JANUARY 9, 2022

Cokriging with Unequally Sampled Data

Luis Davila • University of Alberta

Clayton V. Deutsch • University of Alberta

Learning Objectives

- **Review theory of standardized ordinary cokriging with unequally sampled data**
- **Understand the application and workflow of standardized ordinary cokriging**
- **Compare results of ordinary kriging to standardized ordinary cokriging**
- **Understand how ordinary kriging compares to standardized ordinary cokriging in practice and how results are checked (source code available).**

Introduction

Cokriging with unequally sampled data could be used with drill holes (resource and reserve estimation drilling campaigns - exploration data) and blast holes (or other production samples) to improve short- and medium-term models. Both data types complement each other: (1) drill hole data for exploration and estimation stages typically have better quality control than production sampling, and (2) production sampling is more abundant and closely spaced than exploration data. There are many other situations where multiple data types or variables are available at different locations including different vintages of sampling, different drilling types (reverse circulation and diamond drilling) and relatively inexpensive chip or channel samples together with drill hole data.

Cokriging is recommended when the data types are at different locations. This is sometimes referred to as heterotopic data. When multiple variables are measured at the same locations (equally sampled, or homotopic data) kriging can be used to estimate each variable one at a time. Decorrelation techniques can facilitate simulation of homotopic data, techniques like Projection Pursuit Multivariate Transform (Barnett & Deutsch, 2017) have been presented in previous lessons and the reader is encouraged to review them at GeostatisticsLessons.com. Cokriging optimally combines multiple data types accounting for error, bias and different support of the measurements (Minnitt & Deutsch, 2014). This lesson will present standardized ordinary cokriging for drill holes and blast holes. In practice, any number of data types could be considered and the approach could be applied in a non-mining context.

An essential component of cokriging is to have a valid Linear Model of Coregionalization (LMC) for the direct and cross variograms. For the drill hole and blast hole example considered in this Lesson, there are two direct experimental variograms: one for the drill hole data and one for the blast hole data. There is also a cross-covariance between the drill hole and blast hole data. This cross covariance is made to look like a cross variogram for the conventional practice of variogram fitting. In the end, there are three variograms in three directions that must be fit simultaneously for a valid LMC to solve the cokriging equations.

Multiple elements measured in the drill hole and blast hole data will significantly increase the number of variograms required and automatic fitting would be required. In an estimation context, each variable could be considered separately unless they are correlated and one or more are undersampled relative to the rest.

Theory

Consider two variables: (a) Primary variable from exploration data $\{Z(\mathbf{u}_\alpha), \text{where: } \alpha = 1, \ldots, n_Z\}$, and (b) Secondary variable from production sampling: $\left\{Y\left(\mathbf{u}'_\beta\right), \text{where: } \beta = 1, \ldots, n_Y\right\}$. Primary and secondary data will be used to characterize the spatial variability. Both variables are standardized to have a mean of zero and a variance of one by subtracting the mean and dividing by the standard deviation (Rossi & Deutsch, 2013). The stationary domains have been previously defined for different zones of the mineral deposit. The variables are standardized:

$$Z_{std} = \frac{Z(\mathbf{u}) - m_Z}{\sigma_Z} \qquad Y_{std} = \frac{Y(\mathbf{u}) - m_Y}{\sigma_Y} \tag{1}$$

After the estimation is complete, the values can be back transformed to original units multiplying each estimate by the standard deviation of the primary variable and adding the mean of the primary variable. Care must be taken to have reliable mean and standard deviation values. In the following variograms and covariances, Z and Y denote the standardized variables. The direct variograms are denoted:

$$\begin{aligned} \gamma_{ZZ}(\mathbf{h}) &= \frac{1}{2} E\left\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\right\} \\ \gamma_{YY}(\mathbf{h}) &= \frac{1}{2} E\left\{[Y(\mathbf{u}) - Y(\mathbf{u}+\mathbf{h})]^2\right\} \end{aligned} \tag{2}$$

