that the average error for our model, $\tilde{m}_R$, is exactly 0 and that our modeled error variance, $\tilde{\sigma}_R^2$, is minimized. We will use the same convention as in Chapter 9 where the symbol $\tilde{}$ is used to denote a parameter of a model, and distinguish it from a statistic of the data.

We will be using a random function model since this type of model enables us to express the error, its mean value, and its variance. In this chapter we will begin by reviewing the approach we took earlier in Chapter 9 when we first encountered random functions and tackled the problem of unbiased estimates. After we have developed an expression for the error we will apply an earlier result, also from Chapter 9, that allowed us to express the variance of a weighted linear combination of random variables. We will then be able to develop the ordinary kriging system by using introductory calculus to minimize the error variance. Following a detailed example that illustrates how the ordinary kriging weights are calculated, we will look at how the choice of a model of spatial continuity affects the ordinary kriging weights. Finally, we will extend the point estimation case study of the previous chapter to include ordinary kriging.

The Random Function Model and Unbiasedness

In Chapter 9, we introduced the concept of a random function model and showed how it could help us in deciding how to weight the nearby samples so that our estimates are unbiased. At every point where we do not have a sample, we will estimate the unknown true value using a weighted linear combination of the available samples:

$$\hat{v} = \sum_{j=1}^{n} w_j \cdot v$$

The set of weights is allowed to change as we estimate unknown values at different locations.

If we define the error, r, of any particular estimated value to be the difference between the estimated value and the true value at that same location:

$$\textit{Error of i-th estimate} = r_i = \hat{v}_i - v_i \qquad (12.1)$$

then the average error of a set of k estimates is

$$\textit{Average error} = m_r = \frac{1}{k}\sum_{i=1}^{k} r_i = \frac{1}{k}\sum_{i=1}^{k} \hat{v}_i - v_i \qquad (12.2)$$

Unfortunately, we are unable to make much use of this equation since it involves quantities that we do not know, namely the true values $v_1, \ldots, v_k$.

The probabilistic solution to this problem consists of conceptualizing the unknown values as the outcome of a random process and solving the problem for our conceptual model. For any point at which we attempt to estimate the unknown value, our model is a stationary random function that consists of several random variables, one for the value at each of the seven sample locations, $V(x_1), \ldots, V(x_n)$, and one for the unknown value at the point we are trying to estimate, $V(x_0)$. Each of these random variables has the same probability law; at all locations, the expected value of the random variable is $E\{V\}$. Any pair of random variables has a joint distribution that depends only on the separation between the two points and not on their locations. The covariance between pairs of random variables separated by a particular distance, h, is $\tilde{C}_V(h)$.

Every value in this model is seen as the outcome of a random variable; the samples are outcomes of random variables, as is the unknown true value. Our estimate is also a random variable since it is a weighted linear combination on the random variables at the available sample locations:

$$\hat{V}(x_0) = \sum_{i=1}^{n} w_i \cdot V(x_i)$$

Similarly, the estimation error, defined as the difference between the estimate and the random variable modeling the true value, is also a random variable:

$$R(x_0) = \hat{V}(x_0) - V(x_0)$$

By substituting the previous equation which expressed $\hat{V}(x_0)$ in terms of other random variables, we can express $R(x_0)$ solely in terms of the original $n+1$ random variables in our random function model:

$$R(x_0) = \sum_{i=1}^{n} w_i \cdot V(x_i) \quad - \quad V(x_0) \qquad (12.3)$$

The error that we make when we estimate the unknown value at x_0 is an outcome of the random variable $R(x_0)$.

We can ensure that the error at any particular location has an expected value of 0 by applying the formula for the expected value of a linear combination to Equation 12.3:

$$\begin{aligned} E\{R(x_0)\} &= E\{\sum_{i=1}^{n} w_i \cdot V(x_i) \quad - \quad V(x_0)\} \\ &= \sum_{i=1}^{n} w_i E\{V(x_i)\} \quad - \quad E\{V(x_0)\} \end{aligned}$$

We have already assumed that the random function is stationary, which allows us to express all of the expected values on the right-hand side as $E\{V\}$:

$$E\{R(x_0)\} = \sum_{i=1}^{n} w_i E\{V\} \quad - \quad E\{V\}$$

The expected value of the error at any particular location, $E\{R(x_0)\}$ is often referred to as the bias. Setting this expected value to 0 to ensure unbiasedness results in the following conclusion:

$$\begin{aligned} E\{R(x_0)\} = 0 &= E\{V\}\sum_{i=1}^{n} w_i \quad - \quad E\{V\} \\ E\{V\}\sum_{i=1}^{n} w_i &= E\{V\} \\ \sum_{i=1}^{n} w_i &= 1 \end{aligned}$$

As we noted when we first arrived at this conclusion, all of the common estimation procedures we used in our case studies in the previous chapter all make use of this unbiasedness condition.

The Random Function Model and Error Variance

As an estimation methodology, ordinary kriging distinguishes itself by its attempt to produce a set of estimates for which the variance of the errors is minimum. The error variance, σ_R^2, of a set of k estimates can be written as

$$\sigma_R^2 = \frac{1}{k}\sum_{i=1}^{k}(r_i - m_R)^2$$

$$= \frac{1}{k}\sum_{i=1}^{k}[\hat{v}_i - v_i - \frac{1}{k}\sum_{i=1}^{k}(\hat{v}_i - v_i)]^2$$

$v_1, \ldots, v_n$ are the true values and $\hat{v}_1, \ldots, \hat{v}_n$ are the corresponding estimates. If we are willing to assume that we have a mean error of 0, we can simplify this equation somewhat:

$$\begin{aligned}\sigma_R^2 &= \frac{1}{k}\sum_{i=1}^{k}(r_i - 0)^2 \\ &= \frac{1}{k}\sum_{i=1}^{k}[\hat{v}_i - v_i]^2\end{aligned}$$

As with Equation 12.2, which provided an expression for the mean error, we cannot get very far with this equation for the error variance because it calls for knowledge of the true values.

To get out of this unfortunate dead end, we will again turn to random function models. As in the previous section, we begin with $n+1$ random variables, n of which model the behavior of the phenomenon at the nearby sample locations and one of which models its behavior at the location whose value we are trying to estimate. The available samples will be combined in a weighted linear combination to form our estimate:

$$\hat{V}(x_0) = \sum_{i=1}^{n} w_i V(x_i) \tag{12.4}$$

The difference between the true value and the corresponding estimate will be our error or residual:

$$R(x_0) = \hat{V}(x_0) - V(x_0) \tag{12.5}$$

As we did with the unbiasedness problem, we will transfer the original problem into the corresponding model problem. Though we cannot minimize the variance of our actual errors, we can minimize the variance of our modeled error $R(x_0)$. This minimization will be accomplished by finding an expression for the modeled error variance, $\tilde{\sigma}_R^2$, and setting to 0 the various partial derivatives of this expression.

Our first task, then, is to find an expression for the variance of the error. This error is a random variable, since it is a weighted linear combination of other random variables. In Chapter 9, when we introduced random function models, we gave a formula for the variance of

a weighted linear combination:

$$Var\{\sum_{i=1}^{n} w_i \cdot V_i\} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i \cdot w_j \cdot Cov\{V_i V_j\} \tag{12.6}$$

Using this formula with Equation 12.5, we can express the variance of the error as:

$$\begin{aligned}Var\{R(x_0)\} &= Cov\{\hat{V}(x_0)\hat{V}(x_0)\} - Cov\{\hat{V}(x_0)V(x_0)\} \\ &\quad - Cov\{V(x_0)\hat{V}(x_0)\} + Cov\{V(x_0)V(x_0)\} \\ &= Cov\{\hat{V}(x_0)\hat{V}(x_0)\} - 2Cov\{\hat{V}(x_0)V(x_0)\} \\ &\quad + Cov\{V(x_0)V(x_0)\}\end{aligned} \tag{12.7}$$

The first term $Cov\{\hat{V}(x_0)\hat{V}(x_0)\}$ is the covariance of $\hat{V}(x_0)$ with itself, which is equal to the variance of $\hat{V}(x_0)$, itself a linear combination $\sum_{i=1}^{n} w_i V(x_i)$ of other random variables:

$$Var\{\hat{V}(x_0)\hat{V}(x_0)\} = Var\{\sum_{i=1}^{n} w_i \cdot V_i\} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij}$$

The third term in Equation 12.6, $Cov\{V(x_0)V(x_0)\}$, is the covariance of the random variable $V(x_0)$ with itself and is equal to the variance of $V(x_0)$. If we assume that all of our random variables have the same variance, $\tilde{\sigma}^2$, then this third term can be expressed as

$$Cov\{V(x_0)V(x_0)\} = \tilde{\sigma}^2$$

The second term in Equation 12.6 can be written as

$$\begin{aligned}2Cov\{\hat{V}(x_0)\hat{V}(x_0)\} &= 2Cov\{(\sum_{i=1}^{n} w_i V_i)V_0\} \\ &= 2E\{\sum_{i=1}^{n} w_i V_i \cdot V_0\} - 2E\{\sum_{i=1}^{n} w_i V_i\} \cdot E\{V_0\} \\ &= 2\sum_{i=1}^{n} w_i \cdot E\{V_i, V_0\} - 2\sum_{i=1}^{n} w_i \cdot E\{V_i\} \cdot E\{V_0\} \\ &= 2\sum_{i=1}^{n} w_i \cdot Cov\{V_i V_0\} \\ &= 2\sum_{i=1}^{n} w_i \tilde{C}_{i0}\end{aligned}$$

Combining these three terms again, we now have the following expression for the error variance:

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} - 2\sum_{i=1}^{n} w_i \tilde{C}_{i0} \qquad (12.8)$$

Once we have chosen our random function model parameters, specifically the variance $\tilde{\sigma}^2$ and all the covariances $\tilde{C}_{ij}$, Equation 12.8 gives us an expression for the error variance as a function of n variables, namely the weights $w_1, \ldots, w_n$.

The minimization of a function of n variables is usually accomplished by setting the n partial first derivatives to 0. This produces a system of n equations and n unknowns that can be solved by any one of several methods for solving systems of simultaneous linear equations. Unfortunately, this procedure is not quite correct for the minimization of $\tilde{\sigma}_R^2$ since we have a constraint on our solution. Earlier, we decided to use the unbiasedness condition; this means that we cannot accept any set of n weights as a solution, but must restrict possible solutions to those sets of weights that sum to 1. Such problems of constrained optimization can be solved by the technique of Lagrange parameters described in the next section.

The Lagrange Parameter

The technique of Lagrange parameters is a procedure for converting a constrained minimization problem into an unconstrained one [1]. If we try to tackle the minimization of $\tilde{\sigma}_R^2$, as expressed in Equation 12.8, as an unconstrained problem, we run into difficulties. Setting the n partial first derivatives of $\tilde{\sigma}_R^2$ to 0 will produce n equations and n unknowns. The unbiasedness condition will add another equation without adding any more unknowns. This leaves us with a system of $n+1$ equations and only n unknowns, the solution of which is not straightforward.

To avoid this awkward problem, we introduce another unknown into our equation for $\tilde{\sigma}_R^2$. This new variable is called μ, the Lagrange parameter, and is introduced into Equation 12.8 in the following way:

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij} - 2\sum_{i=1}^{n} w_i \tilde{C}_{i0} + \underbrace{2\mu\left(\sum_{i=1}^{n} w_i - 1\right)}_{0} \qquad (12.9)$$

Adding variables to an equation is a tricky business; we have to be sure we do not upset the equality. The way we have chosen to do it in Equation 12.9 is safe because the term we are adding at the end is 0 due to the unbiasedness condition:

$$\begin{aligned} \sum_{i=1}^{n} w_i &= 1 \\ \sum_{i=1}^{n} w_i - 1 &= 0 \\ 2\mu\left(\sum_{i=1}^{n} w_i - 1\right) &= 0 \end{aligned}$$

The addition of this new term, which does not affect the equality, is all we need to convert our constrained minimization problem into an unconstrained one. The error variance for the model, as expressed in Equation 12.9, is now a function of $n+1$ variables, the n weights and the one Lagrange parameter. By setting the $n+1$ partial first derivatives to 0 with respect to each of these variables, we will have a system of $n+1$ equations and $n+1$ unknowns. Furthermore, setting the partial first derivative to 0 with respect to μ will produce our unbiasedness condition. The first three terms in Equation 12.9 do not contain μ, and do not affect the partial first derivative with respect to μ:

$$\begin{aligned} \frac{\partial(\tilde{\sigma}_R^2)}{\partial\mu} &= \frac{\partial\left(2\mu\left(\sum_{i=1}^{n} w_i - 1\right)\right)}{\partial\mu} \\ &= 2\sum_{i=1}^{n} w_i - 2 \end{aligned}$$

Setting this quantity to 0 produces the unbiasedness condition:

$$\sum_{i=1}^{n} w_i = 1$$

Since the unbiasedness condition is already included in the $n+1$ equations that result from the differentiation of $\tilde{\sigma}_R^2$, the solution of those $n+1$ equations will produce the set of weights that minimizes $\tilde{\sigma}_R^2$ under the constraint that the weights sum to 1. This solution will also provide a value for μ that, as we will see later, is useful for calculating the resulting minimized error variance.

Minimization of the Error Variance

We will now minimize the error variance by calculating the $n+1$ partial first derivatives of Equation 12.9 and setting each one to 0. The differentiation with respect to w_1 is given in detail; the partial first derivatives with respect to the other weights can be calculated in a similar manner.

