Geostatistical Co-Kriging

These notes are intended to explain the confusing topic of co-kriging in geostatistical analysis and provide an example of the procedure as it should be conducted. We begin by specifying the setting in which co-kriging will be considered, discuss several versions of cross-variograms which are the foundation of co-kriging, and develop the predictive kriging equations.

Suppose that we have available geo-referenced data on two responses or attributes of interest which are taken to be realizations of the processes $\{Z_1(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ and $\{Z_2(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$. The data on Z_1 and Z_2 may be available at the same locations $\{\boldsymbol{s}_i : i = 1, \ldots, n\}$ or may be available for Z_1 at locations $\{\boldsymbol{s}_i : i = 1, \ldots, n_1\}$ and available for Z_2 at locations $\{\boldsymbol{s}_g : g = 1, \ldots, n_2\}$. There may be overlap in these sets of locations or there may not be, and n_1 may be the same as n_2, or it may not be. To be as general as possible, we will assume that the sets of locations at which Z_1 is observed, namely $\{\boldsymbol{s}_i : i = 1, \ldots, n_1\}$, is not the same as the set of locations at which Z_2 is observed, namely $\{\boldsymbol{s}_g : g = 1, \ldots, n_2\}$ and that $n_1 \neq n_2$. Everything in what follows may be generalized to more than two responses although the notation needed becomes increasingly more complex as the number of responses increases. We will assume that both the Z_1 and Z_2 processes are intrinsically stationary and that, in addition, the dependence between those processes at various locations depends only on displacement $\boldsymbol{h}_{i,g} = \boldsymbol{s}_i - \boldsymbol{s}_g$. This is not absolutely necessary but will simplify our presentation and is often assumed in practice.

1 Cross-Variograms

There have been several quantities proposed as definitions of cross-variograms for two responses. The original proposal was, in terms of displacement $\boldsymbol{h}$,

$$2\nu_{1,2}(\boldsymbol{h}) = cov\{Z_1(\boldsymbol{s}+\boldsymbol{h}) - Z_1(\boldsymbol{s}),\ Z_2(\boldsymbol{s}+\boldsymbol{h}) - Z_2(\boldsymbol{s})\}. \tag{1}$$

Notice that this definition involves Z_1 and Z_2 at the same locations $\boldsymbol{s}$ and $\boldsymbol{s}+\boldsymbol{h}$. In order to estimate this cross-variogram we need data on the two processes at the same locations.

An alternative definition of a cross-variogram for Z_1 and Z_2 is,

$$2\gamma_{1,2} = var\{Z_1(\boldsymbol{s}+\boldsymbol{h}) - Z_2(\boldsymbol{s})\}. \tag{2}$$

Estimation of this cross-variogram will not require observed data on Z_1 and Z_2 at the same locations, only replication of displacements $\boldsymbol{h}$ (most likely within certain displacement tolerance regions as for typical univariate variograms).

There is an important distinction between either of the cross-variograms in expressions (1) and (2) and the typical univariate variogram for a single response variable Z,

$$2\gamma(\boldsymbol{h}) = var\{Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})\} \tag{3}$$

This distinction is that, by definition, $2\gamma(\mathbf{0}) = 0$, while neither $2\nu_{1,2}(\mathbf{0})$ nor $2\gamma_{1,2}(\mathbf{0})$ are necessarily equal to 0. This will have an impact on kriging equations developed for use with cross-variograms along with ordinary variograms.

Another important point to make note of is that, by definition, $2\nu_{1,2}(\boldsymbol{h})$ is not necessarily equal to $2\nu_{2,1}(\boldsymbol{h})$ nor is $2\gamma_{1,2}(\boldsymbol{h})$ necessarily equal to $2\gamma_{2,1}(\boldsymbol{h})$. However, to make use of $2\nu_{1,2}(\boldsymbol{h})$ in prediction it is necessary to assume that $2\nu_{1,2}(\boldsymbol{h}) = 2\nu_{2,1}(\boldsymbol{h})$ (e.g., Cressie and Wikle 1998). The use of $2\nu_{1,2}(\boldsymbol{h})$ in prediction requires data on Z_1 and Z_2 at the same locations for estimation, and also requires the assumption that $2\nu_{1,2}(\boldsymbol{h}) = 2\nu_{2,1}(\boldsymbol{h})$. Neither of these are needed to make use of $2\gamma_{1,2}(\boldsymbol{h})$ in prediction. Thus, in what follows we will take expression (2) as the basic definition of a cross-variogram.