The cross variogram and cross covariance are denoted:

$$\begin{aligned} \gamma_{ZY}(\mathbf{h}) &= \frac{1}{2} E\left\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]\,[Y(\mathbf{u}) - Y(\mathbf{u}+\mathbf{h})]\right\} \\ C_{ZY}(\mathbf{h}) &= E\left\{Z(\mathbf{u})Y(\mathbf{u}+\mathbf{h})\right\} - E\left\{Z(\mathbf{u})\right\} E\left\{Y(\mathbf{u}+\mathbf{h})\right\} \end{aligned} \tag{3}$$

Under an assumption of first and second-order stationarity and symmetric cross covariance, $C_{ZY}(\mathbf{h}) = C_{ZY}(-\mathbf{h})$, the cross covariance and cross variogram are related:

$$\gamma_{ZY}(\mathbf{h}) = C_{ZY}(\mathbf{0}) - C_{ZY}(\mathbf{h}) \tag{4}$$

This relationship is very important for heterotopic data because the cross variogram can only be calculated for homotopic data. In practice, the cross covariance is calculated and the cross

variogram is then derived from the cross covariance (Donovan, 2015). The cross covariance for a distance of zero ($C_{ZY}(\mathbf{0})$) is estimated by extrapolation. Then the cross variogram can be obtained from Equation 4. Cuba and Deutsch give a detailed explanation of this (Cuba & Deutsch, 2012). Figure 1 presents a schematic illustration. For the cross variograms, the presence of both variables (Z and Y) is required at both ends of the vector $\mathbf{h}$, while for the cross covariance, one variable is required at one end and the other variable at the other end. Notice that different data types could have different support, that is, drilling diameter and length of measurement.

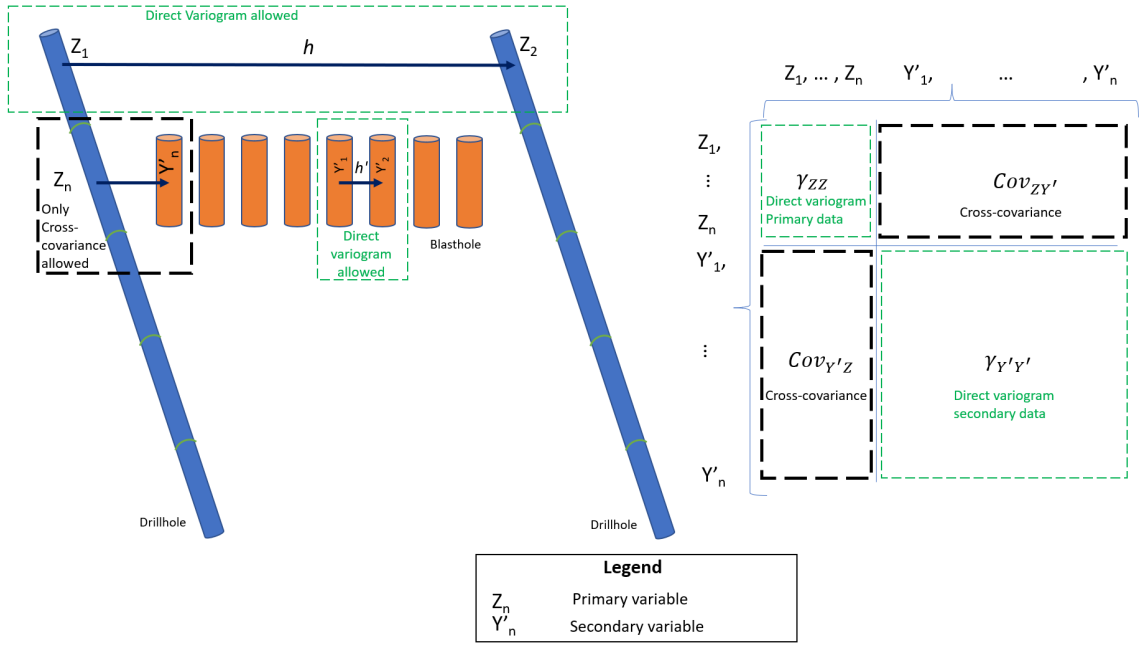


Figure 1: Left: Schematic cross-section with drill holes and some blast holes to show when the cross variogram can be calculated. Right: Schematic representation of relationships between variables to describe the spatial variability, the system of equations is shown later in the text.