The first term on the right-hand side of Equation 12.9 does not depend on w_1, and therefore does not affect the derivative with respect to w_1. Expanding the double summation in the second term and dropping all terms that do not include w_1 gives us

$$\begin{aligned}\frac{\partial(\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{ij})}{\partial w_1} &= \frac{\partial(w_1^2\tilde{C}_{11} + 2w_1\sum_{j=2}^{n} w_j\tilde{C}_{1j})}{\partial w_1} \\ &= 2w_1\tilde{C}_{11} + 2\sum_{j=2}^{n} w_j\tilde{C}_{1j} \\ &= 2\sum_{j=1}^{n} w_j\tilde{C}_{1j}\end{aligned}$$

The third term on the right-hand side of Equation 12.9 contains only one term that involves w_1:

$$\frac{\partial(\sum_{i=1}^{n} w_i\tilde{C}_{i0})}{\partial w_1} = \frac{\partial(w_1\tilde{C}_{10})}{\partial w_1} = \tilde{C}_{10}$$

The last term on the right-hand side of Equation 12.9 also contains only one term that involves w_1:

$$\frac{\partial(\mu(\sum_{i=1}^{n} w_i - 1))}{\partial w_1} = \frac{\partial(\mu w_1)}{\partial w_1} = \mu$$

The first derivative of $\tilde{\sigma}_R^2$ with respect to w_1 can now be written as

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_1} = 2\sum_{j=1}^{n} w_j\tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu$$

Setting this to 0 produces the following equation:

$$\begin{aligned}2\sum_{j=1}^{n} w_j\tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu &= 0 \\ \sum_{j=1}^{n} w_j\tilde{C}_{1j} + \mu &= \tilde{C}_{10} \qquad (12.10)\end{aligned}$$

The differentiation with respect to the other weights produces similar equations:

$$\begin{aligned}\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_1} = 2\textstyle\sum_{j=1}^{n} w_j\tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu = 0 &\Rightarrow \textstyle\sum_{j=1}^{n} w_j\tilde{C}_{1j} + \mu = \tilde{C}_{10} \\ \vdots \qquad\qquad &\qquad\vdots \\ \frac{\partial(\tilde{\sigma}_R^2)}{\partial w_i} = 2\textstyle\sum_{j=1}^{n} w_j\tilde{C}_{ij} - 2\tilde{C}_{i0} + 2\mu = 0 &\Rightarrow \textstyle\sum_{j=1}^{n} w_j\tilde{C}_{ij} + \mu = \tilde{C}_{i0} \\ \vdots \qquad\qquad &\qquad\vdots \\ \frac{\partial(\tilde{\sigma}_R^2)}{\partial w_n} = 2\textstyle\sum_{j=1}^{n} w_j\tilde{C}_{nj} - 2\tilde{C}_{n0} + 2\mu = 0 &\Rightarrow \textstyle\sum_{j=1}^{n} w_j\tilde{C}_{nj} + \mu = \tilde{C}_{n0}\end{aligned}$$

As we noted in the last section, the setting of the partial first derivative to 0 with respect to μ produces the unbiasedness condition.

The set of weights that minimize the error variance under the constraint that they sum to 1 therefore satisfies the following $n+1$ equations:

$$\sum_{j=1}^{n} w_j\tilde{C}_{ij} + \mu = \tilde{C}_{i0} \quad \forall\, i = 1,\ldots,n \qquad (12.11)$$

$$\sum_{i=1}^{n} w_i = 1 \qquad (12.12)$$

This system of equations, often referred to as the *ordinary kriging system*, can be written in matrix notation as

$$\underset{}{\mathbf{C}} \cdot \underset{}{\mathbf{w}} = \underset{}{\mathbf{D}}$$

$$\underbrace{\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{(n+1)\times(n+1)} \cdot \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}_{(n+1)\times 1} = \underbrace{\begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}}_{(n+1)\times 1} \qquad (12.13)$$

To solve for the weights, we multiply Equation 12.13 on both sides by $\mathbf{C}^{-1}$, the inverse of the left-hand side covariance matrix:

$$\begin{aligned}\mathbf{C}\cdot\mathbf{w} &= \mathbf{D} \\ \mathbf{C}^{-1}\cdot\mathbf{C}\cdot\mathbf{w} &= \mathbf{C}^{-1}\cdot\mathbf{D} \\ \mathbf{I}\cdot\mathbf{w} &= \mathbf{C}^{-1}\cdot\mathbf{D} \\ \mathbf{w} &= \mathbf{C}^{-1}\cdot\mathbf{D} \qquad (12.14)\end{aligned}$$

After a considerable amount of mathematics, we have finally arrived at a solution. To minimize the modeled error variance, we first need to choose the $(n+1)^2$ covariances that will describe the spatial continuity in our random function model. In practice this is typically done by choosing a function $\tilde{C}(\mathbf{h})$, and calculating all of the required covariances from this function. Once the $(n+1)^2$ covariances have been chosen, the **C** and **D** matrices can be built. The set of weights that will produce unbiased estimates with the minimum error variance for our random function model is given by Equation 12.14.

Having gone to considerable trouble to minimize the error variance, we may be interested in knowing this minimum value. We could substitute the weights we have obtained into Equation 12.8 to find the actual value of the minimized error variance. There is also a quicker way that avoids the n^2 terms in the double summation. Multiplying each of the n equations given in Equation 12.11 by w_i produces the following result:

$$w_i(\sum_{j=1}^{n} w_j\tilde{C}_{ij} + \mu) = w_i\tilde{C}_{i0} \quad \forall\, i = 1,\ldots,n$$

Summing these n equations leads to an expression for the double summation:

$$\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j\tilde{C}_{ij} + \sum_{i=1}^{n} w_i\mu = \sum_{i=1}^{n} w_i\tilde{C}_{i0}$$

$$\sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j\tilde{C}_{ij} = \sum_{i=1}^{n} w_i\tilde{C}_{i0} - \sum_{i=1}^{n} w_i\mu$$

Since the weights sum to 1, the last term is simply μ, which gives us

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_iw_j\tilde{C}_{ij} = \sum_{i=1}^{n} w_i\tilde{C}_{i0} - \mu$$

Substituting this into Equation 12.8 allows us to express the minimized error variance as

$$\begin{aligned}\tilde{\sigma}_R^2 &= \tilde{\sigma}^2 + \sum_{i=1}^{n} w_i\tilde{C}_{i0} - \mu - 2\sum_{i=1}^{n} w_i\tilde{C}_{i0} \\ &= \tilde{\sigma}^2 - (\sum_{i=1}^{n} w_i\tilde{C}_{i0} + \mu) \qquad (12.15)\end{aligned}$$

Or, in terms of the matrices we defined earlier,

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 - \mathbf{w}\cdot\mathbf{D} \qquad (12.16)$$

This minimized error variance is usually referred to as the *ordinary kriging variance*, for which we will use the notation σ_{OK}^2; though the tilde has been dropped from the notation, the OK subscript should serve as a reminder that this error variance was calculated from a model.

Ordinary Kriging Using γ or ρ

When we derived the expression for the error variance, we assumed that the random variables in our random function model all had the same mean and variance. These two assumptions also allow us to develop the following relationship between the model variogram and the model covariance:

$$\begin{aligned}\gamma_{ij} &= \frac{1}{2}E\{[V_i - V_j]^2\} \\ &= \frac{1}{2}E\{V_i^2\} + \frac{1}{2}E\{V_j^2\} - E\{V_i\cdot V_j\} \\ &= E\{V^2\} - E\{V_i\cdot V_j\} \\ &= E\{V^2\} - \tilde{m}^2 - [E\{V_i\cdot V_j\} - \tilde{m}^2] \\ &= \tilde{\sigma}^2 - \tilde{C}_{ij} \qquad (12.17)\end{aligned}$$

There is also a relationship between the model correlogram and the model covariance:

$$\tilde{\rho}_{ij} = \frac{\tilde{C}_{ij}}{\tilde{\sigma}^2} \qquad (12.18)$$

These relationships are valid for a random function model in which we have made the assumptions that the random variables all have the same mean and variance. This does not entail that the same relationships exist between the variogram, covariance, and correlation functions of an actual data set. Nevertheless, the validity of these relationships for our random function model allows us to express the ordinary kriging equations in terms of the variogram or the correlogram.

In terms of the variogram, the ordinary kriging system can be written as

$$\sum_{j=1}^{n} w_j\tilde{\gamma}_{ij} - \mu = \tilde{\gamma}_{i0} \quad \forall\, i = 1,\ldots,n \qquad (12.19)$$

$$\sum_{i=1}^{n} w_i = 1$$

with the modeled error variance given by

$$\tilde{\sigma}_R^2 = \sum_{i=1}^{n} w_i \tilde{\gamma}_{i0} + \mu \qquad (12.20)$$

In terms of the correlogram, the ordinary kriging system can be written as

$$\sum_{j=1}^{n} w_j \tilde{\rho}_{ij} + \mu = \tilde{\rho}_{i0} \quad \forall\, i = 1, \ldots, n \qquad (12.21)$$

$$\sum_{i=1}^{n} w_i = 1$$

with the modeled error variance given by

$$\tilde{\sigma}_R^2 = \tilde{\sigma}^2 (1 - (\sum_{i=1}^{n} w_i \tilde{\rho}_{i0} + \mu)) \qquad (12.22)$$

The common practice in geostatistics is to calculate modeled variogram values then, for reasons of computational efficiency, to subtract them from some constant, usually $\tilde{\sigma}^2$. The net result is that although geostatisticians eventually resort to solving the ordinary kriging equations in terms of covariances, most of the initial calculations are done in terms of variograms [2].

An Example of Ordinary Kriging

Once the ordinary kriging method has been developed, several small examples will be given to demonstrate how the various model parameters affect the estimates. Finally, to allow a comparison of ordinary kriging with the other point estimation methods we have seen earlier, the case study from the previous chapter will be extended to include estimates calculated by ordinary kriging.

Let us return to the seven sample data configuration we used earlier to see a specific example of how ordinary kriging is done. The data configuration is shown again in Figure 12.1; we have labeled the point we are estimating as location 0, and the sample locations as 1 through

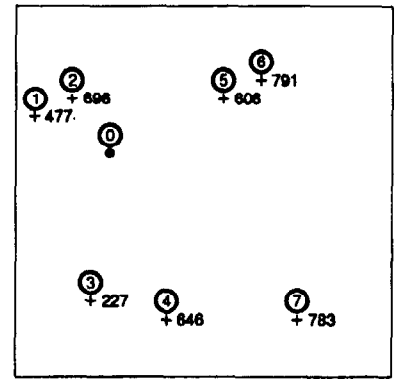


Figure 12.1 An example of a data configuration to illustrate the kriging estimator. This configuration was given earlier in Figure 11.1 where it was used to illustrate other estimation techniques discussed in Chapter 11. The sample value is given immediately to the right of the plus sign.

Table 12.1 Coordinates and sample values for the data shown in Figure 12.1.

Sample No.	X	Y	V	Distance from 65E,137N
1	225	61	139	477	4.5
2	437	63	140	696	3.6
3	367	64	129	227	8.1
4	52	68	128	646	9.5
5	259	71	140	606	6.7
6	436	73	141	791	8.9
7	366	75	128	783	13.5

7. The coordinates of these eight points are given in Table 12.1, along with the available sample values.

To calculate the ordinary kriging weights, we must first decide what pattern of spatial continuity we want our random function model to have. To keep this example relatively simple, we will calculate all of

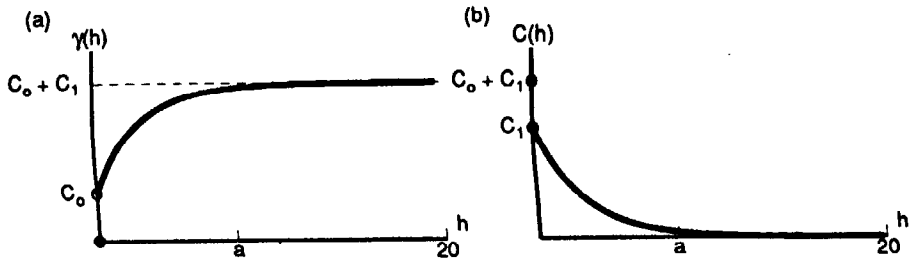


Figure 12.2 An example of an exponential variogram model (a) and an exponential covariance function (b).

our covariances from the following function:

$$\tilde{C}(\mathbf{h}) = \begin{cases} C_0 + C_1 & \text{if } |\mathbf{h}| = 0 \\ C_1 exp(\frac{-3|\mathbf{h}|}{a}) & \text{if } |\mathbf{h}| > 0 \end{cases} \tag{12.23}$$

Using Equation 12.17, this covariance function corresponds to the following variogram:

$$\tilde{\gamma}(\mathbf{h}) = \begin{cases} 0 & \text{if } |\mathbf{h}| = 0 \\ C_0 + C_1(1 - exp(\frac{-3|\mathbf{h}|}{a})) & \text{if } |\mathbf{h}| > 0 \end{cases} \tag{12.24}$$

Both of these functions, shown in Figure 12.2, can be described by the following parameters:

- C_0, commonly called the *nugget effect*, which provides a discontinuity at the origin.
- a, commonly called the *range*, which provides a distance beyond which the variogram or covariance value remains essentially constant.
- $C_0 + C_1$, commonly called the *sill* [3], which is the variogram value for very large distances, $\gamma(\infty)$. It is also the covariance value for $|\mathbf{h}| = 0$, and the variance of our random variables, $\tilde{\sigma}^2$.

Geostatisticians normally define the spatial continuity in their random function model through the variogram and solve the ordinary

Table 12.2 A table of distances, from Figure 12.1, between all possible pairs of the seven data locations.

	distance							
Location	0	1	2	3	4	5	6	7
0	0.00	4.47	3.61	8.06	9.49	6.71	8.94	13.45
1	4.47	0.00	2.24	10.44	13.04	10.05	12.17	17.80
2	3.61	2.24	0.00	11.05	13.00	8.00	10.05	16.97
3	8.06	10.04	11.05	0.00	4.12	13.04	15.00	11.05
4	9.49	13.04	13.00	4.12	0.00	12.37	13.93	7.00
5	6.71	10.05	8.00	13.04	12.37	0.00	2.24	12.65
6	8.94	12.17	10.05	15.00	13.93	2.24	0.00	13.15
7	13.45	17.80	16.97	11.05	7.00	12.65	13.15	0.00

kriging system using the covariance. In this example, we will use the covariance function throughout.

By using the covariance function given in Equation 12.23, we have chosen to ignore the possibility of anisotropy for the moment; the covariance between the data values at any two locations will depend only on the distance between them and not on the direction. Later, when we examine the effect of the various parameters, we will also study the important effect of anisotropy.