Reduced versions of the cross-variograms of expressions (1) and (2) are sometimes presented as the basic definition, although these versions are correct only under certain assumptions. If it is assumed, as we will in what follows, that $E\{Z_1(\boldsymbol{s})\} = \mu_1$ and $E\{Z_2(\boldsymbol{s})\} = \mu_2$ for all $\boldsymbol{s} \in \mathcal{D}$, then expression (1) reduces to

$$2\nu_{1,2}(\boldsymbol{h}) = E[\{Z_1(\boldsymbol{s}+\boldsymbol{h}) - Z_1(\boldsymbol{s})\}\,\{Z_2(\boldsymbol{s}+\boldsymbol{h}) - Z_2(\boldsymbol{s})\}], \tag{4}$$

and expression (2) reduces to

$$2\gamma_{1,2}(\boldsymbol{h}) = E\{Z_1(\boldsymbol{s}+\boldsymbol{h}) - Z_2(\boldsymbol{s})\}^2 - (\mu_1 - \mu_2)^2. \tag{5}$$

For the reasons given previously, we will focus on the use of $2\gamma_{1,2}(\boldsymbol{h})$ in development of cokriging equations. The progression of cokriging is similar to that of ordinary kriging, namely, (1) empirical estimation of variograms and the cross-variogram, (2) fitting theoretical models to those variograms, and (3) prediction to minimize expected squared error loss (i.e., prediction mean squared error). We will first present the form of the optimal predictor based on the variograms and cross-variogram and will then discuss estimation of these functions.

2 Cokriging

In this section we develop cokriging equations that make use of observed values of both $\{Z_1(\boldsymbol{s}_i) : i = 1, \ldots, n_1\}$ and $\{Z_2(\boldsymbol{s}_g) : g = 1, \ldots, n_2\}$ to predict $Z_1(\boldsymbol{s}_0)$ at an unobserved location $\boldsymbol{s}_0$. Parallel versions of what follows can be easily developed for prediction of $Z_2(\boldsymbol{s}_0)$ if that is desired. Assume that variograms $2\gamma_1(\boldsymbol{s}_i - \boldsymbol{s}_j)$ and $2\gamma_2(\boldsymbol{s}_g - \boldsymbol{s}_h)$ are available for the Z_1 and Z_2 processes, respectively and similarly for the cross-variogram $2\gamma_{1,2}(\boldsymbol{s}_i - \boldsymbol{s}_g)$ for all $\boldsymbol{s}_i, \boldsymbol{s}_j, \boldsymbol{s}_g$, and $\boldsymbol{s}_h$ in the domain $\mathcal{D}$.

2.1 Linear, Unbiased Prediction

Our objective is to predict the Z_1 process at an unobserved location $\boldsymbol{s}_0$. As in ordinary kriging we first restrict attention to linear unbiased predictors,

$$p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0) = \sum_{i=1}^{n_1} \lambda_{1,i} Z_1(\boldsymbol{s}_i) + \sum_{g=1}^{n_2} \lambda_{2,g} Z_2(\boldsymbol{s}_g), \tag{6}$$

such that

$$E\{p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\} = E\{Z_1(\boldsymbol{s}_0)\} = \mu_1.$$

The unbiasedness condition will be met if and only if

$$\begin{aligned} \sum_{i=1}^{n_1} \lambda_{1,i} &= 1 \\ \sum_{g=1}^{n_2} \lambda_{2,g} &= 0. \end{aligned} \tag{7}$$