The linear model of coregionalization is used to model the three variograms at the same time to ensure that the corresponding covariances are positive semidefinite (Pyrcz & Deutsch, 2014). The LMC has the following form:

$$\begin{aligned}
\gamma_{\mathrm{ZZ}}(\mathbf{h}) &= \sum_{l=0}^{nst} c_{\mathrm{ZZ}}^{l}\Gamma_l(\mathbf{h}) = c_{\mathrm{ZZ}}^{0} + c_{\mathrm{ZZ}}^{1}\Gamma_1(\mathbf{h}) + c_{\mathrm{ZZ}}^{2}\Gamma_2(\mathbf{h}) + \ldots \\
\gamma_{ZY}(\mathbf{h}) &= \sum_{l=0}^{nst} c_{ZY}^{l}\Gamma_l(\mathbf{h}) = c_{\mathrm{ZY}}^{0} + c_{\mathrm{ZY}}^{1}\Gamma_1(\mathbf{h}) + c_{\mathrm{ZY}}^{2}\Gamma_2(\mathbf{h}) + \ldots \\
\gamma_{YY}(\mathbf{h}) &= \sum_{l=0}^{nst} c_{YY}^{l}\Gamma_l(\mathbf{h}) = c_{\mathrm{YY}}^{0} + c_{\mathrm{YY}}^{1}\Gamma_1(\mathbf{h}) + c_{\mathrm{YY}}^{2}\Gamma_2(\mathbf{h}) + \ldots
\end{aligned} \tag{5}$$

Here the expanded expressions apply for $\mathbf{h} \neq \mathbf{0}$; the nugget structure and all variograms are zero at $\mathbf{h} = \mathbf{0}$.

Where: nst is the number of structures; c_{ZZ}^{l} are the contributions for the l structures; and, $\Gamma_l(\mathbf{h})$ are the variograms of the l nested structures. The contribution coefficients (c_{ZZ}^{l}) could be different for

each model as long they meet the following conditions to ensure positive semidefinite contribution matrices and a valid variogram model (Pyrcz & Deutsch, 2014):

$$\begin{aligned} & c^l_{\mathrm{ZZ}} \geq 0 \\ & c^l_{\mathrm{YY}} \geq 0 \qquad\qquad \forall\, l \\ & c^l_{\mathrm{ZZ}} \cdot c^l_{\mathrm{YY}} \geq c^l_{\mathrm{ZY}} \cdot c^l_{\mathrm{ZY}} \end{aligned} \tag{6}$$

Now, using Z and Y in their original units, the standardized ordinary cokriging estimator can be written as (Minnitt & Deutsch, 2014):

$$\frac{Z^*_{SOCK}\left(\mathbf{u}_0\right) - m_Z}{\sigma_Z} = \sum_{\alpha=1}^{n_Z} \lambda_{\alpha(\mathbf{u}_0)} \left[\frac{Z\left(\mathbf{u}_\alpha\right) - m_Z}{\sigma_Z}\right] + \sum_{\beta=1}^{n_Y} \lambda'_{\beta(\mathbf{u}_0)} \left[\frac{Y\left(\mathbf{u}'_\beta\right) - m_Y}{\sigma_Y}\right] \tag{7}$$

Where $\lambda_{\alpha(\mathbf{u}_0)}$ and $\lambda'_{\beta(\mathbf{u}_0)}$ are the cokriging weights for the primary variable and the secondary variable, respectively at location $\mathbf{u}_0$; m_Z and m_Y are the stationary means of the original, unstandardized variables; n_Z and n_Y are the number of samples inside the ranges of influence of the variogram (or covariance) for the primary and secondary variable, respectively. There is only one condition in order to get an estimator that is unbiased (Deutsch & Journel, 1998; Isaaks & Srivastava, 1989):

$$\sum_{\alpha=1}^{n_Z} \lambda_{\alpha(\mathbf{u}_0)} + \sum_{\beta=1}^{n_Y} \lambda'_{\beta(\mathbf{u}_0)} = 1 \tag{8}$$