To demonstrate how ordinary kriging works, we will use the following parameters for the function given in Equation 12.23:

$$C_0 = 0, \quad a = 10, \quad C_1 = 10$$

These are not necessarily good choices, but they will make the details of the ordinary kriging procedure easier to follow since our covariance model now has a quite simple expression:

$$\tilde{C}(\mathbf{h}) = 10e^{-0.3|\mathbf{h}|} \tag{12.25}$$

Having chosen a covariance function from which we can calculate all the covariances required for our random function model, we can now build the **C** and **D** matrices. Using Table 12.2, which provides the

distances between every pair of locations, and Equation 12.25 above, the **C** matrix is

$$\mathbf{C} = \begin{bmatrix} \tilde{C}_{11} & \tilde{C}_{12} & \tilde{C}_{13} & \tilde{C}_{14} & \tilde{C}_{15} & \tilde{C}_{16} & \tilde{C}_{17} & 1 \\ \tilde{C}_{21} & \tilde{C}_{22} & \tilde{C}_{23} & \tilde{C}_{24} & \tilde{C}_{25} & \tilde{C}_{26} & \tilde{C}_{27} & 1 \\ \tilde{C}_{31} & \tilde{C}_{32} & \tilde{C}_{33} & \tilde{C}_{34} & \tilde{C}_{35} & \tilde{C}_{36} & \tilde{C}_{37} & 1 \\ \tilde{C}_{41} & \tilde{C}_{42} & \tilde{C}_{43} & \tilde{C}_{44} & \tilde{C}_{45} & \tilde{C}_{46} & \tilde{C}_{47} & 1 \\ \tilde{C}_{51} & \tilde{C}_{52} & \tilde{C}_{53} & \tilde{C}_{54} & \tilde{C}_{55} & \tilde{C}_{56} & \tilde{C}_{57} & 1 \\ \tilde{C}_{61} & \tilde{C}_{62} & \tilde{C}_{63} & \tilde{C}_{64} & \tilde{C}_{65} & \tilde{C}_{66} & \tilde{C}_{67} & 1 \\ \tilde{C}_{71} & \tilde{C}_{72} & \tilde{C}_{73} & \tilde{C}_{74} & \tilde{C}_{75} & \tilde{C}_{76} & \tilde{C}_{77} & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 10.00 & 5.11 & 0.44 & 0.20 & 0.49 & 0.26 & 0.05 & 1.00 \\ 5.11 & 10.00 & 0.36 & 0.20 & 0.91 & 0.49 & 0.06 & 1.00 \\ 0.44 & 0.36 & 10.00 & 2.90 & 0.20 & 0.11 & 0.36 & 1.00 \\ 0.20 & 0.20 & 2.90 & 10.00 & 0.24 & 0.15 & 1.22 & 1.00 \\ 0.49 & 0.91 & 0.20 & 0.24 & 10.00 & 5.11 & 0.22 & 1.00 \\ 0.26 & 0.49 & 0.11 & 0.15 & 5.11 & 10.00 & 0.19 & 1.00 \\ 0.05 & 0.06 & 0.36 & 1.22 & 0.22 & 0.19 & 10.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \end{bmatrix}$$

The **D** matrix is

$$\mathbf{D} = \begin{bmatrix} \tilde{C}_{10} \\ \tilde{C}_{20} \\ \tilde{C}_{30} \\ \tilde{C}_{40} \\ \tilde{C}_{50} \\ \tilde{C}_{60} \\ \tilde{C}_{70} \\ 1 \end{bmatrix} = \begin{bmatrix} 2.61 \\ 3.39 \\ 0.89 \\ 0.58 \\ 1.34 \\ 0.68 \\ 0.18 \\ 1.00 \end{bmatrix}$$

The inverse of **C** is

$$\mathbf{C}^{-1} = \begin{bmatrix} 0.127 & -0.077 & -0.013 & -0.009 & -0.008 & -0.009 & -0.012 & 0.136 \\ -0.077 & 0.129 & -0.010 & -0.008 & -0.015 & -0.008 & -0.011 & 0.121 \\ -0.013 & -0.010 & 0.098 & -0.042 & -0.010 & -0.010 & -0.014 & 0.156 \\ -0.009 & -0.008 & -0.042 & 0.102 & -0.009 & -0.009 & -0.024 & 0.139 \\ -0.008 & -0.015 & -0.010 & -0.009 & 0.130 & -0.077 & -0.012 & 0.118 \\ -0.009 & -0.008 & -0.010 & -0.009 & -0.077 & 0.126 & -0.013 & 0.141 \\ -0.012 & -0.011 & -0.014 & -0.024 & -0.012 & -0.013 & 0.085 & 0.188 \\ 0.136 & 0.121 & 0.156 & 0.139 & 0.118 & 0.141 & 0.188 & -2.180 \end{bmatrix}$$

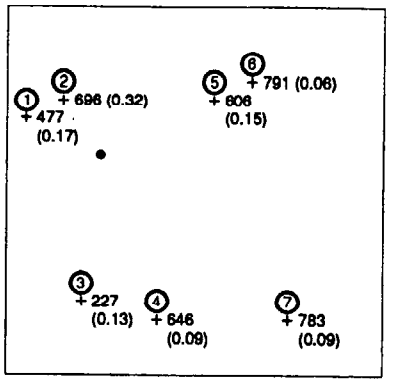


Figure 12.3 The ordinary kriging weights for the seven samples using the isotropic exponential covariance model given in Equation 12.25. The sample value is given immediately to the right of the plus sign while the kriging weights are shown in parenthesis.

The set of weights that will provide unbiased estimates with a minimum estimation variance is calculated by multiplying $\mathbf{C}^{-1}$ by **D**:

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ \mu \end{bmatrix} = \mathbf{C}^{-1} \cdot \mathbf{D} = \begin{bmatrix} 0.173 \\ 0.318 \\ 0.129 \\ 0.086 \\ 0.151 \\ 0.057 \\ 0.086 \\ 0.907 \end{bmatrix}$$

Figure 12.3 shows the sample values along with their corresponding weights. The resulting estimate is

$$\begin{aligned} \hat{v}_0 &= \sum_{i=1}^{n} w_i v_i \\ &= (0.173)(477) + (0.318)(696) + (0.129)(227) + (0.086)(646) + \\ &\quad (0.151)(606) + (0.057)(791) + (0.086)(783) \\ &= 592.7 \text{ ppm} \end{aligned}$$

The minimized estimation variance is

$$\begin{aligned}\tilde{\sigma}_R^2 &= \tilde{\sigma}^2 - \sum_{i=1}^{n} w_i \tilde{C}_{i0} + \mu \\ &= 10 - (0.173)(2.61) - (0.318)(3.39) - (0.129)(0.89) - \\ &\quad (0.086)(0.58) - (0.151)(1.34) - (0.057)(0.68) - \\ &\quad (0.086)(0.18) + 0.907 \\ &= 8.96 \ \text{ppm}^2\end{aligned}$$

Ordinary Kriging and the Model of Spatial Continuity

Earlier, when we tackled the problem of unbiasedness, the unbiasedness condition did not require us to specify any parameters of our random function model. Though we had to assume that the mean of the random variables was the same, we did not have to specify its actual value; the condition that the weights sum to one does not involve the parameter $\tilde{m}$. In the minimization of $\tilde{\sigma}_R^2$, however, our solution does involve model parameters; the ordinary kriging weights and the resulting minimized error variance directly depend on our choice of the covariances for the **C** and **D** matrices.

The choice of a covariance model (or, if one prefers, a variogram model or a correlogram model) is a prerequisite for ordinary kriging. Though this makes ordinary kriging more time consuming than the estimation procedures we looked at in the previous chapter, it also makes it more flexible. We saw earlier that the exponent for inverse distance estimation gave us an ability to modulate the estimation procedure from a polygonal estimation to a moving average estimation. The covariance model in ordinary kriging provides a similar but much more powerful ability to customize the ordinary kriging estimation procedure. In addition to allowing us to modulate between polygonal estimates and moving average estimates, the covariance model also provides a vehicle for incorporating valuable qualitative insights such as the pattern of anisotropy.

In practice, the pattern of spatial continuity chosen for the random function model is usually taken from the spatial continuity evident in the sample data set. Once the sample variogram has been calculated, a function is fit to it; Chapter 16 discusses this procedure in detail. There are two reasons why the sample variogram cannot be used directly in the ordinary kriging system.

First, the **D** matrix may call for variogram values for distances that are not available from the sample data. There are often situations in which the distance from the point being estimated to a particular sample is smaller than the distance between any pair of available samples. Since the sample data set cannot provide any pairs for these small distances, we must rely on a function that provides variogram values for all distances and directions, even those that are not available from sample data.

Second, the use of the sample variogram does not guarantee the existence and uniqueness of the solution to the ordinary kriging system. The system of $n+1$ equations and $n+1$ unknowns described by Equation 12.13 does not necessarily have a unique solution. Certain choices of the covariances in the **C** and **D** matrices may cause the system to have no solution; other choices may cause the system to have several solutions. To be guaranteed of having one and only one solution, we must ensure that our system has a property known as *positive definiteness*. Even if the sample data are regularly gridded and all of the distances for which the **D** matrix requires values are available from sample data, the use of the sample variogram, unfortunately, does not guarantee positive definiteness. There are many ways of checking for positive definiteness [4]; in practice, however, we guarantee the existence and uniqueness of our solution by fitting the sample variogram with functions that are known to be positive definite.

Though fitting a function to the sample variogram is the most common approach to choosing the pattern of spatial continuity for the random function model, it is not the only one nor is it necessarily the best one. There are many situations in which it is better to base the choice of a pattern of spatial continuity on a more qualitative interpretation. Experience with similar data sets may often be a better guide than pattern of spatial continuity shown by too few available samples.

Frequently, the sample data set does not show any clear pattern of spatial continuity. The lack of evident structure in the available samples does not justify using a spatially uncorrelated random function model. In earth science data sets there is nearly always some pattern of spatial continuity. It may not be evident from the available samples due to their insufficient number, sampling error, erratic values, or possible outlier values.

Even in situations where the sample data set does exhibit a clear pattern of spatial continuity, the decision to use the sample spatial

continuity for the random function model should still be considered carefully. If the samples are clustered in particular areas, one should consider how appropriate the sample variogram is for the points that will be estimated. In the Walker Lake sample data set, for example, the preferential clustering of the samples in the Wassuk Range anomaly causes the sample variogram to be more representative of that particular region than of the entire area. If we intend to perform estimation only at locations within this anomalous area, then the sample variogram may be appropriate. If we intend to calculate estimates throughout the entire Walker Lake area, however, then the use of the sample variogram for our random function model is questionable.

The decision to use the sample spatial continuity should be carefully considered even for sample data sets in which clustering is not a significant problem. For example, the anisotropy may not be adequately captured by an analysis of the sample data set. As we will see shortly, anisotropy is an important element of the pattern of spatial continuity in the random function model. The analysis of spatial continuity in a sample data set usually involves the calculation and summary of h-scatterplots for particular directions. As explained in Chapter 7, the use of a tolerance on the direction is necessary in practice; unfortunately, the use of a directional tolerance may cause the anisotropy evident from sample variograms to be weaker than that which would be observed if exhaustive information was available. This is clear from our analysis of the spatial continuity in the exhaustive data set in Chapter 5 and of the sample data set in Chapter 7. The sample variograms and covariance functions show less anisotropy than do the exhaustive ones.

In Chapter 16 we will discuss the practical details of fitting functions to sample variograms and deriving a mathematical expression that provides variogram values for any separation vector $\mathbf{h}$. Though the fitting of functions to sample variograms is certainly the most common approach to choosing the pattern of spatial continuity for the random function model, it should not be viewed as the only correct approach. In every study that uses geostatistical estimation methods, the geostatistician must choose the pattern of spatial continuity. The use of the most common approach does not remove the responsibility of making this choice wisely, nor does it remove the responsibility of understanding the effect of one's chosen model on the estimation procedure.

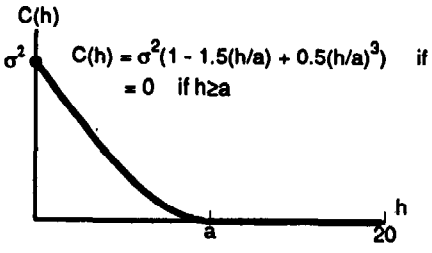


Figure 12.4 The spherical model of a covariance function.

An Intuitive Look at Ordinary Kriging

To understand how parameters such as the nugget effect, the range and the sill affect our estimates, it will help to have a more intuitive understanding of the role of the $\mathbf{C}$ and $\mathbf{D}$ matrices used in the ordinary kriging system. To many practitioners, the mathematical development of the ordinary kriging system presented earlier is tedious and virtually incomprehensible. The following explanation is not rigorous and may provide a more intuitive feel for what the ordinary kriging procedure is doing. While the earlier development provided a rationale for the procedure, the following one probably provides a better understanding of its practical success. Understanding the role of the $\mathbf{C}$ and $\mathbf{D}$ matrices in intuitive terms also allows the practitioner to make ad hoc adjustments that, despite their lack of apparent theoretical rigor, may actually improve the estimation procedure.

Taken by itself, the $\mathbf{D}$ matrix on the right-hand side of Equation 12.13 provides a weighting scheme similar to that of the inverse distance methods. Like an inverse distance weight, the covariance between any particular sample and the point being estimated generally decreases as the sample gets farther away. This can be seen in the example used in the previous section: sample 2 is closest to the point being estimated and $\tilde{C}_{20}$ is the largest covariance in $\mathbf{D}$; sample 7 is the farthest, and $\tilde{C}_{70}$ is the smallest covariance in $\mathbf{D}$. Unlike inverse distance weights, which are limited to the form $|\mathbf{h}|^{-p}$, the covariances calculated for our model can come from a much larger family of functions. For example, the covariance function shown in Figure 12.4 provides decreasing weights up to some distance, a, and provides a weight of 0 for distances greater than a.

The **D** matrix therefore contains a type of inverse distance weighting in which the "distance" is not the geometric distance to the sample but rather its statistical distance. What really distinguishes ordinary kriging from inverse distance methods, however, is not the use of statistical distance instead of geometric distance, but rather the role played by the **C** matrix. From our earlier example, it is clear that the multiplication of **D** by $\mathbf{C}^{-1}$ does considerably more than rescale the covariances in **D** so that they sum to one. For example, sample 4 is farther from the point being estimated than is sample 6; this is recorded in the **D** matrix by the fact that $\tilde{C}_{40}$ is smaller than $\tilde{C}_{60}$. The ordinary kriging weight for sample 4, however, is larger than that for sample 6.