2.2 Prediction Mean Squared Error

Our goal is to find coefficients $\{\lambda_{1,i} : i = 1, \ldots, n_1\}$ and $\{\lambda_{2,g} : g = 1, \ldots, n_2\}$ such that the predictor (6) minimizes

$$E\left[\{Z_1(\boldsymbol{s}_0) - p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\}^2\right] = E\left[\left\{Z_1(\boldsymbol{s}_0) - \sum_{i=1}^{n_1} \lambda_{1,i} Z_1(\boldsymbol{s}_i) - \sum_{g=1}^{n_2} \lambda_{2,g} Z_2(\boldsymbol{s}_g)\right\}^2\right], \tag{8}$$

subject to the restrictions of (7). As in the development of ordinary kriging, we will show that the prediction mean squared error (8) can be written in terms of variograms and cross-variograms. To accomplish this we will make use of several algebraic identities and so we present those first.

2.2.1 Algebraic Identity 1

For any real numbers $\{a_i : i = 1, \ldots, n\}$ and $\{W_i : i = 1, \ldots, n\}$

$$\left(\sum_{i=1}^{n} a_i W_i\right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j W_i W_j. \tag{9}$$

The proof of this first identity is straightforward.

2.2.2 Algebraic Identity 2

For any real numbers $\{a_i : i = 1, \ldots, n\}$ and $\{W_i : i = 1, \ldots, n\}$ such that $\sum a_i = 0$,

$$\left(\sum_{i=1}^{n} a_i W_i\right)^2 = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (W_i - W_j)^2. \tag{10}$$

<u>Proof</u>

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (W_i - W_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (W_i^2 - 2W_i W_j + W_j^2) \\
&= \sum_{i=1}^{n} a_i W_i^2 \left(\sum_{j=1}^{n} a_j\right) - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j W_i W_j + \sum_{j=1}^{n} a_j W_j^2 \left(\sum_{i=1}^{n} a_i\right) \\
&= -2\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j W_i W_j, \qquad (11)
\end{aligned}$$

which proves the identity. The third line of the proof follows because the a_i sum to zero.

2.2.3 Algebraic Identity 3

For any real numbers $\{W_i : i = 0, \ldots, n_1\}$, $\{Z_j : j = 1, \ldots, n_2\}$, $\{a_i : i = 0, \ldots n_1\}$ and $\{b_j : j = 1, \ldots, n_2\}$ such that $\sum_i a_i = 0$, $\sum_j b_j = 0$ and $a_0 = 1$,

$$\left(\sum_{i=0}^{n_1} a_i W_i\right)\left(\sum_{j=1}^{n_2} b_j Z_j\right) = -\frac{1}{2}\sum_{j=1}^{n_2} b_j (W_0 - Z_j)^2 - \frac{1}{2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_i b_j (W_i - Z_j)^2. \tag{12}$$

<u>Proof</u>

$$\begin{aligned}
& \sum_{j=1}^{n_2} b_j (W_0 - Z_j)^2 + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_i b_j (W_i - Z_j)^2 \\
= \ & \sum_{j=1}^{n_2} \left(b_j W_0^2 - 2b_j W_0 Z_j + b_j Z_j^2\right) + \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_i b_j \left(W_i^2 - 2W_i Z_j + Z_j^2\right) \\
= \ & W_0^2 \sum_{j=1}^{n_2} b_j - 2W_0 \sum_{j=1}^{n_2} b_j Z_j + \sum_{j=1}^{n_2} b_j Z_j^2 + \left(\sum_{i=1}^{n_1} a_i W_i^2\right)\left(\sum_{j=1}^{n_2} b_j\right) \\
- \ & 2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_i b_j W_i Z_j + \left(\sum_{i=1}^{n_1} a_i\right)\left(\sum_{j=1}^{n_2} b_j Z_j^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= 0 - 2W_0\sum_{j=1}^{n_2} b_j Z_j + \sum_{j=1}^{n_2} b_j Z_j^2 + 0 - 2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_i b_j W_i Z_j - \sum_{j=1}^{n_2} b_j Z_j^2 \\
&= -2W_0\sum_{j=1}^{n_2} b_j Z_j - 2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} a_i b_j W_i Z_j \\
&= -2W_0\sum_{j=1}^{n_2} b_j Z_j - 2\left(\sum_{i=1}^{n_1} a_i W_i\right)\left(\sum_{j=1}^{n_2} b_j Z_j\right). \qquad (13)
\end{aligned}
$$

