Sequential co-simulation for resources modelling: How good is the reproduction of the joint spatial correlation structure?

N. Madani\textsuperscript{1} and X. Emery\textsuperscript{2}

\textsuperscript{1}School of Mining and Geosciences, Kazakhstan
\textsuperscript{2}University of Chile, Chile

The quantification and classification of mineral resources and ore reserves is often based on an assessment of the certainty (or uncertainty) of the estimates of tonnage and grade of elements of interest (metals and impurities) contained in a mineable deposit. However, these variables are not always sufficient to characterize the factors that impact mining and mineral processing performance and hence the estimated quantity and value of recovered product. Improved understanding of geometallurgical response variables, via spatial simulation, offers one approach to predicting concentration and recovery performance, identifying improved operating parameters, and optimizing criteria that impact on costs of recovery, such as energy and reagent consumption. In this context, the spatial modelling of these responses is challenging. Geostatistical co-simulation techniques can be used to construct high-resolution models of geometallurgical responses that reproduce both the spatial variability and the multivariate relationships between co-regionalized variables. Sequential Gaussian co-simulation is such a technique, in which the variables of interest are simulated hierarchically, using cokriging to determine the distributions of values to be randomly simulated. However, because full cokriging is generally out of reach when the model contains too many locations or blocks to simulate, simplifications based on strictly collocated or multi-collocated cokriging are often adopted. In this study, the quality of such simplifications is investigated, with an application to a porphyry copper deposit. Sampling data from two cross-correlated geometallurgical variables, the iron grade and the concentrate copper grade, have been selected and sequentially co-simulated using the abovementioned cokriging strategies. The results show that both collocated and multi-collocated cokriging succeed in reproducing the spatial correlation structure of the variables to be simulated.

INTRODUCTION

The joint spatial simulation of cross-correlated variables is widely used to assess local uncertainty of geometallurgical response variables. The quantification of their uncertainty using many realizations allows for significant improvement in further analysis of downstream activities of a mining project e.g., mining rates, fragmentation (Boisvert \textit{et al.}, 2013; Rossi and Deutsch, 2014; Emery 2012; Hosseini and Asghari, 2015; Deutsch, 2013). The method requires approximate algorithms to be employed based on the Gaussian assumption of the underlying random fields. For instance, sequential Gaussian co-simulation (Verly 1993; Deutsch and Journel 1998) and turning bands co-simulation (Emery, 2008) have received wide acceptance among practitioners. Paravarzar, Emery, and Madani (2015) showed that the latter outperforms the former in terms of reproduction of the first- and second-order moments. These algorithms are based on cokriging to produce simulations conditioned to existing data, in which a search neighbourhood can be simplified depending on the availability of conditioning information.
In this paper, the sequential Gaussian co-simulation paradigms through two corresponding simplified versions of cokriging are compared. Their validity is assessed through an application to the modelling of geometallurgical variables (concentrate copper and iron grades as the target and auxiliary variables, respectively) in a porphyry copper deposit. The two versions of cokriging correspond to collocated cokriging, only where the auxiliary data at the target grid node is considered, and multi-collocated cokriging that also incorporates the auxiliary data collocated with the target data (Chilès and Delfiner, 2012).

THEORY

Sequential Gaussian Co-simulation

Sequential Gaussian co-simulation (SGCOSIM) is a generalization of sequential Gaussian simulation under the multi-normality assumption aimed at reproducing the spatial cross-correlation of several variables (Gómez-Hernández and Journel, 1993; Verly, 1993; Almeida and Journel, 1994). This paradigm aims to directly simulate a vector random field \( Y(u) = \{Y_1(u), Y_2(u), \ldots, Y_k(u)\} \) in lieu of a scalar random field \( Y_1(u) \) for sequential Gaussian simulation. The main difference between these two algorithms is that the simple kriging used in the latter is substituted by simple cokriging in the former. Depending on the availability of data and size of the vector random field (in the present case with two components: target and auxiliary attributes), one is able to choose an appropriate search strategy for the implementation of cokriging. Among others, Almeida and Journel (1994) proposed a hierarchical approach to reduce the simple cokriging system to a collocated cokriging approximation. In this case, the auxiliary data must be available at each target location and it is assumed that this data screens out the influence of the auxiliary data located farther away (Journel, 1999). Alternatively, another modification of simple cokriging is the so-called multi-collocated cokriging, where the auxiliary data collocated with the data of the target variable is retained as well (Chilès and Delfiner, 2012). In the following subsection, we give a brief review on these two cokriging paradigms.