Some historical implementations of ordinary cokriging considered the condition that the sum of the weights to the secondary data should be zero. This is a severe constraint that undermines the improvement brought by cokriging (Deutsch & Journel, 1998; Minnitt & Deutsch, 2014). A single constraint ensures that the secondary data are weighted appropriately (Goovaerts, 1998). Minimizing the mean squared error (MSE) generates the cokriging system of equations. Since there is only one constraint (see Equation 8) there is one Lagrange parameter. The system of equations:

$$\begin{aligned} & \sum_{i=1}^{n_Z} \lambda_{i(\mathbf{u}_0)} C_{ZZ}\left(\mathbf{u}_\alpha - \mathbf{u}_i\right) + \sum_{j=1}^{n_Y} \lambda'_{j(\mathbf{u}_0)} C_{ZY}\left(\mathbf{u}_\alpha - \mathbf{u}'_j\right) + \mu_{(\mathbf{u}_0)} = C_{ZZ}\left(\mathbf{u}_\alpha - \mathbf{u}_0\right) \\ & \text{where } \alpha = 1, \ldots, n_Z \\ & \sum_{i=1}^{n_Z} \lambda_{i(\mathbf{u}_0)} C_{YZ}\left(\mathbf{u}'_\beta - \mathbf{u}_i\right) + \sum_{j=1}^{n_Y} \lambda'_{j(\mathbf{u}_0)} C_{YY}\left(\mathbf{u}'_\beta - \mathbf{u}'_j\right) + \mu_{(\mathbf{u}_0)} = C_{YZ}\left(\mathbf{u}'_\beta - \mathbf{u}_0\right) \\ & \text{where } \beta = 1, \ldots, n_Y \\ & \sum_{\alpha=1}^{n_Z} \lambda_{\alpha(\mathbf{u}_0)} + \sum_{\beta=1}^{n_Y} \lambda'_{\beta(\mathbf{u}_0)} = 1 \end{aligned} \tag{9}$$

This can be expressed in matrix form as:

$$\begin{bmatrix} C_{ZZ}(\mathbf{u}_1-\mathbf{u}_1) & \dots & C_{ZZ}(\mathbf{u}_1-\mathbf{u}_{n_Z}) & C_{ZY}(\mathbf{u}_1-\mathbf{u}'_1) & \dots & C_{ZY}(\mathbf{u}_1-\mathbf{u}'_{n_Y}) & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ C_{ZZ}(\mathbf{u}_{n_Z}-\mathbf{u}_1) & \dots & C_{ZZ}(\mathbf{u}_{n_Z}-\mathbf{u}_{n_Z}) & C_{ZY}(\mathbf{u}_{n_Z}-\mathbf{u}'_1) & \dots & C_{ZY}(\mathbf{u}_{n_Z}-\mathbf{u}'_{n_Y}) & 1 \\ C_{YZ}(\mathbf{u}'_1-\mathbf{u}_1) & \dots & C_{YZ}(\mathbf{u}'_1-\mathbf{u}_{n_Z}) & C_{YY}(\mathbf{u}'_1-\mathbf{u}'_1) & \dots & C_{YY}(\mathbf{u}'_1-\mathbf{u}'_{n_Y}) & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \\ C_{YZ}(\mathbf{u}'_{n_Y}-\mathbf{u}_1) & \dots & C_{YZ}(\mathbf{u}'_{n_Y}-\mathbf{u}_{n_Z}) & C_{YY}(\mathbf{u}'_{n_Y}-\mathbf{u}'_1) & \dots & C_{YY}(\mathbf{u}'_{n_Y}-\mathbf{u}'_{n_Y}) & 1 \\ 1 & \dots & 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_{n_Z} \\ \lambda'_1 \\ \vdots \\ \lambda'_{n_Y} \\ \mu \end{bmatrix}$$

$$= \begin{bmatrix} C_{ZZ}(\mathbf{u}_1-\mathbf{u}_0) \\ \vdots \\ C_{ZZ}(\mathbf{u}_{n_Z}-\mathbf{u}_0) \\ C_{YZ}(\mathbf{u}'_1-\mathbf{u}_0) \\ \vdots \\ C_{YZ}(\mathbf{u}'_{n_Y}-\mathbf{u}_0) \\ 1 \end{bmatrix} \tag{10}$$

Since weights are calculated for each location being estimated, a consistent notation would be $\lambda_1(\mathbf{u}_0)$ instead of λ_1 (as in Equation 9), nevertheless, for the matrix form, the latter is being used because of space.