Now, with $a_0 = 1$,

$$
\left(\sum_{i=0}^{n_1} a_i W_i\right)\left(\sum_{j=1}^{n_2} b_j Z_j\right) = W_0\sum_{j=1}^{n_2} b_j Z_j + \left(\sum_{i=1}^{n_1} a_i W_i\right)\left(\sum_{j=1}^{n_2} b_j Z_j\right), \qquad (14)
$$

and proof of the identity follows from the last lines of (13) and (14).

2.2.4 Back to Prediction Mean Squared Error

To make use of the algebraic identities presented previously, note that if we define $\lambda_{1,0} = 1$, $\lambda'_{1,i} = -\lambda_{1,i}$ for $i = 1, \ldots, n_1$, and $\lambda_{2,g} = -\lambda_{2,g}$ for $g = 1, \ldots, n_2$, then the squared error term of expression (8) may be written as,

$$
\begin{aligned}
\{Z_1(\boldsymbol{s}_0) - p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\}^2 &= \left\{\sum_{i=0}^{n_1} \lambda'_{1,i} Z_1(\boldsymbol{s}_i) + \sum_{g=1}^{n_2} \lambda'_{2,g} Z_2(\boldsymbol{s}_g)\right\}^2 \\
&= \left[\sum_{i=0}^{n_1} \lambda'_{1,i} Z_1(\boldsymbol{s}_i)\right]^2 \\
&+ 2\left[\sum_{i=0}^{n_1} \lambda'_{1,i} Z_1(\boldsymbol{s}_i)\right]\left[\sum_{g=1}^{n_2} \lambda'_{2,g} Z_2(\boldsymbol{s}_g)\right] \\
&+ \left[\sum_{g=1}^{n_2} \lambda'_{2,g} Z_2(\boldsymbol{s}_g)\right]^2. \qquad (15)
\end{aligned}
$$

Note that $\sum_{i=0}^{n_1} \lambda'_{1,i} = 0$ and $\sum_{i=1}^{n_1} \lambda'_{1,i} = -1$ like the real numbers a_i in Algebraic Identity 3 and $\sum_{g=1}^{n_2} \lambda'_{2,g} = 0$ like the real numbers b_j in that identity.

Now, applying Algebraic Identity 2 to the second and fourth lines of expression (15), and Algebraic Identity 3 to the third line results in
$\{Z_1(\boldsymbol{s}_0) - p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\}^2 =$

$$
- \quad \frac{1}{2}\sum_{i=0}^{n_1}\sum_{j=0}^{n_1} \lambda'_{1,i}\lambda'_{1,j}\{Z_1(\boldsymbol{s}_i) - Z_1(\boldsymbol{s}_j)\}^2
$$

$$- \sum_{g=1}^{n_2} \lambda'_{2,g}\{Z_1(\boldsymbol{s}_0) - Z_2(\boldsymbol{s}_g)\}^2 - \sum_{i=1}^{n_1}\sum_{g=1}^{n_2} \lambda'_{1,i}\lambda'_{2,g}\{Z_1(\boldsymbol{s}_i) - Z_2(\boldsymbol{s}_g)\}^2$$
$$- \frac{1}{2}\sum_{g=1}^{n_2}\sum_{h=1}^{n_2} \lambda'_{2,g}\lambda'_{2,h}\{Z_2(\boldsymbol{s}_g) - Z_2(\boldsymbol{s}_h)\}^2. \tag{16}$$