The **C** matrix records distances between each sample and every other sample, providing the ordinary kriging system with information on the clustering of the available sample data. If two samples are very close to each other, this will be recorded by a large value in the **C** matrix; if two samples are far apart, this will be recorded by a low value. The multiplication of **D** by $\mathbf{C}^{-1}$ adjusts the raw inverse statistical distance weights in **D** to account for possible redundancies between the samples.

In our earlier example, though sample 6 was the closer than sample 4 to the point we were trying to estimate, its usefulness was reduced by its proximity to sample 5. When **D** is multiplied by the inverse of **C**, the net result is that some of the weight that was allocated to sample 6 because of its closeness to the point being estimated is redistributed to other samples that are farther away yet less redundant.

Like the **D** matrix, the **C** matrix records the distance in terms of statistical distance rather than geometric distance. The possible redundancy between samples depends not simply on the distance between them but also on the spatial continuity. For example, two measurements of the elevation of the water table taken 10 m apart are likely to be much more redundant than two measurements of the gold concentration in a vein-type gold deposit also taken 10 m apart. The fact that the elevation of the water table is much more spatially continuous than the concentration of gold would be captured by the covariance functions of the two data sets. It makes good sense, therefore, that clustering information be recorded in terms of the statistical distance, using some measure of spatial continuity such as the covariance or the variogram.

The ordinary kriging system therefore takes into account two of the

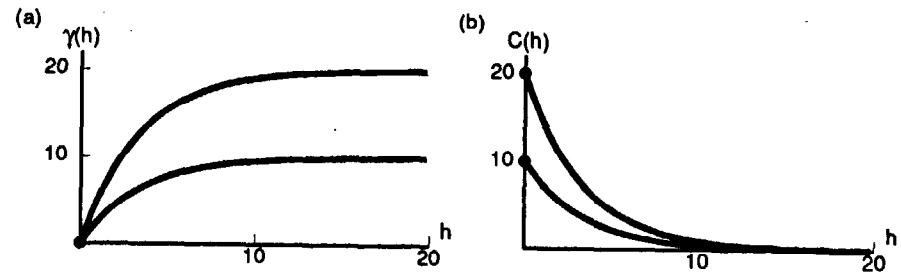


Figure 12.5 Two variograms and their corresponding covariance functions that differ only by their scale.

important aspects of estimation problems, distance, and clustering:

$$\mathbf{w} = \underbrace{\mathbf{C}^{-1}}_{\text{Clustering}} \cdot \underbrace{\mathbf{D}}_{\text{Distance}}$$

The information on the distances to the various samples and the clustering between the samples is all recorded in terms of a statistical distance, thereby customizing the estimation procedure to a particular pattern of spatial continuity.

Variogram Model Parameters

We will now look at how the various parameters of the covariance or variogram model affect the ordinary kriging weights. Much of the terminology for these parameters has evolved from the traditional use of the variogram, and we will refer to variogram models throughout this section. The same remarks apply to covariance functions and correlograms; in the following examples we will show the covariance functions that correspond to the variograms we discuss so that it is clear how the changes in various parameters manifest themselves on the covariance function.

The following observations serve as additional support for the argument that even with regularly gridded and well-behaved sample data, the exercise of fitting a function to the sample variogram model involves important choices on the part of the practitioner. As we noted earlier, the sample variogram does not provide any information for distance shorter than the minimum spacing between the sample data. Unless

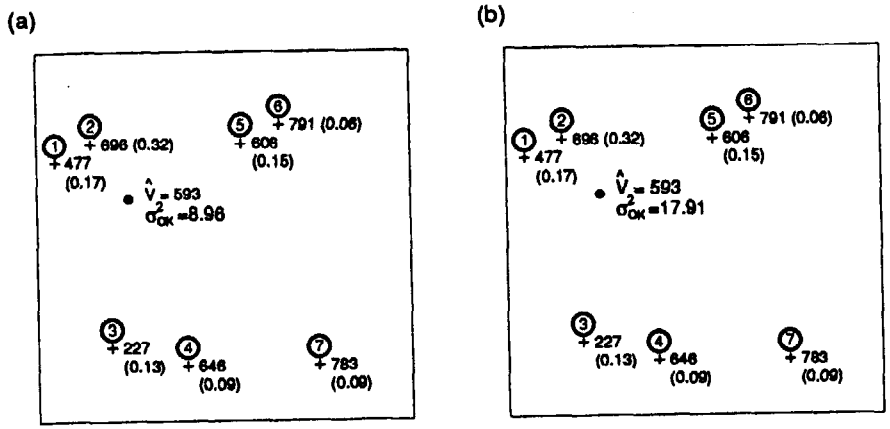


Figure 12.6 Ordinary kriging results using two different covariance functions that differ only in their scale. (a) shows the kriging weights for the variogram with a sill of 20 while (b) shows the weights for a sill of 10. The two covariance functions are given in Figure 12.5.

the sampling includes duplicates at the same location, the nugget effect and the behavior of the variogram near the origin can not be determined from the sample variogram. Yet the following examples demonstrate that these two parameters have the biggest effect on the ordinary kriging weights and on the resulting estimate.

The Effect of Scale. Figure 12.5 shows two variogram models that differ only in their scale. $\gamma_1(h)$ is the variogram that corresponds to the covariance function we chose for our earlier detailed example of ordinary kriging; $\gamma_2(h)$ has the same shape as $\gamma_1(h)$, but it is exactly twice as big:

$$\begin{aligned} \gamma_1(h) &= 10(1 - e^{-.3|h|}) \\ \gamma_2(h) &= 20(1 - e^{-.3|h|}) = 2\gamma_1(h) \end{aligned}$$

The results of using these two covariance models for ordinary kriging are shown in Figure 12.6. Rescaling the variogram values has not affected the ordinary kriging weights or the ordinary kriging estimate; however, it has affected the ordinary kriging variance. These effects will be observed with any rescaling; while the estimate itself is unchanged,

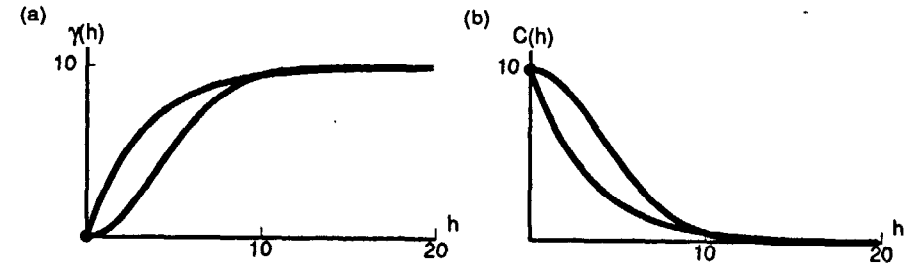


Figure 12.7 Two variograms and their corresponding covariance functions that differ only in their shape. The model that has a parabolic behavior near the origin is the Gaussian model, while the other is the exponential model.

the estimation variance increases by the same factor that was used to scale the variogram [5].

The Effect of Shape. Figure 12.7 shows two variogram models that reach the same sill but have different shapes. $\gamma_1(h)$ is the same as the first variogram model from the previous example:

$$\gamma_1(h) = 10(1 - exp(-3\frac{|h|}{10}))$$

$\gamma_2(h)$ has a similar expression, but the square in the exponent causes it to behave more like a parabola near the origin:

$$\gamma_2(h) = 10(1 - exp(-3(\frac{|h|}{10})^2))$$

The results of using these two variogram models for ordinary kriging are shown in Figure 12.8. With the second model, more weight is given to the three values that surround the point being estimated (the same three that were used for the triangulation estimate in Chapter 11); the remaining points all receive less weight, with most of them actually receiving a negative weight. A parabolic behavior near the origin is indicative of a very continuous phenomena so the estimation procedure makes much more use of the closest samples.

The appearance of negative weights is a result of an effect often referred to as the screen effect. A particular sample is said to be screened if another sample falls between it and the point being estimated. For the data configuration we are using, sample 6 is strongly screened by

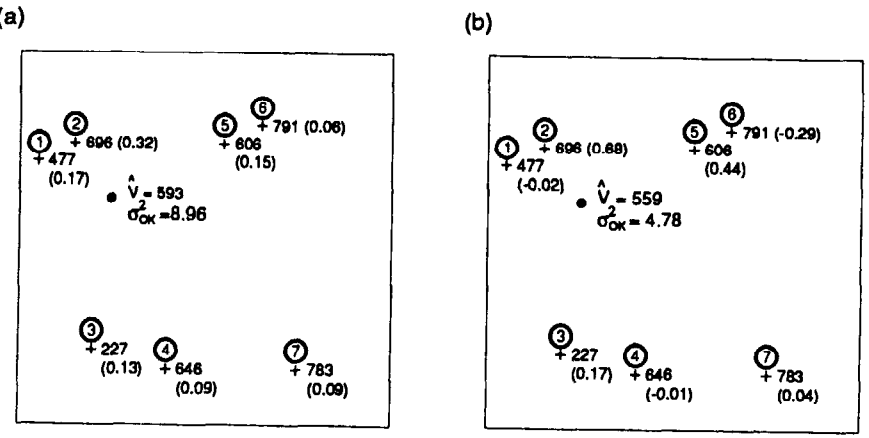


Figure 12.8 Ordinary kriging results using two different covariance functions that differ only in their shape. (a) shows the kriging weights for a exponential model while (b) shows the weights for a Gaussian model. The two covariance functions are given in Figure 12.7.

sample 5; to a lesser extent, sample 1 is partially screened by sample 2. It makes sense to reduce the weights of samples that are partially or totally screened by others; this is part of what the multiplication by $\mathbf{C}^{-1}$ accomplishes. The degree to which screened samples lose their influence depends on the pattern of spatial continuity. The use of a variogram with a parabolic behavior near the origin will cause the screen effect to be much more pronounced, often producing negative weights even larger than the ones we calculated in our last example.

Even with variogram models that are linear rather than parabolic near the origin, it is possible to produce negative weights for samples that are strongly screened by others. None of the other estimation procedures we looked at earlier can produce negative weights. The advantage of a procedure that can assign weights smaller than 0 or greater than 1 (but still respect the constraint that the sum of the weights is 1) is that it can yield estimates larger than the largest sample value or smaller than the smallest sample value. All procedures that restrict the weights to be between 0 and 1 can only produce estimates that are between the minimum and maximum sample values. It is unlikely that the sample data set includes the most extreme values and it is reasonable to imagine that the true values we are trying to estimate may be beyond the extremes of the available sample values. The disadvantage of negative weights is that they also create the possibility of negative estimates if a particularly high sample value is associated with a negative weight. In most earth science applications, the variable being estimated is necessarily positive. Ore grades and tonnages, porosities, permeabilities, pollutant concentrations, densities, depths to geologic horizons, and thicknesses of strata are common examples of variables that one may be interested in estimating and that are never negative. For such variables, if ordinary kriging produces estimates that are negative, one is perfectly justified in setting such estimates to 0.

For data sets in which the variable of interest is indeed extremely continuous, such as the depth to a particular horizon or the thickness of a certain zone, the sample variogram often shows a definite parabolic behavior near the origin. Even in such situations where the spatial continuity of the sample data set can correctly be extended to the points at which estimates will be required, variogram models with parabolic behavior near the origin are avoided in practice since the negative weights they may produce tend to make the estimation very erratic.

The Nugget Effect. Figure 12.9 shows two variogram models that differ only in their nugget effect. While $\gamma_1(h)$ has no nugget effect, $\gamma_2(h)$ has a nugget effect that is 50% of the sill:

$$\gamma_1(h) = 10(1 - e^{-.3|h|})$$

$$\gamma_2(h) = \begin{cases} 0 & \text{if } h = 0 \\ 5 + 5(1 - e^{-.3|h|}) & \text{if } h > 0 \end{cases}$$

The results of using these two variogram models for ordinary kriging are shown in Figure 12.10. The weights calculated using $\gamma_2(h)$ are more similar to one another than are those calculated using $\gamma_1(h)$. With $\gamma_2(h)$, the smallest weight is 0.125 and the largest weight is 0.178; with $\gamma_1(h)$, the smallest weight is 0.057 and the largest weight is 0.318. The more equal distribution of weight causes the estimated value to be somewhat higher. The other noticeable result of using a higher nugget effect is that the ordinary kriging variance is higher.

Increasing the nugget effect makes the estimation procedure become more like a simple averaging of the available data. If the vari-

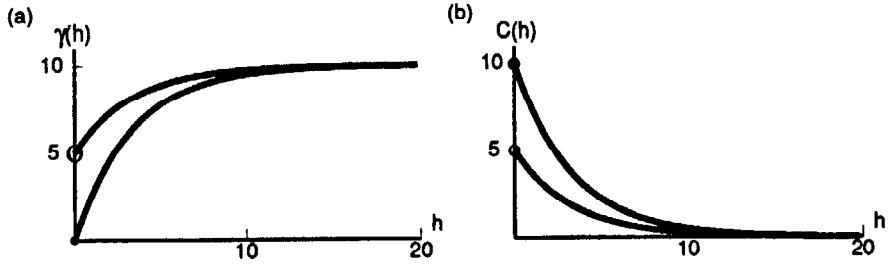


Figure 12.9 Two variograms and their corresponding covariance functions that differ only in their nugget effect.