The cokriging estimator of Equation 7 will account for bias through the different mean values and account for error and different support through the covariance values. This estimator could be used directly in short- or medium-term modeling. In the case of simulation, the use of simple cokriging would be recommended to be consistent with the multivariate Gaussian distribution. This Lesson is aimed at estimation.

Implementation

The first step is to perform an exploratory data analysis to recognize the different characteristics of the data. Care should be taken if the information comes from different laboratories. The data are commonly processed separately within different domains. Basic statistics and visualization of the data are minimum steps. The next step is to standardize both data types by subtracting the mean and dividing by the standard deviation for each value. There will be different mean and standard deviation values for each data type.

The principal directions of continuity (major, minor and tertiary) are identified and defined. Experimental variograms of the drill holes and blast holes are calculated with different distance and direction tolerances since the data are spaced differently. The cross-covariance can be calculated with its own tolerance parameters, however, the value at $\mathbf{h} = 0$ ($C(0)$) will be inferred from the experimental model of the covariance. The cross variogram is then inferred.

The three experimental variograms are modeled at the same time with the LMC. The same models and structures are used with different contributions. The practitioner can consider the direct variogram of drill holes as first priority, then the cross variogram and finally the direct variogram of the blast holes. Once the variograms have been modeled, the cokriging program can be used. It is useful to perform ordinary kriging with the primary data only to ensure that the cokriging is performing well, the results of cross-validation can be assessed.

Example

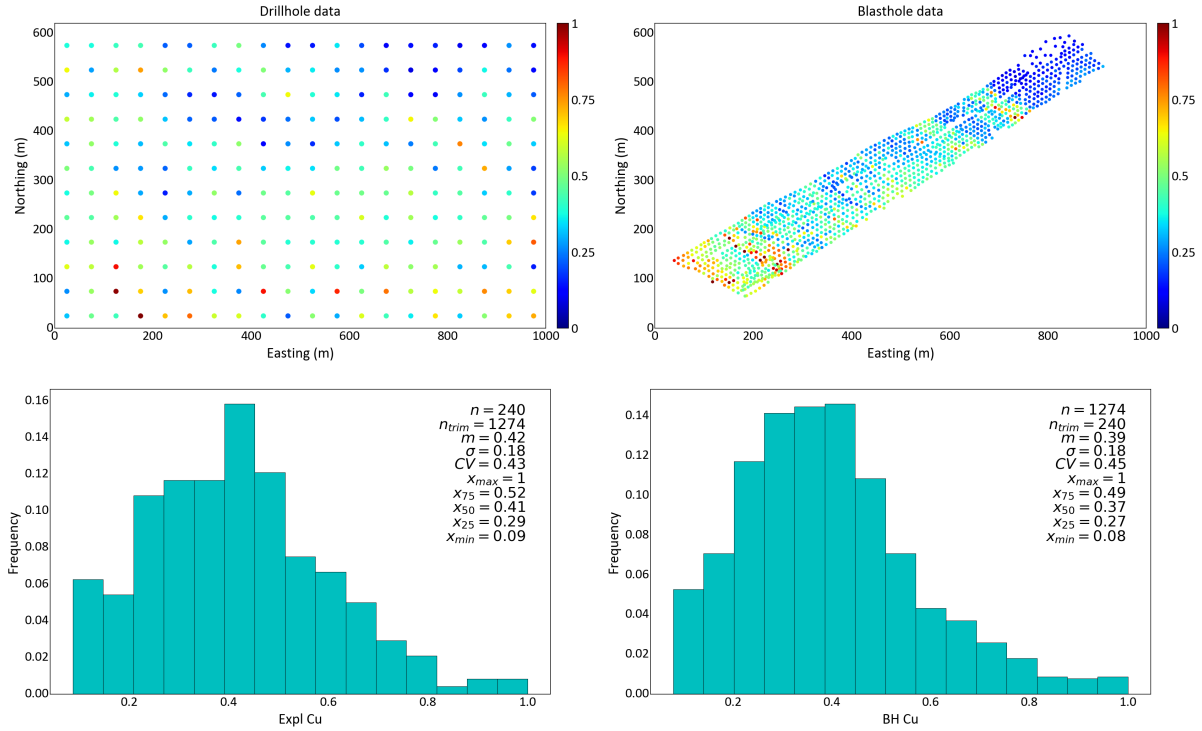


Figure 2: Plan view of drill hole and blast hole data and its histograms in original units.