Replacing the $\lambda'_{1,i}$ and $\lambda'_{2,g}$ with their definitions in terms of the original $\lambda_{1,i}$ and $\lambda_{2,g}$ then gives
$\{Z_1(\boldsymbol{s}_0) - p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\}^2 =$

$$\sum_{i=1}^{n_1} \lambda_{1,i}\{Z_1(\boldsymbol{s}_0) - Z_1(\boldsymbol{s}_i)\}^2 - \frac{1}{2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} \lambda_{1,i}\lambda_{1,j}\{Z_1(\boldsymbol{s}_i) - Z_1(\boldsymbol{s}_j)\}^2$$
$$+ \sum_{g=1}^{n_2} \lambda_{2,g}\{Z_1(\boldsymbol{s}_0) - Z_2(\boldsymbol{s}_g)\}^2 - \sum_{i=1}^{n_1}\sum_{g=1}^{n_2} \lambda_{1,i}\lambda_{2,g}\{Z_1(\boldsymbol{s}_i) - Z_2(\boldsymbol{s}_g)\}^2$$
$$- \frac{1}{2}\sum_{g=1}^{n_2}\sum_{h=1}^{n_2} \lambda_{2,g}\lambda_{2,h}\{Z_2(\boldsymbol{s}_g) - Z_2(\boldsymbol{s}_h)\}^2. \tag{17}$$

We are now ready to produce the prediction mean squared error of expression (8) by taking expectations in (17). With constant means $E\{Z(\boldsymbol{s})\} = \mu_1$ and $E\{Z_2(\boldsymbol{s})\} = \mu_2$ (refer back to expression (5)),

$$\begin{aligned}
E\{Z_1(\boldsymbol{s}_0) - Z_1(\boldsymbol{s}_i)\}^2 &= 2\gamma_1(\boldsymbol{s}_0 - \boldsymbol{s}_i)\\
E\{Z_1(\boldsymbol{s}_i) - Z_1(\boldsymbol{s}_j)\}^2 &= 2\gamma_1(\boldsymbol{s}_i - \boldsymbol{s}_j)\\
E\{Z_1(\boldsymbol{s}_0) - Z_2(\boldsymbol{s}_g)\}^2 &= 2\gamma_{1,2}(\boldsymbol{s}_0 - \boldsymbol{s}_g) + (\mu_1 - \mu_2)^2\\
E\{Z_1(\boldsymbol{s}_i) - Z_2(\boldsymbol{s}_g)\}^2 &= 2\gamma_{1,2}(\boldsymbol{s}_i - \boldsymbol{s}_g) + (\mu_1 - \mu_2)^2\\
E\{Z_2(\boldsymbol{s}_g) - Z_2(\boldsymbol{s}_h)\}^2 &= 2\gamma_2(\boldsymbol{s}_g - \boldsymbol{s}_h).
\end{aligned} \tag{18}$$

Using these expectations in (17) we are able to write the prediction mean squared error (8) in the form
$E\left[\{Z_1(\boldsymbol{s}_0) - p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\}^2\right] =$

$$\sum_{i=1}^{n_1} \lambda_{1,i} 2\gamma_1(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \frac{1}{2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} \lambda_{1,i}\lambda_{1,j} 2\gamma_1(\boldsymbol{s}_i - \boldsymbol{s}_j)$$
$$+ \sum_{g=1}^{n_2} \lambda_{2,g}\left\{2\gamma_{1,2}(\boldsymbol{s}_0 - \boldsymbol{s}_g) + (\mu_1 - \mu_2)^2\right\} - \sum_{i=1}^{n_1}\sum_{g=1}^{n_2} \lambda_{1,i}\lambda_{2,g}\left\{2\gamma_{1,2}(\boldsymbol{s}_i - \boldsymbol{s}_g) + (\mu_1 - \mu_2)^2\right\}$$
$$- \frac{1}{2}\sum_{g=1}^{n_2}\sum_{h=1}^{n_2} \lambda_{2,g}\lambda_{2,h} 2\gamma_2(\boldsymbol{s}_g - \boldsymbol{s}_h). \tag{19}$$

