

a few of the summary statistics of the different polygons along with the correlation coefficient of the true and estimated values to obtain the results in Table 11.1, it appears that incorporating additional nearby samples does improve the estimation. The larger distance between points correlates best with the variance and they also have a large effect on the error.

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 \sim 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.

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

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:

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

As was noted when we first conclusion, all of the components of the random function model have assumed arbitrary values.

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:

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

(E.31)

then the average error of a set of k estimates is

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

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, \dots, 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, $V(x_1), \dots, V(x_n)$, 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:

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

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:

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

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 $E\{V\}$:

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

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:

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

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

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

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

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

$$\begin{aligned}\sigma_R^2 &= \frac{1}{k} \sum_{i=1}^k (r_i - 0)^2 \\ &= \frac{1}{k} \sum_{i=1}^k [\hat{v}_i - v_i]^2\end{aligned}\quad (12.4)$$

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:

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

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

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

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, $\hat{\sigma}_R^2$, 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:

$$Var\left\{\sum_{i=1}^n w_i \cdot V_i\right\} = \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot Cov\{V_i V_j\} \quad (12.6)$$

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

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

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:

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

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, $\hat{\sigma}^2$, then this third term can be expressed as

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

The second term in Equation 12.6 can be written as

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

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

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

Once we have chosen our random function model parameters, specifically the variance $\hat{\sigma}^2$ 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, \dots, 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 $\hat{\sigma}_R^2$ 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 $\hat{\sigma}_R^2$, as expressed in Equation 12.8, as an unconstrained problem, we run into difficulties. Setting the n partial first derivatives of $\hat{\sigma}_R^2$ 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 $\hat{\sigma}_R^2$. This new variable is called μ , the Lagrange parameter, and is introduced into Equation 12.8 in the following way:

$$\hat{\sigma}_R^2 = \hat{\sigma}^2 + \sum_{i=1}^n \sum_{j=1}^n w_i w_j \tilde{C}_{ij} - 2 \underbrace{\sum_{i=1}^n w_i \tilde{C}_{i0}}_0 + 2\mu \left(\sum_{i=1}^n w_i - 1 \right) \quad (12.9)$$

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:

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 :

$$\begin{aligned} \text{unbiasedness cond. } & \frac{\partial(\hat{\sigma}_R^2)}{\partial \mu} = \frac{\partial(2\mu(\sum_{i=1}^n w_i - 1))}{\partial \mu} \quad \text{no effect on } \hat{\sigma}_R^2 \\ & = 2 \sum_{i=1}^n w_i - 2 \end{aligned} \quad (12.13)$$

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 $\hat{\sigma}_R^2$, the solution of those $n+1$ equations will produce the set of weights that minimizes $\hat{\sigma}_R^2$ 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:

$$2 \sum_{j=1}^n w_j \tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu = 0$$

$$\sum_{j=1}^n w_j \tilde{C}_{1j} + \mu = \tilde{C}_{10} \quad (12.10)$$

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

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_i} = 2 \sum_{j=1}^n w_j \tilde{C}_{1j} - 2\tilde{C}_{10} + 2\mu = 0 \Rightarrow \sum_{j=1}^n w_j \tilde{C}_{1j} + \mu = \tilde{C}_{10}$$

$$\vdots$$

$$\frac{\partial(\tilde{\sigma}_R^2)}{\partial w_n} = 2 \sum_{j=1}^n w_j \tilde{C}_{nj} - 2\tilde{C}_{n0} + 2\mu = 0 \Rightarrow \sum_{j=1}^n w_j \tilde{C}_{nj} + \mu = \tilde{C}_{n0}$$

$$\vdots$$

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:

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

$$\sum_{i=1}^n w_i = 1 - E(V_i, V_i) \quad (12.12)$$

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

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

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

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:

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

$$\mathbf{C}^{-1} \cdot \mathbf{C} \cdot \mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{D}$$

$$\mathbf{I} \cdot \mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{D}$$

$$\mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{D} \quad (12.14)$$

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 \mathbf{C} and \mathbf{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:

$$(12.15) \quad w_i \left(\sum_{j=1}^n w_j \tilde{C}_{ij} + \mu \right) = w_i \tilde{C}_{i0} \quad \forall i = 1, \dots, n$$

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

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

The last term of this equation is zero because the weights sum to one. The first term is zero because the covariances \tilde{C}_{ij} are symmetric. The middle term is zero because the weights are unbiased. This leaves us with

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

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

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

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

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

Or, in terms of the matrices we defined earlier,

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

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:

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

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

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

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

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

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

with the modeled error variance given by

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

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

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

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

with the modeled error variance given by

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

The common practice in geostatistics is to calculate modeled variogram values then, for reasons of computational efficiency, to subtract them from some constant, usually $\hat{\sigma}^2$. 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