A set of copper values from exploration data (drill holes) and production data (blast holes) are considered. The location of the data and summary statistics and histograms are presented in Figure 2. As a detail, the file to be used with the GSLIB program ‘cokb3d’ should contain both data types, one column with exploration data and another column with production data, this implies that half of the entries in these two data columns will be missing values. Then, it is necessary to standardize both data types.

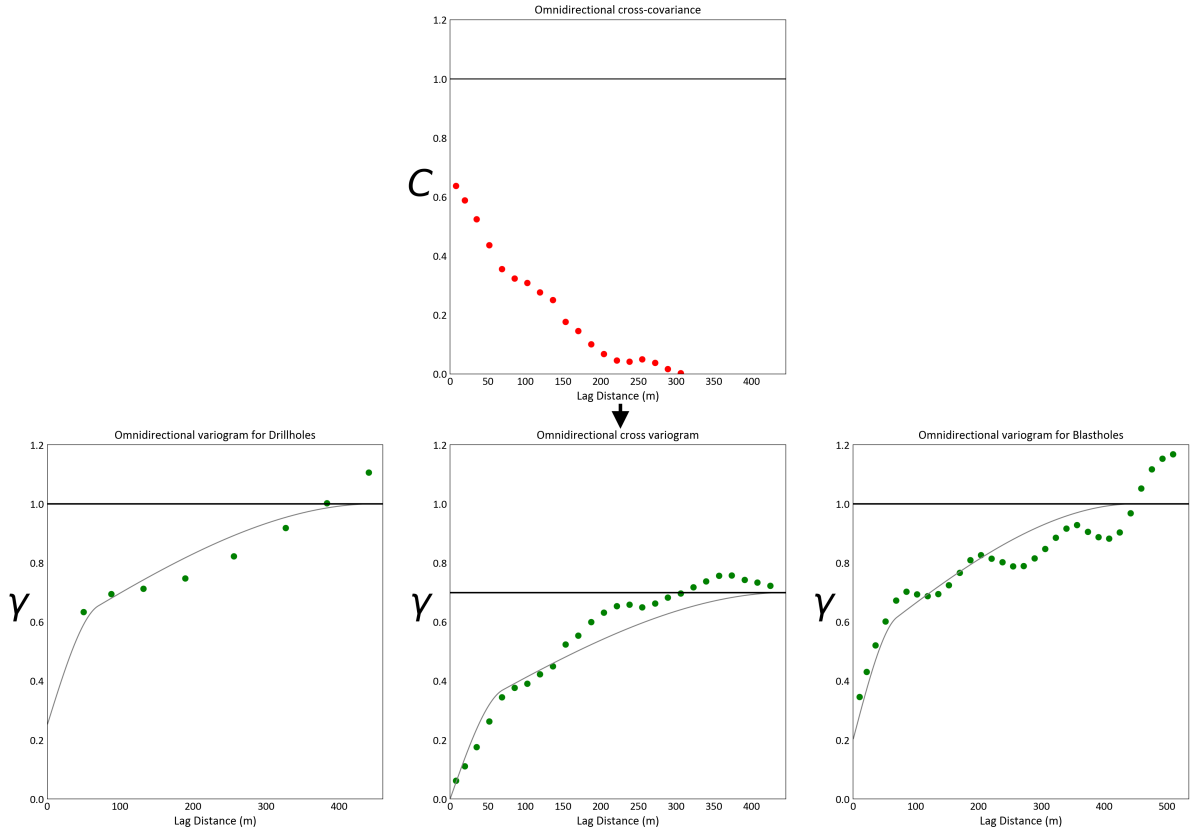


Figure 3: Experimental cross covariance calculation to get the experimental cross variogram.

