20. Kriging

Kriging is based on the correlogram $C(h)$ (or for isotropic data $C(h)$) which may have been estimated from our data (either directly or from the variogram and $C(0)$ using equation 20-12) or known a priori. There are basically two basic sorts of kriging, kriging with a known mean $\mu$ or model $m(x)$ and kriging in which the functional form of $m(x)$ is known but the values are unknown.

**Kriging with a Known Model or Mean (Simple Kriging)**

If the model for long wavelength trends (most simply just the mean) is known, can write the residual values

$$Y^*(x_0) = Z^*(x_0) - m(x_0)$$  \hspace{1cm} (20-1)

and

$$Y^*(x_0) = \sum_{i=1}^{m} \lambda_i Y(x_i)$$  \hspace{1cm} (20-2)

where the asterisk indicate estimates

The goal of Kriging is to determine the values of $\lambda_i$ that minimize the expected error variance of the estimate $Y^*$. We can write error variance as

$$e^2 = E \left[ Y^*(x_0) - Y(x_0) \right]^2$$  \hspace{1cm} (20-3)

where $E \{ \}$ is used to denote the expected value. We can expand equation (20-3)

$$e^2 = E \left[ Y^*(x_0) \right]^2 - 2E \left\{ Y^*(x_0) Y(x_0) \right\} + E \left[ Y(x_0) \right]^2$$  \hspace{1cm} (20-4)

Substituting for $Y^*$ using equation (20-2) and recognizing the last term as the variance of $Y$ yields

$$e^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j E \left\{ Y(x_i) Y(x_j) \right\} - 2 \sum_{i=1}^{n} \lambda_i E \left\{ Y(x_i) Y(x_0) \right\} + C(0)$$  \hspace{1cm} (20-5)

Now if we substitute the spatial covariance for the expected values we get

$$e^2 = \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j C(x_i - x_j) - 2 \sum_{i=1}^{n} \lambda_i C(x_i - x_0) + C(0)$$  \hspace{1cm} (20-6)

If we have knowledge of the spatial covariance based on the analysis of our data, the only unknowns are the $\lambda_i$ values. We wish to find those that minimize the error variance of our estimate $e$. To do this, we differentiate equation (20-18) with respect to the weights and set them to zero

$$\frac{de^2}{d\lambda_i} = 2 \sum_{j=1}^{m} \lambda_j C(x_i - x_j) - 2C(x_i - x_0) = 0$$  \hspace{1cm} (20-7)

Note that the 2 in front of the first term on the right hand term arises because the double summation yields two pairs of $\lambda_i \lambda_j$ for a given choice of $i$ and $j$ except for $i = j$ where the derivative of $\lambda_i^2$ is $2\lambda_i$. Equation (20-7) yields a set of equations
\[
\sum_{j=1}^{m} \lambda_j C(x_i - x_j) = C(x_i - x_0), \quad i = 1, 2, ..., m \tag{20-8}
\]

Now in the case of an isotropic correlogram we can write the equations as
\[
\sum_{j=1}^{m} \lambda_j C(h_{ij}) = C(h_{i0}), \quad i = 1, 2, ..., m
\]

where \(h_{ij} = h_{ji}\) is just the distance between the \(i^{th}\) and \(j^{th}\) point \(h_{i0}\) is the distance from the \(i^{th}\) point to the location of our estimate. We can write this in matrix notation as
\[
\begin{bmatrix}
C(h_{11}) & C(h_{12}) & C(h_{13}) & \cdots & C(h_{1m}) \\
C(h_{21}) & C(h_{22}) & C(h_{23}) & \cdots & C(h_{2m}) \\
C(h_{31}) & C(h_{32}) & C(h_{33}) & \cdots & C(h_{3m}) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
C(h_{m1}) & C(h_{m2}) & C(h_{m3}) & \cdots & C(h_{mm})
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\vdots \\
\lambda_m
\end{bmatrix}
= \begin{bmatrix}
C(h_{10}) \\
C(h_{20}) \\
C(h_{30}) \\
\vdots \\
C(h_{m0})
\end{bmatrix}
\tag{20-9}
\]