Finally, noting again that $\sum \lambda_{1,i} = 1$and $\sum \lambda_{2,g} = 1$, and bringing the multipliers 2 out front of the summations we arrive at
$E\left[\{Z_1(\boldsymbol{s}_0) - p_1(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)\}^2\right] =$

$$\begin{aligned}
& 2\sum_{i=1}^{n_1} \lambda_{1,i}\gamma_1(\boldsymbol{s}_0 - \boldsymbol{s}_i) - \sum_{i=1}^{n_1}\sum_{j=1}^{n_1} \lambda_{1,i}\lambda_{1,j}\gamma_1(\boldsymbol{s}_i - \boldsymbol{s}_j) \\
+ \quad & 2\sum_{g=1}^{n_2} \lambda_{2,g}\gamma_{1,2}(\boldsymbol{s}_0 - \boldsymbol{s}_g) - 2\sum_{i=1}^{n_1}\sum_{g=1}^{n_2} \lambda_{1,i}\lambda_{2,g}\gamma_{1,2}(\boldsymbol{s}_i - \boldsymbol{s}_g) \\
- \quad & \sum_{g=1}^{n_2}\sum_{h=1}^{n_2} \lambda_{2,g}\lambda_{2,h}\gamma_2(\boldsymbol{s}_g - \boldsymbol{s}_h).
\end{aligned} \tag{20}$$

2.3 Cokriging Equations

Recall our objective is to minimize the prediction mean squared error (20) subject to the restrictions given in expression (7). Using two Lagrange multipliers the objective function becomes expression (2) plus the two additional terms

$$-2m_1\left(\sum_{i=1}^{n_1} \lambda_{1,i} - 1\right) - 2m_2\left(\sum_{g=1}^{n_2} \lambda_{2,g}\right).$$

Taking derivatives and setting them equal to zero yields the sets of equations. For $i = 1, \ldots, n_1$,

$$\sum_{j=1}^{n_1} \lambda_{1,j}\gamma_1(\boldsymbol{s}_i - \boldsymbol{s}_j) + \sum_{g=1}^{n_2} \lambda_{2,g}\gamma_{1,2}(\boldsymbol{s}_i - \boldsymbol{s}_g) + m_1 = \gamma_1(\boldsymbol{s}_0 - \boldsymbol{s}_i), \tag{21}$$

and $\sum \lambda_{1,i} = 1$. For $g = 1, \ldots, n_2$,

$$\sum_{h=1}^{n_2} \lambda_{2,h}\gamma_2(\boldsymbol{s}_g - \boldsymbol{s}_h) + \sum_{i=1}^{n_1} \lambda_{1,i}\gamma_{1,2}(\boldsymbol{s}_i - \boldsymbol{s}_g) + m_2 = \gamma_2(\boldsymbol{s}_0 - \boldsymbol{s}_g), \tag{22}$$

and $\sum \lambda_{2,g} = 0$. These equations are gathered into matrix form as follows. Let Γ_1 be the $n_1 \times n_1$ matrix

$$\Gamma_1 = \begin{pmatrix} \gamma_1(\boldsymbol{s}_1 - \boldsymbol{s}_1) & \gamma_1(\boldsymbol{s}_1 - \boldsymbol{s}_2) & \ldots & \gamma_1(\boldsymbol{s}_1 - \boldsymbol{s}_{n_1}) \\ \gamma_1(\boldsymbol{s}_2 - \boldsymbol{s}_1) & \gamma_1(\boldsymbol{s}_2 - \boldsymbol{s}_2) & \ldots & \gamma_1(\boldsymbol{s}_2 - \boldsymbol{s}_{n_1}) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_1(\boldsymbol{s}_{n_1} - \boldsymbol{s}_1) & \gamma_1(\boldsymbol{s}_{n_1} - \boldsymbol{s}_2) & \ldots & \gamma_1(\boldsymbol{s}_{n_1} - \boldsymbol{s}_{n_1}) \end{pmatrix}$$