The next step is to find the major directions of continuity. In this example an omnidirectional variogram is used, although anisotropy can be recognized in blast hole data (major direction: N 120, minor direction N 30); however, it cannot be clearly recognized in the drill hole data. Figure 3 shows the calculation of the experimental variograms and its models. The cross-covariance was calculated first and used to derive the cross-variogram. The three experimental variograms are modeled at the same time, only the contributions of each structure for every variogram model can be different (see Equation 5). The model is checked to ensure it is positive semidefinite (see Equation 6). A spherical model is used, the LMC is shown:

$$
\begin{aligned}
\gamma_{ZZ}(\mathbf{h}) &= 0.25\ \text{Nugget} + 0.30\ \text{Sph}_{a=70}(\mathbf{h}) + 0.45\ \text{Sph}_{a=450}(\mathbf{h}) \\
\gamma_{YZ}(\mathbf{h}) &= 0.00\ \text{Nugget} + 0.27\ \text{Sph}_{a=70}(\mathbf{h}) + 0.43\ \text{Sph}_{a=450}(\mathbf{h}) \\
\gamma_{YY}(\mathbf{h}) &= 0.20\ \text{Nugget} + 0.30\ \text{Sph}_{a=70}(\mathbf{h}) + 0.50\ \text{Sph}_{a=450}(\mathbf{h})
\end{aligned}
$$

Ordinary kriging and standardized ordinary cokriging are performed to obtain estimates in original units. Cross validation results are shown. The fit for kriging is reasonably good; however, the fit for cokriging is better. Maps of both estimates are displayed with subtle, yet important, local changes in the area of the blast holes that shows a better definition of grades.

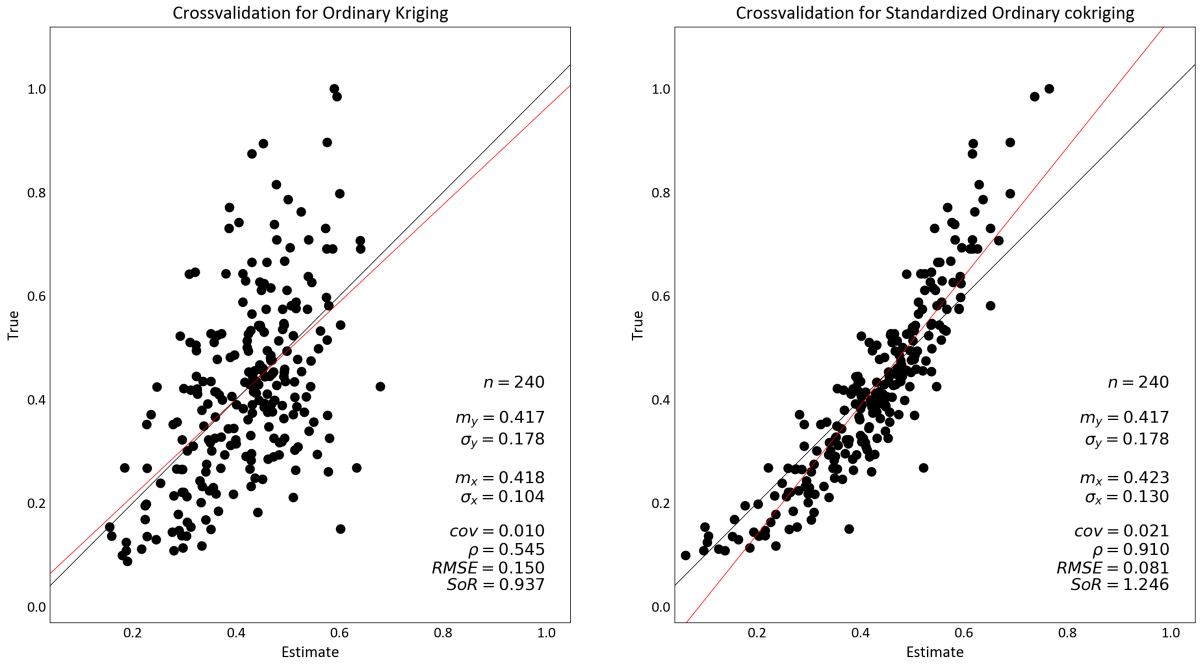


Figure 4: Cross validation between estimates of ordinary kriging (left) and standardized ordinary cokriging (right) compared with the true values of drill hole data.