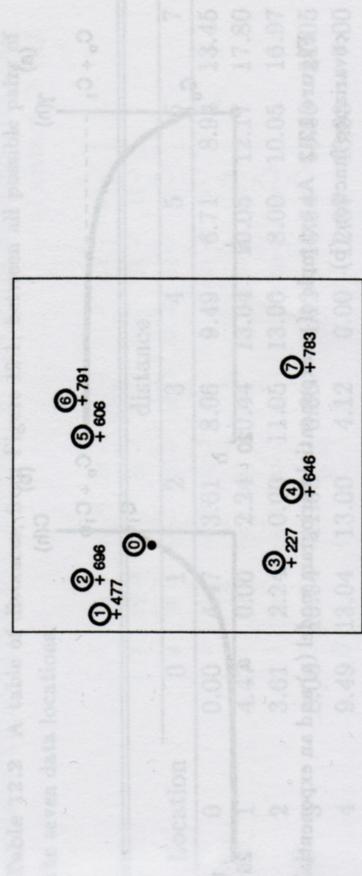


Figure 12.1 An example of a data configuration to illustrate the kriging estimator. This configuration was given earlier in Figure 11.1 where it was used to illustrate other estimation techniques discussed in Chapter 11. The sample value is given immediately to the right of the plus sign.

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

	Sample				Distance from
No.	X	Y	V	0	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

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

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

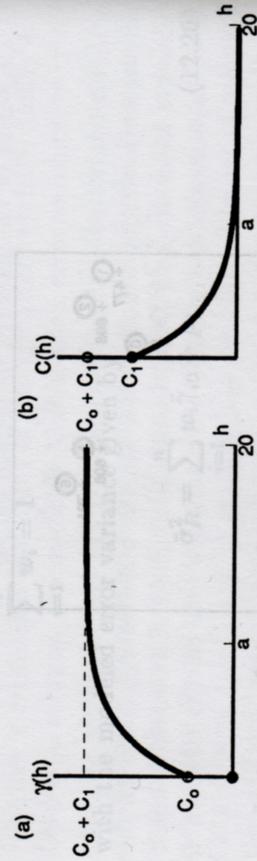


Figure 12.2 An example of an exponential variogram model (a) and an exponential covariance function (b).

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

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

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

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

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:

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

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

distances between every pair of locations, and Equation 12.25 above, the \mathbf{C} matrix is

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

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

The \mathbf{D} matrix is

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

The inverse of \mathbf{C} is

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

Ordinary Kriging

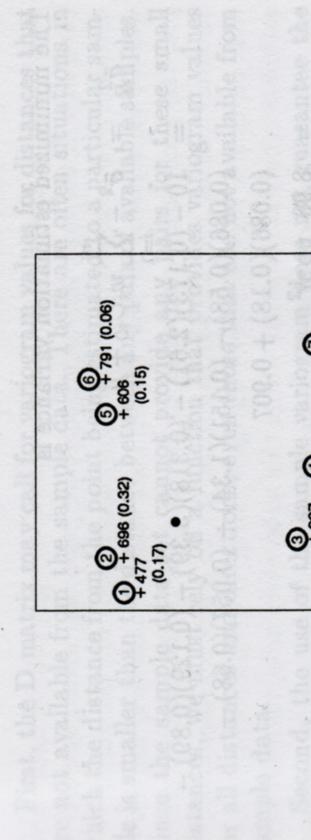


Figure 12.3 The ordinary kriging weights for the seven samples using the isotropic exponential covariance model given in Equation 12.25. The sample value is given immediately to the right of the plus sign while the kriging weights are shown in parenthesis.

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

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

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

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

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.

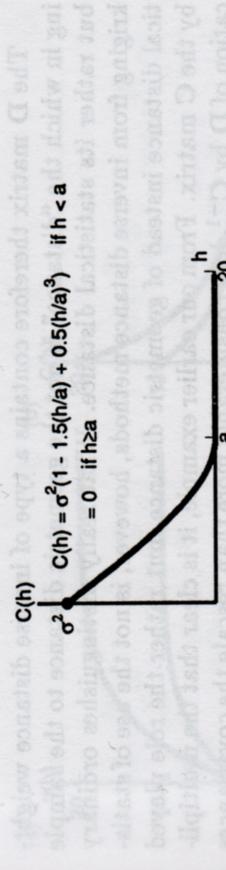


Figure 12.4 The spherical model of a covariance function.

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 minimized estimation variance is

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

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 $\hat{\sigma}_R^2$, 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 \mathbf{C} and \mathbf{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 \mathbf{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 \mathbf{C} and \mathbf{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 \mathbf{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