Collocated Cokriging

Provided that the target and auxiliary variables are represented by second-order stationary random fields, the collocated cokriging predictor and the variance of the prediction error (known as cokriging variance) for the target variable (hereafter denoted with index 1) given one auxiliary variable (denoted with index 2) are defined as (Xu et al., 1992; Almeida and Journel, 1994):

\[
Y_{cck}(u_0) = m_1 + \sum_{\alpha=1}^{n_1} \omega_\alpha Y_1(x_{1,\alpha}) - m_1 + \omega_0 Y_2(u_0) - m_2
\]

\[
\sigma_{cck}^2(u_0) = C_{11}(u_0 - u_0) - \sum_{\alpha=1}^{n_1} \omega_\alpha C_{11}(u_{1,\alpha} - u_0) - \omega_0 C_{21}(u_0 - u_0)
\]

where \( \omega_\alpha \) (\( i = 1, 2 \)) is the weight assigned to the data \( Y_i(u_{i,\alpha}) \) of the \( i - th \) variable \( Y_i \) at the \( \alpha - th \) data location \( x_{1,\alpha} \) \( (\alpha = 1, \ldots, n_1) \) of this variable, \( u_0 \) is the location targeted for prediction, \( m_i \) is the mean value of the \( i - th \) variable \( Y_i, C_{ij} \) is the direct \( (i = j) \) or cross \( (i \neq j) \) covariance between variables \( Y_i \) and \( Y_j \) \( (i, j = 1, 2) \). The previous equations can be generalized to the case with more than one auxiliary variable, which will not be considered in this work. The simple collocated cokriging system in order to calculate the weights \( \omega_\alpha \) required in Equations [1] and [2] is:

\[
\left( \sum_{\alpha=1}^{n_1} \omega_\alpha C_{11}(u_{1,\beta} - u_{1,\alpha}) + \omega_0 C_{12}(u_{1,\beta} - u_0) = C_{11}(u_{1,\beta} - u_0), \quad \beta = 1, \ldots, n_1 \right)
\]

\[
\left( \sum_{\alpha=1}^{n_1} \omega_\alpha C_{21}(u_{2,\beta} - u_{1,\alpha}) + \omega_0 C_{22}(u_{2,\beta} - u_0) = C_{21}(u_{2,\beta} - u_0) \right)
\]

where index 0 is used to numerate the target location \( u_{2,0} = u_0 \) and the weight \( \omega_\alpha \) assigned to the collocated auxiliary data \( Y_2(u_0) \).

Multi-collocated Cokriging

In multi-collocated cokriging, the retained auxiliary data is that available at the target location \( x_0 \) and at the locations of the data of the target variable \( x_{1,\alpha}, \alpha = 1, \ldots, n_1 \). In the case of a single auxiliary variable, the cokriging predictor and the error variance are given by (Wackernagel, 2003; Chilès and Delfiner, 2012):
Hierarchical Joint Simulation

The simulation of the two variables (target and auxiliary) is carried out as follows (Almeida and Journel, 1994; Goovaerts, 1997).