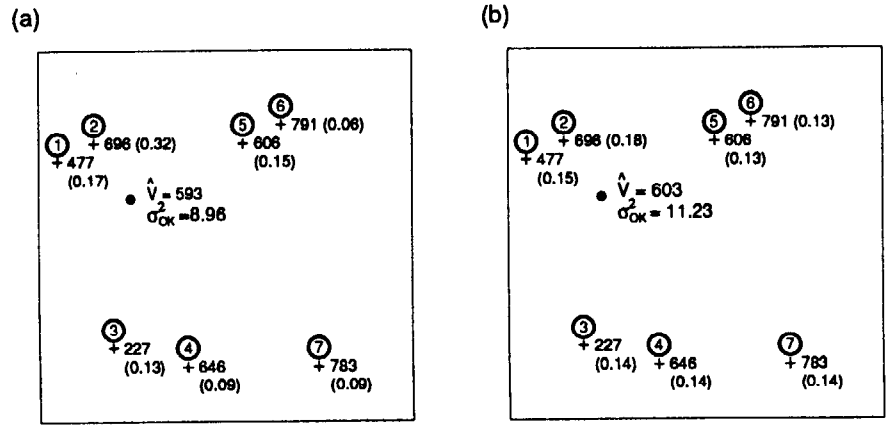


Figure 12.10 Ordinary kriging results using two different covariance functions that differ only in their nugget effect. (a) shows the kriging weights for no nugget effect while (b) shows the weights for a relative nugget of one-half. The two covariance functions are given in Figure 12.9.

ogram model is a pure nugget effect:

$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ C_0 & \text{if } h > 0 \end{cases}$$

there is no redundancy between any of the samples and, in terms of statistical distance, none of the samples is any closer to the point being estimated than any other. The result is that for ordinary kriging with a pure nugget effect model of spatial continuity, all weights are equal

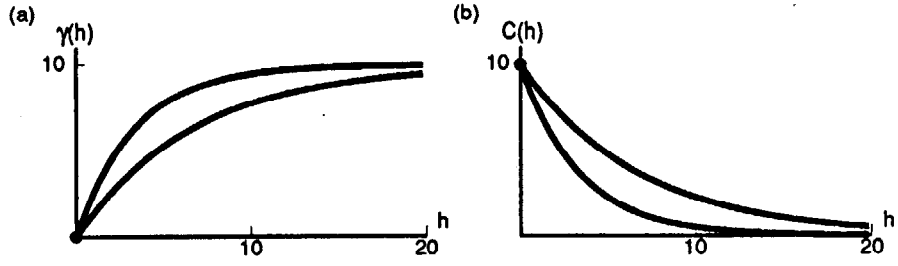


Figure 12.11 Two variograms and their corresponding covariance functions that differ only in their range.

to $\frac{1}{n}$. A pure nugget effect model entails a complete lack of spatial correlation; the data value at any particular location bears no similarity even to very nearby data values. While this produces a simple solution to the ordinary kriging system, it is not a desirable situation in terms of the ordinary kriging variance. The only use of additional samples is to reduce the uncertainty about the unknown mean of the random variables in our random function model. The ordinary kriging variance for spatially uncorrelated phenomena is the variance we have assumed for our random variables, plus the variance of the unknown mean:

$$\sigma^2_{OK} = \underbrace{\tilde{\sigma}^2}_{\text{Variance of random variables}} + \underbrace{\frac{\tilde{\sigma}^2}{n}}_{\text{Variance of unknown mean}}$$

The Effect of the Range. Figure 12.11 show two variogram models that differ only in their ranges. $\gamma_2(h)$ has a range twice that of $\gamma_1(h)$:

$$\gamma_1(h) = 10(1 - e^{-.3|h|})$$
$$\gamma_2(h) = 10(1 - e^{-.15|h|}) = \gamma_1(\frac{1}{2}h)$$

The change of the range has a relatively minor effect on the ordinary kriging weights; none of the weights changes by more than 0.06. Even so, these relatively small adjustments in the weights do cause a noticeable change in the estimate. The ordinary kriging variance is lower since the effect of doubling the range in $\gamma_2(h)$ is to make the

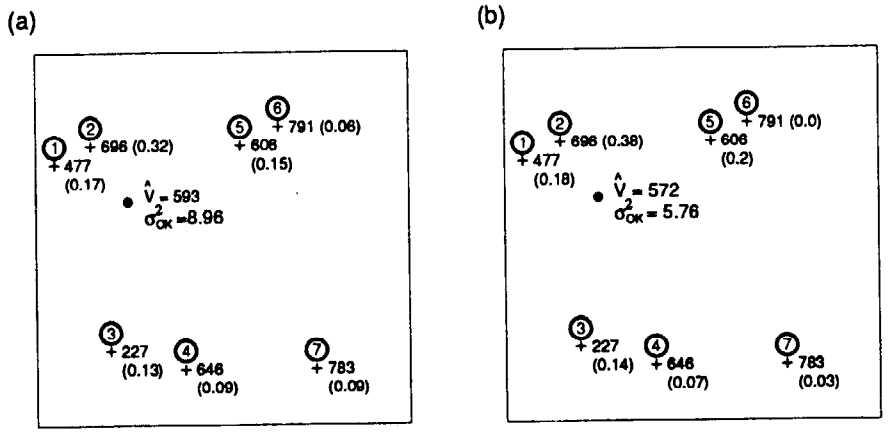


Figure 12.12 Ordinary kriging results using two different covariance functions that differ only in their range. (a) shows the kriging weights for a range of 10 while (b) shows the weights for a range of 20. The two covariance functions are given in Figure 12.11.

samples appear to be twice as close, in terms of statistical distance, as they originally were with $\gamma_1(h)$. If the range becomes very small, then all samples appear to be equally far away from the point being estimated and from each other, with the result being similar to that of a pure nugget effect model: the weights all become $\frac{1}{n}$ and the estimation procedure becomes a simple average of the available sample data.

The Effect of Anisotropy. In all of the examples of ordinary kriging we have looked at so far, we have used only the magnitude of the vector **h**, thus ignoring the influence of direction. All of our variogram models have been isotropic; a contour map of the variogram or covariance surface, such as the one shown in Figure 12.13, would show circular contour lines. In many data sets the data values are more continuous along certain directions than along others. The covariance surface contoured in Figure 12.14 rises more rapidly in the N45°E direction than in the N45°W direction. The directional covariance functions and variograms along these axes of maximum and minimum continuity are shown in Figure 12.15; the anisotropy ratio is 2:1. In Chapter 16 we will show how two variogram models for perpendicular directions can be combined into a single function that

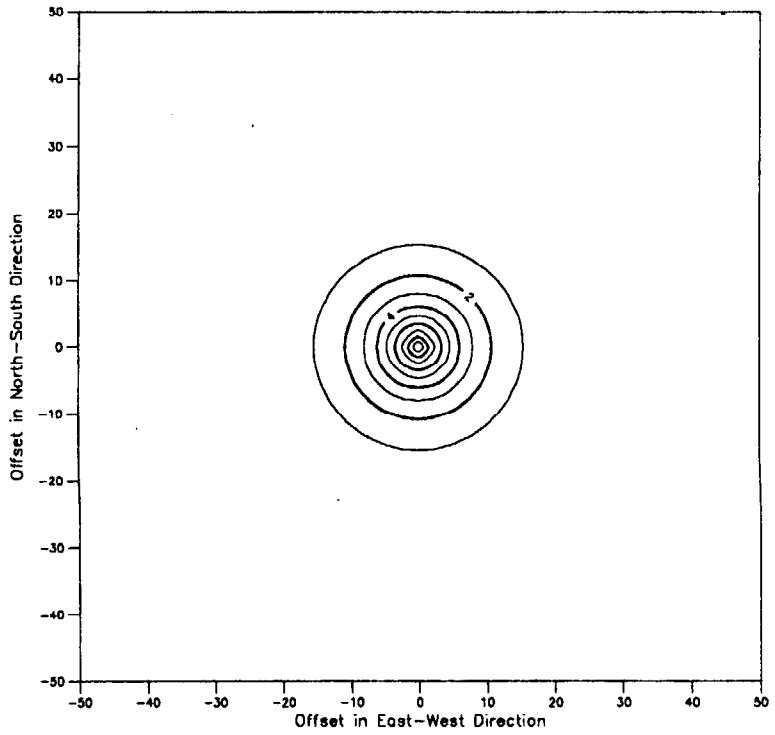


Figure 12.13 A contour map of an isotropic covariance surface. The contour map of the corresponding variogram surface appears identical except that the contours would show a hole rather than a peak.

describes the spatial continuity for all distances and directions. For the moment, however, let us concentrate on the effect of anisotropy and not worry about the precise mathematical description of the variogram model shown in Figure 12.14.

The results of using the isotropic variogram model shown in Figure 12.13 and the anisotropic variogram model shown in Figure 12.14 are shown in Figure 12.16. With the anisotropic model, more of the weight is given to samples 1 and 2, which lie in the direction of maximum continuity and considerably less is given to sample 5, which lies in the direction of minimum continuity.

If we rotate the axes of the anisotropy, as shown in Figure 12.17, so that the N45°E direction is now the major direction of continuity,

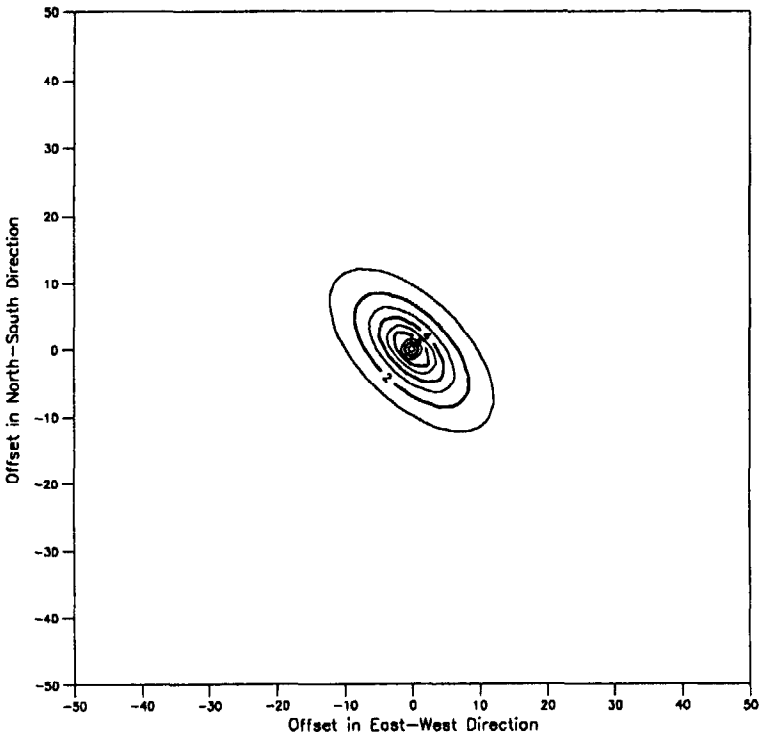


Figure 12.14 A contour map of an anisotropic covariance surface whose major axis of continuity is oriented along N45°W. The contour map of the corresponding variogram surface appears identical except that the contours would show a hole rather than a peak.

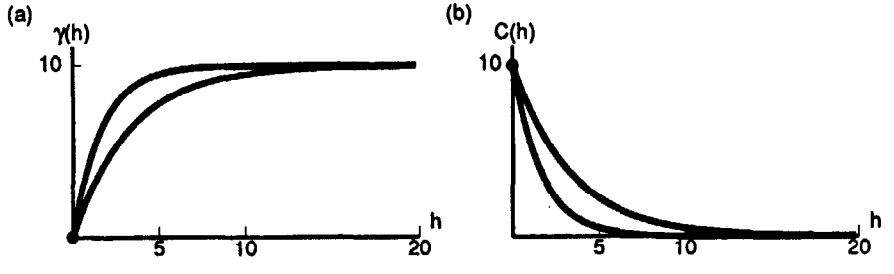


Figure 12.15 Directional variograms of the surface shown in Figure 12.14 along the axes of maximum and minimum continuity.

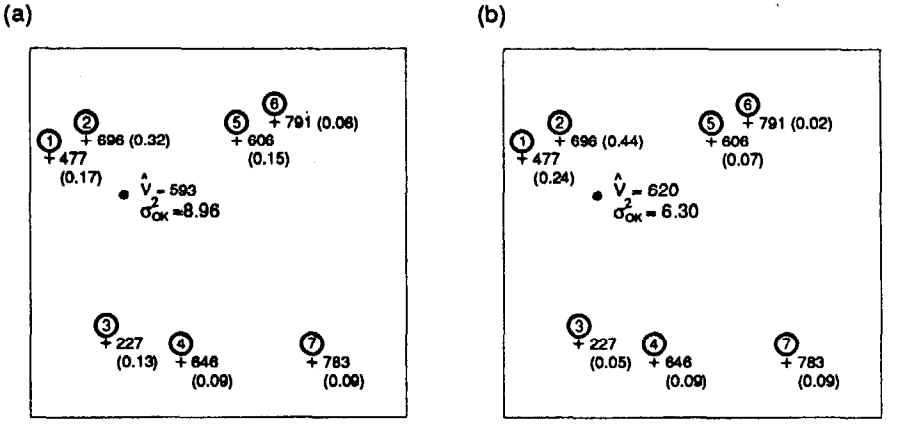


Figure 12.16 Ordinary kriging results using two different covariance functions that differ only in their anisotropy. (a) shows the kriging weights for the isotropic model shown in Figure 12.13 and (b) for the anisotropic model shown in Figure 12.14.

the ordinary kriging weights will reflect this new choice. Figure 12.18 shows the result of using the variogram model shown in Figure 12.17. The choice of a pattern of spatial continuity, which identifies N45°E as the direction of maximum continuity, causes sample 5 to receive the largest weight despite the fact that sample 2 is much closer in terms of geometric distance to the point being estimated.

Finally, the anisotropy ratio plays as important a role as the direction of anisotropy. Figure 12.19 shows a variogram surface that, like the one shown in Figure 12.14, has N45°W as the direction of maximum continuity. The variogram model shown in Figure 12.19, however, has a much higher anisotropy ratio of 10:1. The ordinary kriging weights, which are calculated using this model, are shown in Figure 12.20. Nearly all of the weight is now given to three samples: samples 1 and 2 to the northwest of the point we are estimating and sample 7 far to the southeast. In terms of geometric distance, sample 7 is the farthest from the point we are estimating, but since it lies in the direction of maximum continuity and since we have chosen a high anisotropy ratio, it becomes one of the most influential samples.