Recognizing that \(C(h_i) = C(0)\) and \(C(h_{ij}) = C(h_{ji})\) this becomes,
\[
\begin{bmatrix}
C(0) & C(h_{12}) & C(h_{13}) & \cdots & C(h_{1m}) \\
C(h_{12}) & C(0) & C(h_{23}) & \cdots & C(h_{2m}) \\
C(h_{13}) & C(h_{23}) & C(0) & \cdots & C(h_{3m}) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
C(h_{1m}) & C(h_{2m}) & C(h_{3m}) & \cdots & C(0)
\end{bmatrix}
\begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\lambda_3 \\
\vdots \\
\lambda_m
\end{bmatrix}
= \begin{bmatrix}
C(h_{10}) \\
C(h_{20}) \\
C(h_{30}) \\
\vdots \\
C(h_{m0})
\end{bmatrix}
\tag{20-10}
\]

In matrix notation we have
\[
C\lambda = \sigma_0
\tag{20-11}
\]
or
\[
\lambda = C^{-1}\sigma_0
\tag{20-12}
\]

Because \(C\) is symmetric and has positive values down the diagonal (i.e., it is positive definite) it will always be non-singulare invertible provided two points are not repeated. It can be solved to yield the coefficients for our estimate \(Y'(x_0)\) as given by equation (20-2). This is known as simple kriging. In practice only nearby points would be incorporated into each estimate.

We can also estimate the variance in \(Y'(x_0)\) which is \(e^2\) from (20-6) by substituting (20-8)
\[
e^2 = \sum_{i=1}^{m} \lambda_i \sum_{j=1}^{m} \lambda_j C(x_i - x_j) - 2 \sum_{i=1}^{m} \lambda_i C(x_i - x_0) + C(0)
\]
\[
= \sum_{i=1}^{m} \lambda_i C(x_i - x_0) - 2 \sum_{i=1}^{m} \lambda_i C(x_i - x_0) + C(0) + C(0)
\tag{20-13}
\]
\[
= C(0) - \sum_{i=1}^{m} \lambda_i C(x_i - x_0)
\]
or in the case of an isotropic correlogram
$$e^2 = C(0) - \sum_{i=1}^{m} \lambda_i C(h_{i0}) \quad (20-14)$$

**Kriging with an Unknown Mean or Model (Ordinary and Universal Kriging)**

In many practical cases the parameter values for the model of long wavelength variations are not known. One approach is to estimate them from the data – indeed, this is generally the approach that would be taken to estimate the correlogram from the data. However, an alternative approach is to include the unknown model in the Kriging.

We write our estimate in terms of \(Z(x_0)\) (rather than in terms of the residual \(Y(x_0)\) as for Simple Kriging (see equation 20-2))

$$Z^*(x_0) = \sum_{i=1}^{m} \lambda_i Z(x_i) \quad (20-15)$$

We can write our general model for the long wavelength trends as

$$m(x) = \sum_{l=0}^{L} a_l f^l(x) \quad (20-16)$$

where \(a_l\) are constants multiplying model functions \(f^l(x)\). In the simplest case of a model which comprises the mean, we would have \(L = 1, f^1(x) = 1\) and \(a_1\) the unknown mean \(\mu\).

As before we seek to minimize the expected squared difference between the estimated value and the true value which we can write

$$e^2 = E\left[ (Z^*(x_0) - Z(x_0))^2 \right] = \text{Variance}(Z^*(x_0) - Z(x_0)) + \left( E\left[ Z^*(x_0) - Z(x_0) \right] \right)^2 \quad (20-17)$$

The first term is just the normally distributed variance of the estimate relative to the true values while the second term accounts for a systematic offset or bias between the two, a quantity that would be set to zero if the long wavelength model were known.

This bias term can be expanded using equation (20-16)

$$E\left[ Z^*(x_0) - Z(x_0) \right] = \sum_{m=1}^{M} \sum_{l=1}^{m} \lambda_i a_l f^l(x_i) - \sum_{l=1}^{m} a_l f^l(x_0)$$

$$= \sum_{l=1}^{L} a_l \left( \sum_{i=1}^{m} \lambda_i f^l(x_i) - f^l(x_0) \right) \quad (20-18)$$

where in the final expression we have reversed the order of summation. To minimize equation (20-17) it is clear that the bias which requires that terms in parentheses of the final expression be zero