To avoid notational confusion, now take the set of n_2 locations at which Z_2 is observed to be denoted as $\{\boldsymbol{v}_g : g = 1, \ldots, n_2\}$. Then let Γ_2 be the $n_2 \times n_2$ matrix

$$\Gamma_2 = \begin{pmatrix} \gamma_2(\boldsymbol{v}_1 - \boldsymbol{v}_1) & \gamma_2(\boldsymbol{v}_1 - \boldsymbol{v}_2) & \ldots & \gamma_2(\boldsymbol{v}_1 - \boldsymbol{v}_{n_2}) \\ \gamma_2(\boldsymbol{v}_2 - \boldsymbol{v}_1) & \gamma_2(\boldsymbol{v}_2 - \boldsymbol{v}_2) & \ldots & \gamma_2(\boldsymbol{s}_2 - \boldsymbol{v}_{n_2}) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_2(\boldsymbol{v}_{n_2} - \boldsymbol{v}_1) & \gamma_2(\boldsymbol{v}_{n_2} - \boldsymbol{v}_2) & \ldots & \gamma_2(\boldsymbol{v}_{n_2} - \boldsymbol{v}_{n_2}) \end{pmatrix}$$

Also define the $n_1 \times n_2$ matrix $\Gamma_{1,2}$ as,

$$\Gamma_{1,2} = \begin{pmatrix} \gamma_{1,2}(\boldsymbol{s}_1 - \boldsymbol{v}_1) & \gamma_{1,2}(\boldsymbol{s}_1 - \boldsymbol{v}_2) & \ldots & \gamma_{1,2}(\boldsymbol{s}_1 - \boldsymbol{v}_{n_2}) \\ \gamma_{1,2}(\boldsymbol{s}_2 - \boldsymbol{v}_1) & \gamma_{1,2}(\boldsymbol{s}_2 - \boldsymbol{v}_2) & \ldots & \gamma_{1,2}(\boldsymbol{s}_2 - \boldsymbol{v}_{n_2}) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{1,2}(\boldsymbol{s}_{n_1} - \boldsymbol{v}_1) & \gamma_{1,2}(\boldsymbol{s}_{n_1} - \boldsymbol{v}_2) & \ldots & \gamma_{1,2}(\boldsymbol{s}_{n_1} - \boldsymbol{v}_{n_2}) \end{pmatrix}$$

Let $\mathbf{1}_k$ denote a $k-$vector of 1s and $\mathbf{0}_k$ a $k-$vector of 0s, both as column vectors. Construct the $(n_1 + n_2 + 2) \times (n_1 + n_2 + 2)$ matrix Γ_c as

$$\Gamma_c = \begin{pmatrix} \Gamma_1 & \Gamma_{1,2} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \Gamma_{1,2}^T & \Gamma_2 & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \\ \mathbf{1}_{n_1}^T & \mathbf{0}_{n_2}^T & 0 & 0 \\ \mathbf{0}_{n_1}^T & \mathbf{1}_{n_2}^T & 0 & 0 \end{pmatrix} \tag{23}$$

Further define

$$\begin{aligned} \lambda_c &= (\lambda_{1,1}, \ldots, \lambda_{1,n_1}, \lambda_{2,1}, \ldots, \lambda_{2,n_2}, m_1, m_2)^T \\ \gamma_c &= (\gamma_1(\boldsymbol{s}_0 - \boldsymbol{s}_1), \ldots, \gamma_1(\boldsymbol{s}_0 - \boldsymbol{s}_{n_1}), \gamma_{1,2}(\boldsymbol{s}_0 - \boldsymbol{v}_1), \ldots, \gamma_{1,2}(\boldsymbol{s}_0 - \boldsymbol{v}_{n_2}), 1, 0)^T \end{aligned} \tag{24}$$