The possibility of choosing strongly anisotropic patterns of spatial continuity for our random function model gives us a powerful ability

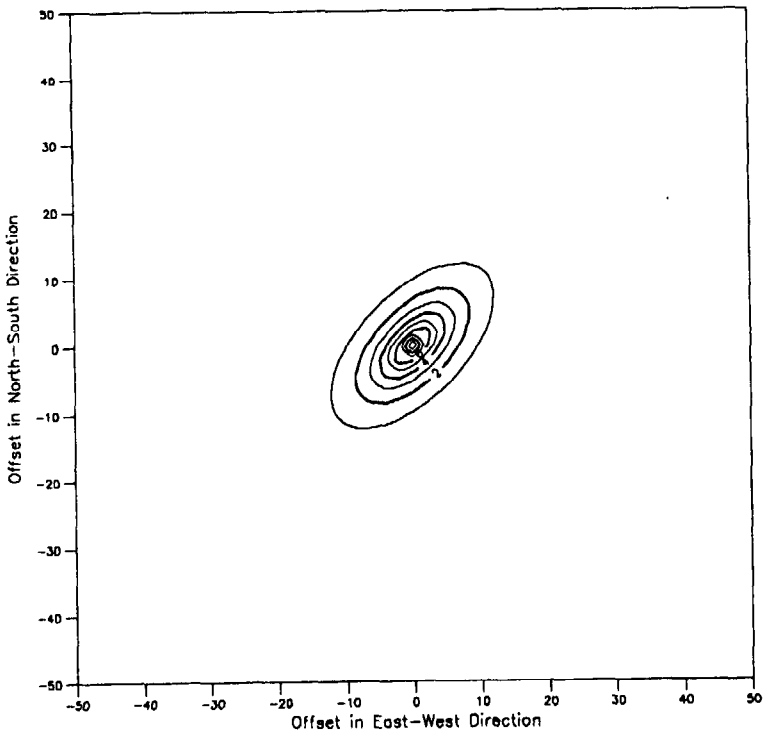


Figure 12.17 A contour map of an anisotropic covariance surface. The direction of maximum continuity is N45°E. The contour map of the corresponding variogram surface appears identical except that the contours would show a hole rather than a peak.

to customize the estimation procedure. Qualitative information such as a geologic interpretation for an ore deposit, a knowledge of the prevailing wind direction in a study of airborne pollution, or tracer tests that reveal preferred flow directions in a reservoir, can be incorporated through the anisotropy of the variogram model. In many data sets, the direction of maximum continuity is not the same throughout the area of interest; there may be considerable local fluctuations in the direction and the degree of the anisotropy. In such situations, the sample variograms may appear isotropic only because we are unable to sort out the undulating character of the anisotropy. If qualitative information offers a way to identify the direction and the degree of the anisotropy,

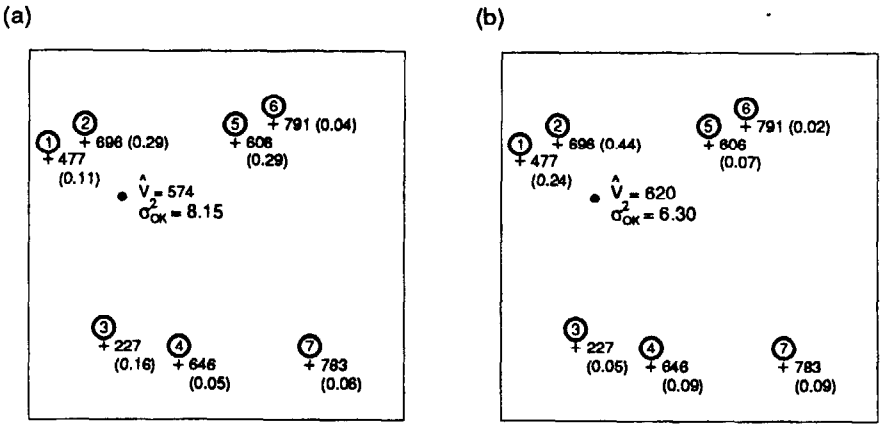


Figure 12.18 Ordinary kriging results using two different covariance functions that differ only in their anisotropy. (a) shows the kriging weights for the anisotropic model shown in Figure 12.17 and (b) for the anisotropic model shown in Figure 12.14.

then the estimation procedure will benefit greatly from a decision to base the choice of the spatial continuity model on qualitative evidence rather than on the quantitative evidence of the sample variogram.

Comparison of Ordinary Kriging to Other Estimation Methods

To compare ordinary kriging to the other estimation methods we looked at in the last chapter, we have repeated the exercise of estimating the V value at points located on a regular square grid. As before, the origin of this grid is 5E,5N and the spacing between points is 10 m in both the north-south and the east-west directions.

For the spatial continuity of our random function model, we have chosen the traditional method of fitting a function to our sample variogram. The directional sample variograms calculated in Chapter 7 as part of the analysis of the spatial continuity in the sample data set are shown again in Figure 12.22. These sample variograms show an anisotropy that is preserved in the complete variogram model shown in Figure 12.21. The variogram models for the directions of maximum and minimum continuity shown in Figure 12.22 are: Direction of

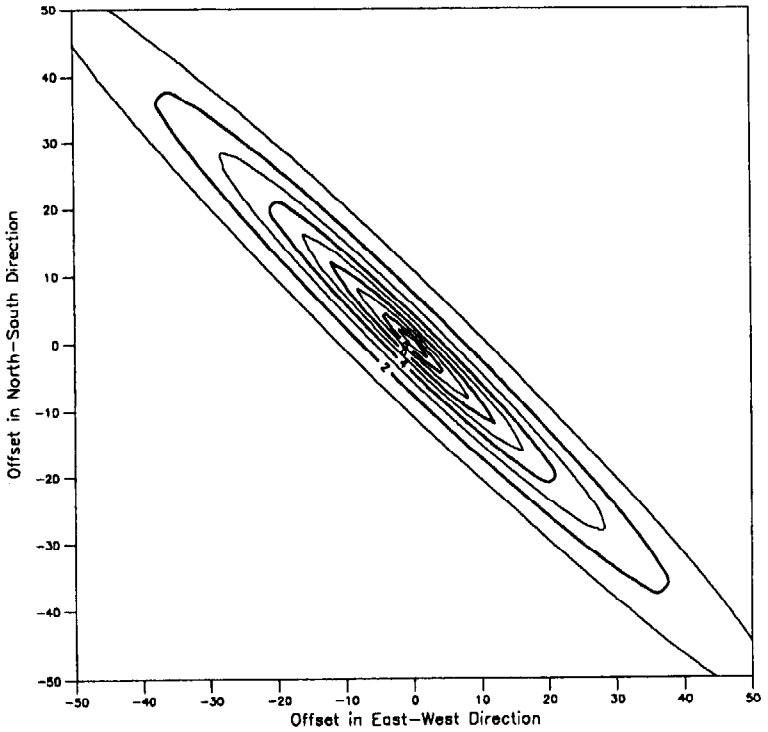


Figure 12.19 A contour map of a severe anisotropic covariance surface. The direction of maximum continuity is N45°W and the anisotropy ratio is 10:1. The contour map of the corresponding variogram surface appears identical except that the contours would show a hole rather than a peak.

maximum continuity (N14°W):

$$\gamma_{max}(h) = \begin{cases} 0 & \text{if } h = 0 \\ 22{,}000 + 40{,}000\text{Sph}_{30}(h) + 45{,}000\text{Sph}_{150}(h) & \text{if } h > 0 \end{cases} \quad (12.26)$$

Direction of minimum continuity (N76°E):

$$\gamma_{min}(h) = \begin{cases} 0 & \text{if } h = 0 \\ 22{,}000 + 40{,}000\text{Sph}_{25}(h) + 45{,}000\text{Sph}_{50}(h) & \text{if } h > 0 \end{cases} \quad (12.27)$$

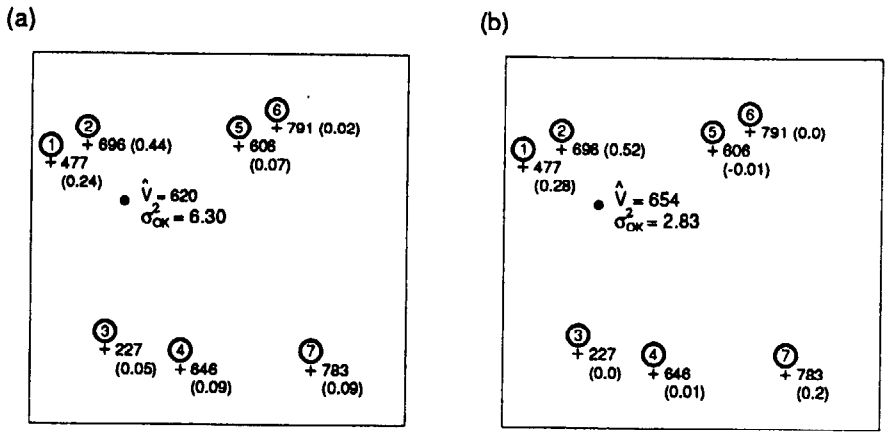


Figure 12.20 Ordinary kriging results using two different covariance functions that differ only in their anisotropy ratio. (a) shows the kriging weights for the anisotropic model shown in Figure 12.14 and (b) for the anisotropic model shown in Figure 12.19.

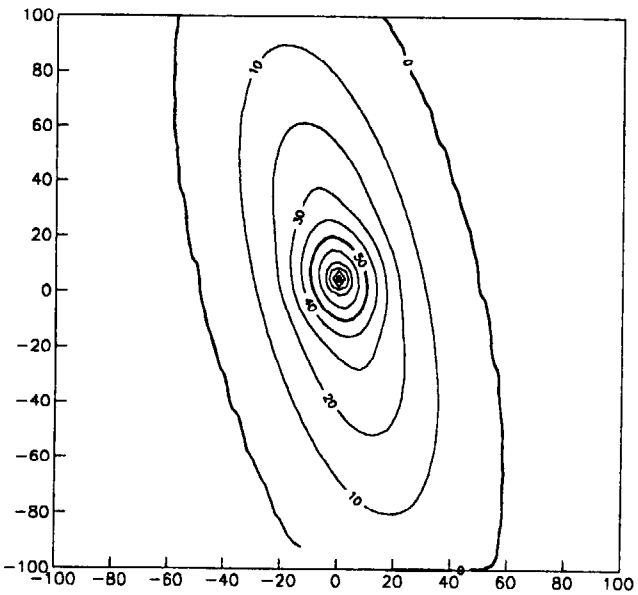
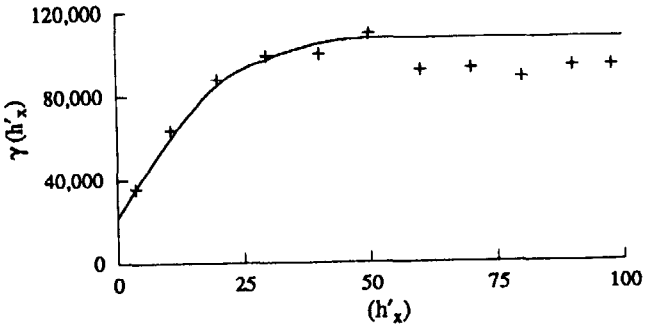
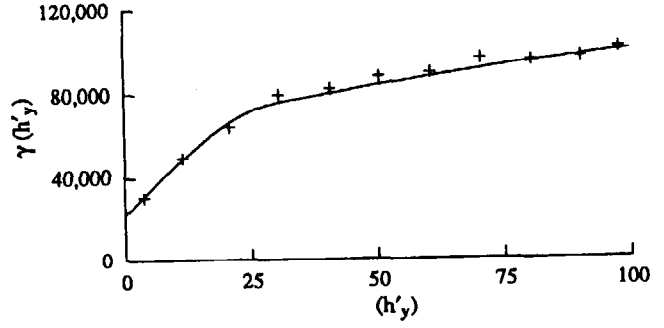


Figure 12.21 A contour map of the anisotropic model fitted to the sample variogram of V.

(a) $\gamma_V(h'_x) = 22{,}000 + 40{,}000\ \mathrm{Sph}_{25}(h'_x) + 45{,}000\ \mathrm{Sph}_{50}(h'_x)$



(b) $\gamma_V(h'_y) = 22{,}000 + 40{,}000\ \mathrm{Sph}_{30}(h'_y) + 45{,}000\ \mathrm{Sph}_{150}(h'_y)$



(c)

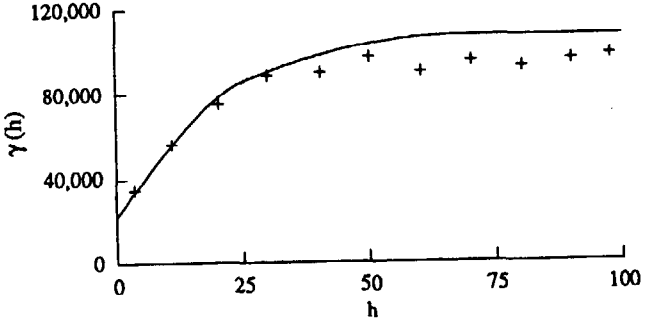


Figure 12.22 The directional sample variograms of V with their fitted model.

Table 12.3 Univariate statistics of ordinary kriged estimates compared to the true values.

	True	OK
n	780	780
m	283	283
σ	251	202
min	0	0
Q_1	70	127
M	219	251
Q_3	446	392
max	1323	1227

Table 12.4 Summary statistics for the error distributions from OK and each of the previous four point estimation methods.

	Polygonal	Triangulation	Local Sample Mean	Inverse Distance Squared	OK
n	780	672	780	780	780
m	1.1	−16.8	51.1	26.8	-0.2
σ	175.2	153.7	187.5	156.0	144.2
min	−651	−546	−618	−630	−472
Q_1	−97	−58	−66	−65	−86.8
M	0	3.9	48.9	24.4	9.1
Q_3	95	105	175	121	79.9
max	595	524	537	473	657
MAE	128	111	154	121	108
MSE	30,642	23,885	37,741	25,000	20,769
ρ_{tv}	0.69	0.80	0.67	0.78	0.82

$\mathrm{Sph}_a(h)$ is a positive definite function that is commonly used in practice for fitting sample variograms and is defined as follows:

$$\mathrm{Sph}_a(h) = \begin{cases} 0 & \text{if } h = 0 \\ 1.5\frac{h}{a} - 0.5(\frac{h}{a})^3 & \text{if } 0 < h < a \\ 1 & \text{if } h \geq a \end{cases} \qquad (12.28)$$

The details of how these two directional variogram models are combined to make the complete model shown in Figure 12.21 will be deferred to Chapter 16.

Like the case studies presented in the last chapter, the ordinary

Table 12.5 Comparison of OK to the previous four estimation methods for the 50 least clustered sample data configurations that contained at least 10 samples.

		Polygonal	Triangulation	Local Sample Mean	Inverse Distance Squared	OK
Error Distribution	n	50	50	50	50	50
	m	−22.6	−49.5	−56.3	−30.1	-31.7
	σ	187	142	212	152	126
	IQR	227	201	327	221	203
	MAE	151	127	178	123	109
	MSE	34,750	22,140	47,194	23,562	23,562
Correlation	ρ	0.845	0.907	0.834	0.921	0.930

kriging exercise used all samples that fell within 25 m of the point being estimated.