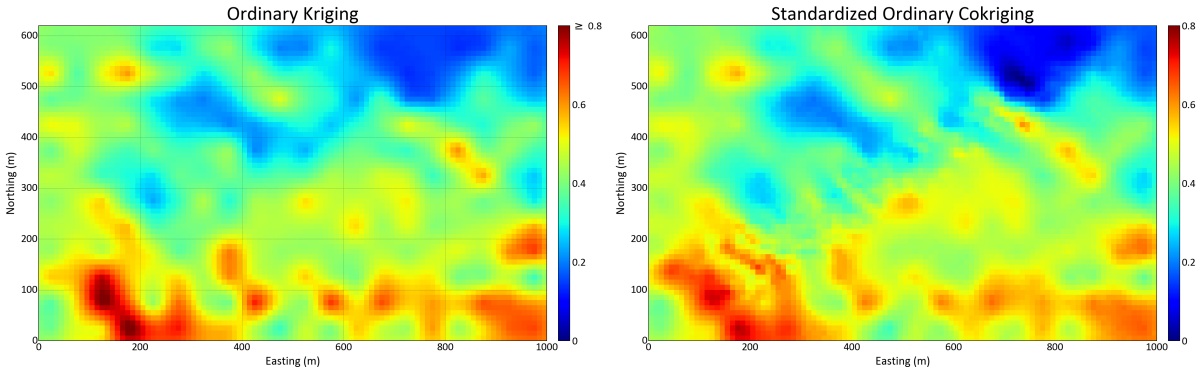


Figure 5: Plan view of the results of estimation of ordinary kriging and standardized ordinary cokriging estimates.

Discussion

Theoretically, with a correct covariance model and the same primary data, cokriging has a mean squared error no greater than kriging; the improvement need not be strict. This is shown in the example. The good results depend on reasonable variogram fitting and a reasonable assessment of the blast hole data. The main application of the methodology presented, in mining geostatistics applications, will be short- and medium term modeling. There are similar applications in environmental geostatistics with data of different quality. An important outstanding topic is simulation and the assessment of uncertainty with multiple data types. This is an evident extension of the theory and example presented here, but not developed in this lesson. Another related topic

is the possibility of using data imputation instead of cokriging, that is, impute the true values given collocated measured values with error. This imputation approach would be suited to simulation.

Summary

Standardized ordinary cokriging (SOCK) is a useful tool to apply in mining with higher quality exploration data and short term production data. The cross variogram of heterotopic data cannot be directly obtained; however, the cross covariance can be used as an estimator. The LMC must be positive semidefinite and all variograms are fitted simultaneously. In this example, the final cokriging outperforms ordinary kriging.

References

Barnett, R. M., & Deutsch, C. V. (2017). Projection pursuit multivariate transform. Retrieved from https://geostatisticslessons.com/lessons/ppmt

Cuba, M. A., & Deutsch, C. V. (2012). *Cross variography calculation of exploratory, infill, and blasthole drilling* (CCG Annual Report 14, Paper No. 311). Centre for Computational Geostatistics, University of Alberta. Retrieved from https://www.ccgalberta.com/ccgresources/report14/2012-311_cross-variography.pdf

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.

Donovan, P. N. (2015). *Resource estimation with multiple data types* (Master's thesis). University of Alberta. Retrieved from https://www.ccgalberta.com/ccgresources/theses/2015_-_donovan_-_unequal_sampling.pdf

Goovaerts, P. (1998). Ordinary cokriging revisited. *Mathematical Geology*, *30*(1), 21--42.

Isaaks, E. H., & Srivastava, M. R. (1989). *An introduction to applied geostatistics*. New York: Oxford University Press.

Minnitt, R. C. A., & Deutsch, C. V. (2014). Cokriging for optimal mineral resource estimates in mining operations. *Journal of the Southern African Institute of Mining and Metallurgy*, *114*(3), 189--203.

Pyrcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. Oxford university press.

Rossi, M. E., & Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.

Cite this lesson

Davila, L., & Deutsch, C. V. (2022). Cokriging with Unequally Sampled Data. In J. L. Deutsch (Ed.), *Geostatistics Lessons*. Retrieved from https://geostatisticslessons.com/lessons/cokrigingunequal