The co-kriging equations are then given through the quantities in (23) and (24) as $\Gamma_c \lambda_c = \gamma_c$, and the co-kriging weights for use in the predictor of expression (6) are the first $n_1 + n_2$ elements of

$$\lambda_c = \Gamma_c^{-1} \gamma_c. \tag{25}$$

3 Empirical Variograms and Fitting Variogram Models

Now that we have seen the way that cross-variograms enter into the minimum mean squared error predictor for $Z_1(\boldsymbol{s}_0)$ we turn our attention to estimation of the empirical cross-variogram and fitting theoretical models to those empirical quantities. It is natural to think of using Matheron's moment-based estimator for empirical cross-variograms. Recall that for univariate variograms, Matheron's estimator assumes constant mean $E\{Z(\boldsymbol{s})\} = \mu$ for all $\boldsymbol{s} \in \mathcal{D}$. This is so that

$$2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j) = E\left[\{Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)\}^2\right]$$

and the expected value is replaced with an averaging operator in estimation. Expression (5) indicates that for a cross-variogram to equal an expected squared difference we need not only constant means but equal means for $Z_1(\boldsymbol{s})$ and $Z_2(\boldsymbol{s})$. The squared difference in means in the rightmost term of (5) is not an expected value and cannot be directly estimated with a sample average.

Suppose, however, that $E\{Z_1(\boldsymbol{s})\} = E\{Z_2(\boldsymbol{s})\}$ for all $\boldsymbol{s} \in \mathcal{D}$. Then we could use a moment estimator for a displacement class $N(\boldsymbol{h})$,

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} \{Z_1(\boldsymbol{s}_i) - Z_2(\boldsymbol{s}_g)\}^2, \tag{26}$$

where $N(\boldsymbol{h})$ is the set of all pairs of locations $\boldsymbol{s}_i, \boldsymbol{s}_g$ that have displacement within a defined tolerance region. Often, as for univariate variograms, we will take $N(\boldsymbol{h})$ to be defined by a distance bin.

In this fortunate situation we could then also proceed to fit theoretical variogram models (e.g., spherical, exponential) to the empirical cross-variogram values, and replace $\gamma_{1,2}(\cdot)$ in the matrix $\Gamma_{1,2}$ of expression (23) with estimated values. Note, however, that because the cross-variogram does not necessarily have a value of 0 at displacement $\mathbf{0}$, the diagonal values of $\Gamma_{1,2}$ will not be zero as they are by definition for Γ_1 and Γ_2.

Now, is it possible to use this simple version of empirical cross-variograms and procedure for fitting theoretical models? Fortunately, the answer is yes. Cressie and Wikle(1998) show that co-kriging predictors are equivariant to location and scale transformations. A statistical operator ϕ (estimator, predictor) is equivariant to location and scale transformation if

$$\phi(Z) = s\phi(Y) + m,$$

where $Y = (Z - m)/s$ for constants m and s.

This means that, if $p(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)$ is the predictor of $Z(\boldsymbol{s}_0)$ and we take standardized versions of Z_1 and Z_2 as

$$\begin{aligned} Y_1(\boldsymbol{s}_i) &= \frac{Z_1(\boldsymbol{s}_i) - \bar{Z}_1}{S_1} \\ Y_2(\boldsymbol{s}_g) &= \frac{Z_2(\boldsymbol{s}_g) - \bar{Z}_2}{S_2} \end{aligned} \tag{27}$$

where $\bar{Z}_1$ and S_1 are the usual sample mean and standard deviation of observed values $\{Z_1(\boldsymbol{s}_i) : i = 1, \ldots, n\}$ and similarly for $\bar{Z}_2$ and S_2, then we can compute $p(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0)$ by working with the standardized variables and then back-transforming the result. Specifically, let $p(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{s}_0)$ be the cokriging predictor of $Y_1(\boldsymbol{s}_0)$. Then the cokriging predictor of $Z_1(\boldsymbol{s}_0)$ is,

$$p(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \boldsymbol{s}_0) = S_1 p(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{s}_0) + \bar{Z}_1. \tag{28}$$

Because Y_1 and Y_2 have equal means (i.e., zero) we can use the estimator (26) with these quantities, fit standard variogram models to the results, and produce cokriging predictors of $Y_1(\boldsymbol{s}_0)$ as in the previous section. Estimation of empirical cross-variograms and fitting appropriate theoretical models becomes more difficult if this standardization is not undertaken. Another benefit of the standardization is that the variances of Y_1 and Y_2 are also equal, which prevents "swamping" of spatial structure by variance in estimation, which can occur if the variances of Z_1 and Z_2 are vastly different.