Table 12.3 presents a comparison of the univariate distributions of the estimates and the true values. The results of the estimation studies in the previous chapter have been repeated in Tables 12.4, 12.5, and 12.6 so that we can compare the ordinary kriging results with the results of those earlier methods. The mean of the ordinary kriged estimates is the same as the true mean; this nearly exact match between the estimated and true means is quite fortuitous and one should not expect such a close agreement in all situations. The ordinary kriging estimates are less variable than the true values; their standard deviation and interquartile range are both lower than those of the true values, and the maximum estimated value is lower than the maximum true value. The degree of smoothing of the ordinary kriging estimates is more severe than that of the polygonal and triangulation estimates, and similar in magnitude to the degree of smoothing of the inverse distance squared estimates. The q-q plot of the estimated and true distributions shown in Figure 12.23 further reveals the effects of the smoothing. There are fewer extremely low estimates than there are extremely low true value; very high values are also more numerous in the true values than in the estimates.

Table 12.4, which summarizes the error distributions for the various methods, shows that the standard deviation of the errors is lower for the ordinary kriging estimates than for any of the other techniques.

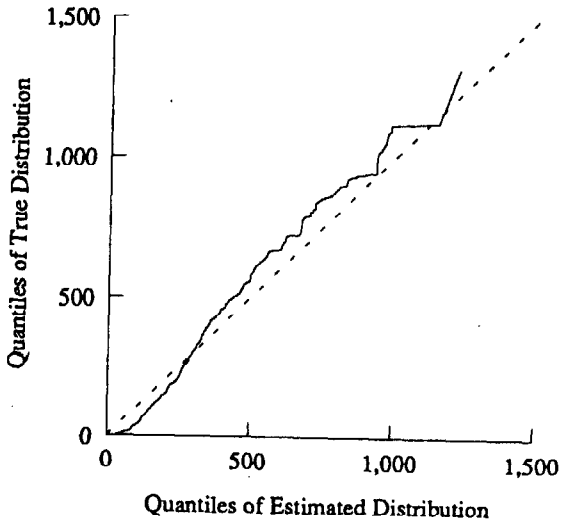


Figure 12.23 A q-q plot of the distribution of OK estimates and the distribution of their true values.

Table 12.6 Comparison of OK to the previous four estimation methods for the 50 most clustered sample data configurations that contained at least 10 samples.

		Polygonal	Triangulation	Local Sample Mean	Inverse Distance Squared	OK
Error Distribution	n	50	43	50	50	50
	m	17.3	−60.5	103.7	58.0	-29.6
	σ	182	138	221	161	125
	IQR	213	202	330	248	177
	MAE	138	106	198	142	102
	MSE	32,586	22,416	58,415	28,779	16,099
Correlation	ρ	0.674	0.794	0.260	0.774	0.853

Though the specific aim of ordinary kriging was only to minimize the error variance (or, equivalently, the error standard deviation), we can see that the ordinary kriging estimates are also very good according to many other criteria. They have the lowest mean absolute error and also the lowest mean squared error. The distribution of the errors

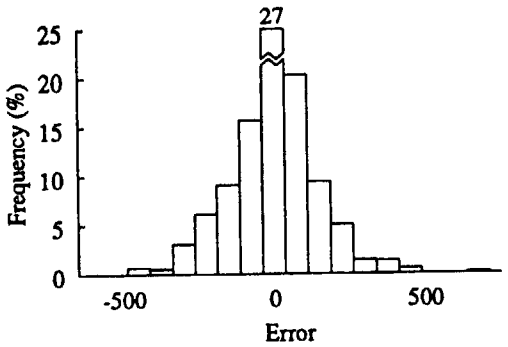


Figure 12.24 A histogram of the 780 OK estimation errors.

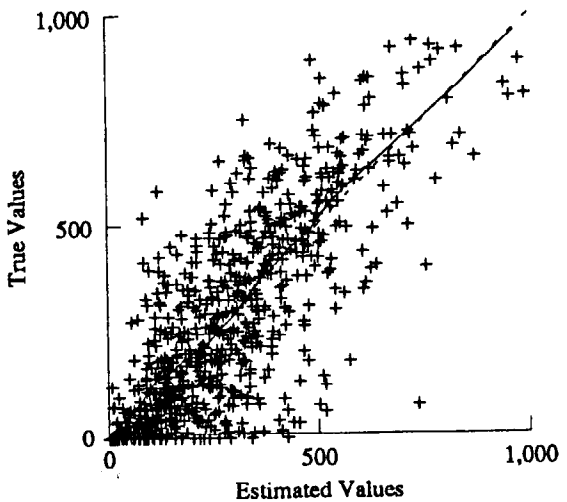


Figure 12.25 A scatterplot of the 780 OK estimates and their true values. The conditional expectation line is also included in the figure.

shown by the histogram in Figure 12.24 is fairly symmetric and does not have long tails. The correlation coefficient between the true values and the estimates is higher for ordinary kriging than for any of the other method we have tried.

The scatterplot of the 780 estimated and true values shown in Figure 12.25 along with the conditional expectation curve shows that there is very little conditional bias for the entire range of estimated values.

In the previous chapter we saw that all of the estimation methods suffered from the effects of clustering. We saw that our estimates were generally poorer for the 50 most clustered data configurations than for the 50 least clustered configurations. In Tables 12.5 and 12.6 this comparison is extended to include the ordinary kriging exercise. Though the ordinary kriging estimates are definitely adversely affected by clustering, they do not deteriorate as much as the estimates for other techniques. For example, the triangulation estimates for the least clustered configurations correlate almost as well with the true values as do the ordinary kriging estimates; both have a correlation coefficient of about 0.9. For the most clustered configurations, however, the correlation coefficient for the triangulation estimates drops to 0.78 while the correlation coefficient for the ordinary kriging estimates drops slightly to 0.85.

By trying to account for the possibility of redundancy in the sample data set through the covariances in the **C** matrix, ordinary kriging handles the adverse effects of clustering much better than other methods. Though it is certainly a more time-consuming procedure, it does generally produce better estimates.

The success of ordinary kriging is due to its use of a customized statistical distance rather than a geometric distance and to its attempt to decluster the available sample data. Its use of a spatial continuity model that describes the statistical distance between points gives it considerable flexibility and an important ability to customize the estimation procedure to qualitative information.

Notes

[1] A good presentation of Lagrange multipliers and constrained maximization-minimization problems with theorems and proofs is given in:
Edwards, C. and Penney, D. , *Calculus and Analytical Geometry.* Englewood Cliffs, N.J.: Prentice-Hall, 1982.

[2] The covariance is used by many algorithms in solving the ordinary kriging matrices for the sake of convenience. By using the covariance, the largest elements of the covariance matrix will be located on the diagonal. Thus for a system solver based on Gaussian elim-

ination, for example, there is no need for a pivot search and the exchange of rows.

[3] The name *sill* comes from the variogram, which typically reaches a plateau; although it makes less sense when discussing a covariance function, we will still refer to $C_0 + C_1$ as the sill.

[4] Each of the following tests is a necessary and sufficient condition for a real symmetric matrix $\mathbf{C}$ to be positive definite.

$\mathbf{w}^T\mathbf{C}\mathbf{w} > 0$ for all nonzero vectors $\mathbf{w}$.

All the eigenvalues of $\mathbf{C}$ are greater than 0.

All the submatrices of $\mathbf{C}$ have positive determinants.

All the pivots (without row exchanges) are greater than 0.

[5] The fact that the variogram can be rescaled by any constant without changing the estimate enables one to use the relative variogram without fear of altering the estimate. In the case where local relative variograms differ one from another by only a rescaling factor, only the kriging variance will be affected. Each one of the local relative variograms will provide an identical kriging estimate.

Further Reading

Brooker, P. , "Kriging," *Engineering and Mining Journal*, vol. 180, no. 9, pp. 148–153, 1979.

Buxton, B. , *Coal Reserve Assessment: A Geostatistical Case Study.* Master's thesis, Stanford University, 1982.

David, M. , *Geostatistical Ore Reserve Estimation.* Amsterdam: Elsevier, 1977.

Journel, A. G. and Huijbregts, C. J. , *Mining Geostatistics.* London: Academic Press, 1978.

Royle, A. , "Why geostatistics?," *Engineering and Mining Journal*, vol. 180, no. 5, pp. 92–102, 1979.

13

BLOCK KRIGING

In the last two chapters, we have concentrated on the goal of point estimation. Often, however, we require a block estimate or, more precisely, an estimate of the average value of a variable within a prescribed local area [1].

One method for obtaining such an estimate is to discretize the local area into many points and then average the individual point estimates all together to get the average over the area. Though conceptually simple, this procedure may be computationally expensive. For example, in mining applications hundreds of thousands of block estimates may be required. If each block is discretized into 100 points, and each point estimate is made by ordinary kriging, there will be several million kriging systems to solve. In this chapter we will see how the number of computations can be significantly reduced by constructing and solving only one kriging system for each block estimate.

We begin this chapter with the development of the block kriging system and follow with an example demonstrating the equivalence between block kriging and the averaging of kriged point estimates within the block. We then use the sample V data to explore how the discretization of the block affects the estimates. The chapter concludes with a case study that compares block estimates calculated by kriging and by inverse distance squared to the true values.

The Block Kriging System

The block kriging system is similar to the point kriging system given in the previous chapter:

$$\underset{\mathbf{C}}{\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{(n+1)\times(n+1)} \cdot \underset{\mathbf{w}}{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}_{(n+1)\times 1} = \underset{\mathbf{D}}{\begin{bmatrix} \tilde{C}_{10} \\ \vdots \\ \tilde{C}_{n0} \\ 1 \end{bmatrix}}_{(n+1)\times 1} \tag{13.1}$$

The matrix **C** consists of the covariance values $\tilde{C}_{ij}$ between the random variables V_i and V_j at the sample locations. The vector **D** consists of the covariance values $\tilde{C}_{i0}$ between the random variables V_i at the sample locations and the random variable V_0 at the location where we need an estimate. The vector **w** consists of the kriging weights $w_1, \ldots, w_n$ and the Lagrange parameter μ. It should be remembered that the random variables V_i, V_j, and V_0 are models of the phenomenon under study and the tilde above the C reminds us that these are parameters of a random function model.

However, suppose we wish to estimate the mean value of some phenomenon over a local area, rather than at a point location. Within the framework of the random function model that we used earlier, the mean value of a random function over a local area is simply the average (a linear combination) of all the point random variables contained within the local area. Recall from Chapter 9 that a linear combination of random variables is also a random variable, thus the mean value over a local area can be described as follows:

$$V_A = \frac{1}{|A|} \sum_{j|j\in A} V_j \tag{13.2}$$

where V_A is a random variable corresponding to the mean value over an area A and V_j are random variables corresponding to point values within A.

If we examine the ordinary point kriging system given in Equation 13.1 with a view towards modifying it for block estimation, we will soon see that the location of the point or block that we are estimating has absolutely nothing to do with the construction of the

has absolutely nothing to do with the construction of the covariance matrix **C**. This matrix is independent of the location at which the estimate is required and so we can correctly conclude that the matrix **C** does not require any modifications for block kriging. However, the covariance vector **D** consists of covariance values between the random variables at the sample locations and the random variable at the location that we are trying to estimate. For point estimation, these covariances are point-to-point covariances. By analogy, for block estimation, the covariance values required for the covariance vector **D** are the point-to-block covariances. In fact, by making this single alteration, we can convert the ordinary point kriging system to a ordinary block kriging system.

The point-to-block covariances that are required for block kriging can be developed as follows:

$$\begin{aligned} \tilde{C}_{iA} &= Cov\{V_A V_i\} \\ &= E\{V_A V_i\} - E\{V_A\}E\{V_i\} \\ &= E\{\frac{1}{|A|}\sum_{j|j\in A} V_j V_i\} - E\{\frac{1}{|A|}\sum_{j|j\in A} V_j\}E\{V_i\} \\ &= \frac{1}{|A|}\sum_{j|j\in A} E\{V_j V_i\} - \frac{1}{|A|}\sum_{j|j\in A} E\{V_j\}E\{V_i\} \\ &= \frac{1}{|A|}\sum_{j|j\in A} [E\{V_j V_i\} - E\{V_j\}E\{V_i\}] \\ &= \frac{1}{|A|}\sum_{j|j\in A} Cov\{V_j V_i\} \end{aligned}$$

The covariance between the random variable at the ith sample location and the random variable V_A representing the average value of the phenomenon over the area A is the same as the average of the point-to-point covariances between V_i and the random variables at all the points within A. The block kriging system can therefore be written

as

$$\underset{}{\mathbf{C}} \cdot \mathbf{w} = \mathbf{D}$$

$$\underbrace{\begin{bmatrix} \tilde{C}_{11} & \cdots & \tilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n1} & \cdots & \tilde{C}_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{(n+1)\times(n+1)} \cdot \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}_{(n+1)\times 1} = \underbrace{\begin{bmatrix} \bar{\tilde{C}}_{1A} \\ \vdots \\ \bar{\tilde{C}}_{nA} \\ 1 \end{bmatrix}}_{(n+1)\times 1} \tag{13.3}$$

The bar above the covariances on the right-hand side reminds us that the covariance is no longer a point-to-point covariance, but the average covariance between a particular sample location and all of the points within A:

$$\bar{\tilde{C}}_{iA} = \frac{1}{|A|} \sum_{j|j\in A} \tilde{C}_{ij} \tag{13.4}$$

Later in this chapter we will determine the number of discretizing points that are needed within A to give us an adequate approximation of $\bar{\tilde{C}}_{iA}$.

The block kriging variance is given by:

$$\tilde{\sigma}^2_{OK} = \bar{\tilde{C}}_{AA} - (\sum_{i=1}^{n} w_i \bar{\tilde{C}}_{iA} + \mu) \tag{13.5}$$

The value $\bar{\tilde{C}}_{AA}$ in this equation is the average covariance between pairs of locations within A:

$$\bar{\tilde{C}}_{AA} = \frac{1}{|A|^2} \sum_{i|i\in A} \sum_{j|j\in A} \tilde{C}_{ij} \tag{13.6}$$

In practice, this average block-to-block covariance is approximated by discretizing the area A into several points. It is important to use the same discretization for the calculations of the point-to-block covariances in $\mathbf{D}$ and for the calculation of the block-to-block covariance in Equation 13.5 [2].