$$\sum_{i=1}^{m} \lambda_i f^l(x_i) - f^l(x_0) = 0 \quad l = 0, 1, 2, \ldots L \quad (20-19)$$

or

20-3
We can write the first term in equation (20-17) as before (see equations 20-3 to 20-6)

\[
\text{Variance}\left(Z^*\left(x_0\right) - Z\left(x_0\right)\right) = \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j C\left(x_i - x_j\right) - 2 \sum_{i=1}^{m} \lambda_i C\left(x_i - x_0\right) + C\left(0\right)
\]

where the covariances are those we have estimated for the data after removing the long-wavelength model. Our goal is to minimize the variance given by equation (20-21) subject to the constraints of equation (20-20). This is done by the method of Lagrange multipliers. We can write a function \(Q\) comprising the variance and the zero constraints multiplied by \(L\) additional unknowns \(2\mu_l\)

\[
Q = \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j C\left(x_i - x_j\right) - 2 \sum_{i=1}^{m} \lambda_i C\left(x_i - x_0\right) + C\left(0\right) + 2 \sum_{l=0}^{L} \mu_l \left[ \sum_{i=1}^{m} \lambda_i f^l\left(x_i\right) - f^l\left(x_0\right) \right]
\]

We minimize this function by differentiating with respect to \(\lambda_i\) and \(\mu_l\)

\[
\frac{\partial Q}{\partial \lambda_i} = 2 \sum_{j=1}^{m} \lambda_j C\left(x_i - x_j\right) - 2 C\left(x_i - x_0\right) + 2 \sum_{l=0}^{L} \mu_l f^l\left(x_i\right) = 0 \quad i = 1, 2, \ldots, m
\]

\[
\frac{\partial Q}{\partial \mu_l} = 2 \left[ \sum_{i=1}^{m} \lambda_i f^l\left(x_i\right) - f^l\left(x_0\right) \right] = 0 \quad l = 1, 2, \ldots, L
\]

We have \(m + L\) equations for \(m + L\) unknowns that we can write

\[
\sum_{j=1}^{m} \lambda_j C\left(x_i - x_j\right) + \sum_{l=0}^{L} \mu_l f^l\left(x_i\right) = 2 C\left(x_i - x_0\right) \quad i = 1, 2, \ldots, m
\]

\[
\sum_{i=1}^{m} \lambda_i f^l\left(x_i\right) = f^l\left(x_0\right) \quad l = 1, 2, \ldots, L
\]

which we can write in matrix notation in the form \(A\mathbf{x} = \mathbf{b}\)

\[
\begin{bmatrix}
C & F \\
F' & 0
\end{bmatrix}
\begin{bmatrix}
\lambda \\
\mu
\end{bmatrix} =
\begin{bmatrix}
\sigma_0 \\
\mathbf{f}_0
\end{bmatrix}
\]

where \(C\) and \(\sigma_0\) are as in equation (20-11), \(\lambda\) and \(\mu\) are vectors of the unknown coefficients and \(F\) and \(\mathbf{f}_0\) are given by

\[
F =
\begin{bmatrix}
f^1\left(x_1\right) & f^2\left(x_1\right) & f^3\left(x_1\right) & \cdots & f^L\left(x_1\right) \\
f^1\left(x_2\right) & f^2\left(x_2\right) & f^3\left(x_2\right) & \cdots & f^L\left(x_2\right) \\
f^1\left(x_3\right) & f^2\left(x_3\right) & f^3\left(x_3\right) & \cdots & f^L\left(x_3\right) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
f^1\left(x_m\right) & f^2\left(x_m\right) & f^3\left(x_m\right) & \cdots & f^L\left(x_m\right)
\end{bmatrix}
\]

and
These equations can be solved using standard inverse theory methods. The variance of the estimate is given by
\[ e^2 = C(0) - \sum_{i=1}^{m} \lambda_i C(x_i - x_0) - \sum_{i=1}^{l} \mu_i f^i(x_0) \]  
(20-30)

The solution of equation (20-27) is known as universal kriging. If our model of \( m(x) \) is just the mean then universal kriging reduces to solving equation (20-11) with the additional constraint
\[ \sum_{j=1}^{n} \lambda_j = 1 \]  
(20-31)

This is known as ordinary kriging.