The advantage of using the block kriging system given in Equation 13.3 is that it produces an estimate of the block average with the solution of only one kriging system. The disadvantage is that the calculation of the average covariances involves slightly more computation than the calculation of the point-to-point covariances in the point kriging system. However, the computational savings in requiring only one set of simultaneous linear equations far outweighs the additional cost of calculating average point-to-block covariances.

The convenience of estimating block values directly is a feature of the ordinary kriging system that is not shared by other estimation methods. Though some methods can be adapted in a similar manner, the results are not consistent. For example, inverse distance methods can be adapted so that the weight is proportional to the average distance between a nearby sample and the block. Unfortunately, the estimate calculated by this procedure is not the same as the one calculated by averaging individual point estimates within the block.

In the next section, we will give an example that demonstrates that the block kriging system given in Equation 13.3 produces an estimate identical to that obtained by averaging the point estimates produce by Equation 13.1.

Block Estimates Versus the Averaging of Point Estimates

An example of block kriging is shown in Figure 13.1a. Five sampled locations are marked by plus signs; the sample value is given immediately to the right of each sign with the corresponding block kriging weight enclosed in parentheses. The variogram model used to build the kriging matrices was the one used earlier in the case study in Chapter 12 and developed in detail in Chapter 16. The block whose average value we wish to estimate is shown as a shaded square. For the purposes of calculating the various average covariances, this block has been discretized by four points shown as black dots in Figure 13.1a. The block estimate is $\hat{V}_A = 337$ ppm. Figures 13.1b through 13.1e show the kriging weights and the point estimates for each one of the four point locations within the shaded square. As shown in Table 13.1, the average of the four point estimates is the same as the direct block estimate and the average of the point kriging weights for a particular sample is the same as the block kriging weight for that sample.

Varying the Grid of Point Locations Within a Block

When using the block kriging approach, one has to decide how to discretize the local area or block being estimated. The grid of discretizing points should always be regular; the spacing between points, however,

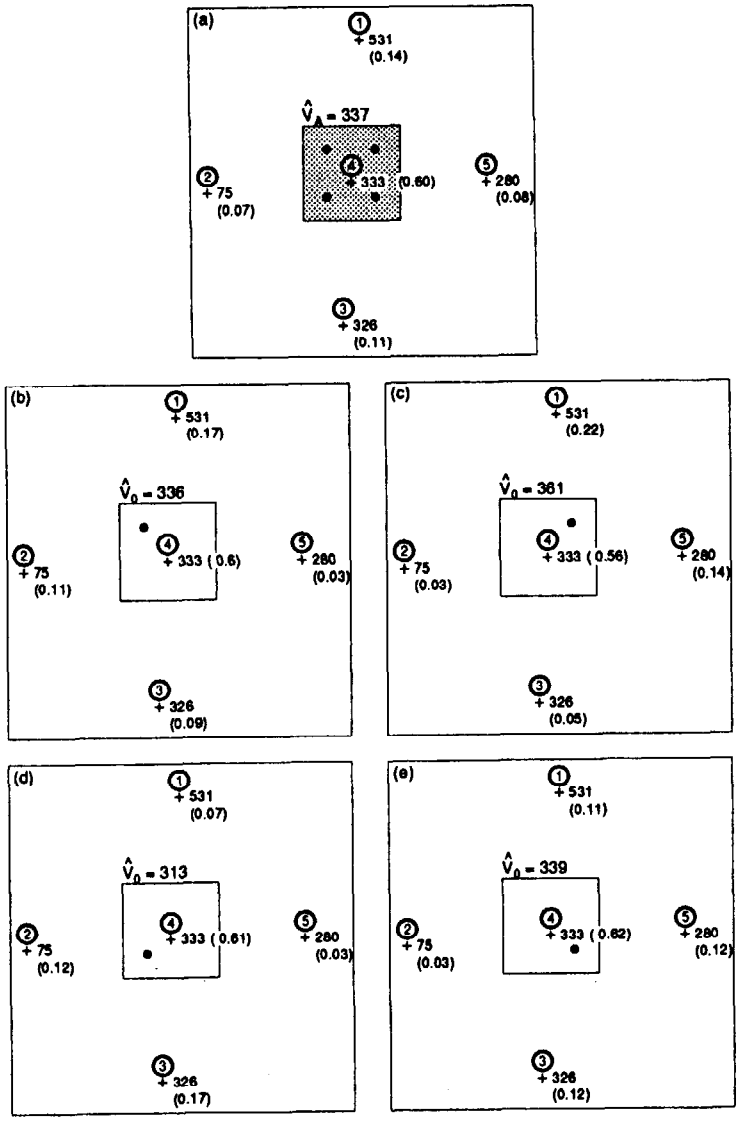


Figure 13.1 In (a) the shaded block is estimated directly using block kriging, with the block being approximated by the four points shown as dots. The nearby sample locations are marked with a plus sign. The value posted immediately to the right of the plus sign is the sample value and the value enclosed in parentheses is the corresponding kriging weight. Figures (b) to (e) show the results for the point kriging of each one of the four points within the shaded square. The average of the point estimates in (b) to (e) is identical to the block estimate $\hat{V}_A$=337 in (a).

Table 13.1 The point estimates and weights shown in Figure 13.1(b) - (e) are tabulated and averaged, demonstrating their equivalence to the direct block estimate shown in Figure 13.1(a).

		Kriging weights for samples				
Figure	Estimate	1	2	3	4	5
13.1(b)	336	0.17	0.11	0.09	0.60	0.03
13.1(c)	361	0.22	0.03	0.05	0.56	0.14
13.1(d)	313	0.07	0.12	0.17	0.61	0.03
13.1(e)	339	0.11	0.03	0.12	0.62	0.12
Average	337	0.14	0.07	0.11	0.60	0.08
13.1(a)	337	0.14	0.07	0.11	0.60	0.08

may be larger in one direction than the other if the spatial continuity is anisotropic. An example of such an anisotropic grid is given in Figure 13.2. The shaded block is approximated by six points located on a regular 2 x 3 grid. The closer spacing of the points in the north-south direction reflects a belief that there is less continuity in this direction than in the east-west direction. Despite the differences in the east-west and north-south spacing, the regularity of the grid ensures that each discretizing point accounts for the same area, as shown by the dashed lines.

If one chooses to use fewer discretizing points, less computer memory is required and the computations are faster. This computational efficiency must be weighed against the desire for accuracy, which calls for as many points as possible.

Table 13.2 provides several examples of the effect of the number of discretizing points on the final estimate. The table shows estimates of the average V value within 10 x 10 m^2 blocks using ordinary block kriging to calculate weights for the nearby samples. In these examples, the search strategy included all samples within a 25 m radius of the center of the block; the variogram model used to calculate the various covariances is the same one used in earlier studies and given in Equations 12.25–12.27. The only parameter changed from one kriging to the next is the number of points used to discretize a 10 x 10 m^2 block.

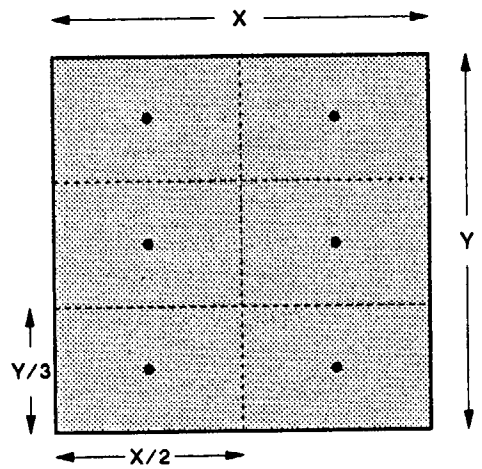


Figure 13.2 An example showing the design of a regular 2 x 3 grid of point locations within a block. The points are located within the square so that each discretizing point accounts for the same area, as shown by the dashed lines.

The entries in Table 13.2 show significant differences in the estimates using grids containing fewer than 16 discretizing points; with more than 16 points, however, the estimates are all very similar. For a two-dimensional block, sufficient accuracy can usually be obtained with a 4 x 4 grid containing 16 points. For a three dimensional block, more points are usually required; a 4 x 4 x 4 grid containing 64 points is usually sufficient.

A Case Study

Next we present a case study that compares ordinary block kriging estimates of the average value of V within 10 x 10 m^2 blocks to block estimates using the inverse distance squared method.

For this study, we have used a very fine discretization, with each 10 x 10 m^2 block being discretized by 100 points. Ordinary block kriging estimates were calculated using all samples within 25 m. The variogram model used to calculate the various covariances required by Equation 13.3 was the same one that we used for the point kriging case study in Chapter 12.

The inverse distance squared estimates were calculated using the same search strategy as that used for kriging. For each block, 100 inverse distance point estimations were averaged together to obtain the block estimate.

Table 13.2 Examples of estimates of the average value of V within 10 x 10 m^2 blocks using ordinary block kriging and various discretizing grids within the block.

Block center		Grid size within the block				
E	N	1x1	2x2	4x4	6x6	10x10
80	80	584.67	576.41	574.30	573.98	573.81
100	80	408.53	418.29	419.19	419.38	419.47
80	90	538.36	519.89	520.58	520.53	520.47
100	90	460.13	479.73	480.35	480.52	480.61
80	100	497.66	547.87	549.40	550.13	550.51
100	100	530.37	513.32	513.56	513.47	513.42
80	110	781.17	737.04	732.29	731.06	730.41
100	110	591.13	580.73	578.75	578.74	578.72

Summary statistics for the estimates are given in Table 13.3. The average of the kriged block estimates is closer to the true mean than the inverse distance squared estimates. The correlation of the kriged estimates with the true values is also larger. In this particular case, the kriged block estimates have a standard deviation closer to the true one and are therefore not as smoothed as the inverse distance estimates.

The summary statistics of the estimation error are given in Table 13.4. The errors from ordinary block kriging have a mean closer to 0 than those from inverse distance squared. Their spread, as measured by the standard deviation or the interquartile range, is also lower. The MAE and MSE, which provide measures of the combination of bias and spread, both favor the ordinary block kriging estimates.

Figures 13.3 and 13.4 provide a comparison of the two sets of errors using grayscale maps. A plus symbol denotes a positive estimation error while a minus symbol denotes a negative estimation error. The relative magnitude of the error corresponds to the degree of shading indicated by the grey scale at the top of the figure.

The most striking feature of these maps is the consistency of the overestimations in the area of Walker Lake itself and the corresponding

Table 13.3 Summary statistics for the ordinary kriged block estimates, the inverse distance squared block estimates, and the true block values of V.

	True Values	Ordinary Block Kriging	Inverse Distance Squared
n	780	780	780
m	278	284	319
σ	216	194	186
CV	0.77	0.68	0.58
min	0	5	7
Q_1	103	136	178
M	239	258	310
Q_3	404	389	431
max	1,247	1,182	1,112
$\rho_{\hat{V}V}$		0.90	0.87

Table 13.4 Summary statistics for the error distributions from the ordinary block kriging and the inverse distance squared method.

	Ordinary Block Kriging	Inverse Distance Squared
n	780	780
m	6.2	40.1
σ	92.9	105.4
min	-405.6	-336.5
Q_1	-45.5	-21.4
M	17.6	42.6
Q_3	65.3	116.8
max	286.8	392.1
MAE	71.9	90.2
MSE	8,674	12,763

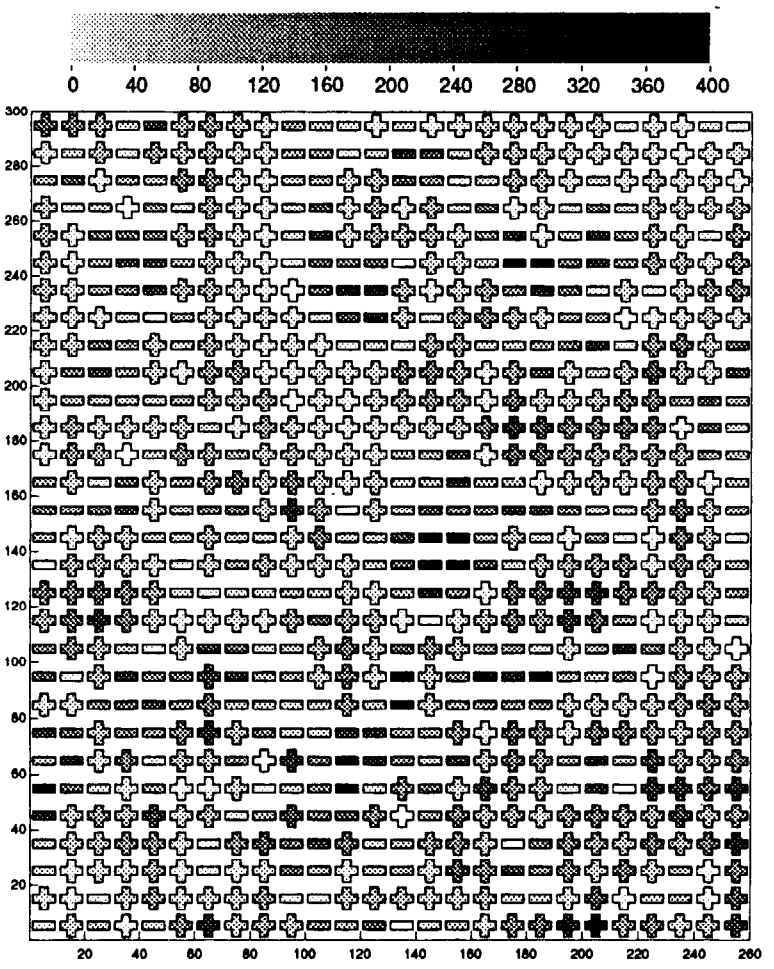


Figure 13.3 A map of the V block kriging errors. Positive errors are plotted as plus symbols while negative errors are plotted as minus symbols. The magnitude of the error is given by the grey scale at the top of the figure.

underestimations in the Wassuk range area. This pattern of the over- and underestimations illustrates the smoothing inherent in both the inverse distance squared method and in kriging; high values tend to be underestimated while low values tend to be overestimated.

Though the pattern of over and underestimation is the same in both maps, it is clear that the inverse distance squared method produces