1. Transform the data of the original variables \( Y_i \) \((i = 1, 2)\) into normal score values \( Z_i \) with a zero mean and unit variance: this can be implemented by a Gaussian anamorphosis modelling (Rivoirard, 1994) or by a quantile-based approach (Deutsch and Journel, 1998):

\[
Z_i(x) = G^{-1}(F(Y_i(x))) \quad i = 1, 2
\]

where \( x \) is a sample location; \( G^{-1}(\cdot) \) is the standard normal inverse cumulative distribution function, \( F(\cdot) \) is the cumulative distribution function of the original variable \( Y_i(x) \), and \( Z_i(x) \) is the normal score value.

2. Define a random path visiting once each location (node) targeted for simulation. Multiple-grid sequences or midpoint sequences (Gómez-Hernández, 1991; Gómez-Hernández and Cassiraga, 1994; Tran, 1994) are often used to improve the reproduction of the long-range structure of the predefined spatial continuity model without the need for cokriging neighbourhoods that convey excessive conditioning sample information. These techniques consist in visiting the nodes of a coarse grid using a relatively large neighbourhood in order to reproduce the underlying long-range correlation structure, and then proceeding successively to the nodes of finer grids using a small search neighbourhood with the focus on reproducing the short-range correlation structure. Within each grid, a random path is followed.

3. Use simple cokriging to infer the parameters required for simulating the value of the auxiliary variable at a location \( u \) from its conditional cumulative distribution function (hereinafter, ccdf). To this end, for each node \( u_i \), \( i \in \{1, \ldots, k\} \), we put:

\[
Z_2^{(l)}(u_i) = Z_{2,SCS}^{*}(u_i) + \sigma_{2,SCS}(u_i)\Lambda_0^{(l)}, \quad l = 1, \ldots, L
\]

where \( Z_{2,SCS}^{*}(u_i) \) is the simple cokriging of \( Z_2^{(l)}(u_i) \) obtained by using the conditioning data and the previously simulated values of the auxiliary variable \( \{Z_2^{(l-1)}(u_i), \ldots, Z_2^{(l-1)}(u_{i-1})\} \), \( L \) is the number of realizations, \( \sigma_{2,SCS}(u_i) \) is the simple cokriging standard deviation, and \( \Lambda_0^{(l)} \) is a standard Gaussian random value (independent of \( \Lambda_0^{(l-1)} \)). It is worth mentioning that ordinary cokriging is not recommended as it results in a deterioration of the reproduction of the second-order moments (Deutsch and Journel, 1999; Emery, 2004).

4. Once the auxiliary variable has been simulated at all the grid nodes, the target variable is now simulated, starting from location \( u_1 \) repeating steps 3 and 4. But at this stage, simple cokriging is replaced by collocated or multi-collocated cokriging in Equation \([8]\):

\[
\begin{align*}
\{ & Z_2^{(l)}(u_i) = Z_{2,CCS}^{*}(u_i) + \sigma_{2,CCS}(u_i)\Lambda_0^{(l)}, \quad i \in \{2, \ldots, k\} \\
& Z_2^{(l)}(u_1) = Z_{2,MCSS}^{*}(u_1) + \sigma_{2,MCSS}(u_1)\Lambda_0^{(l)}, \quad l = 1, \ldots, L
\end{align*}
\]

with the same notations as in step 4. The definition of the cokriging neighbourhood in either case is different in terms of selection of conditioning information. In simply collocated
cokriging, the data participating in the process of cokriging consists of the values of the target variable (at both the original data locations and previously simulated nodes) and the auxiliary value at the target node $Z_2^{(i)}(u_i)$. Through multi-collocated cokriging, the neighbourhood is the same, except that the auxiliary values at the same locations as that of the target variable are incorporated in the cokriging for drawing the conditional cumulative distribution function. This process is looped until all the nodes are simulated.

5. Back-transform the realizations $Z_2^{(i)}(u)$ and $Z_1^{(i)}(u)$ to the scale of the original variables $Y_2^{(i)}(u)$ and $Y_1^{(i)}(u)$, respectively.

Coregionalization Modelling

Constructing the variance-covariance matrix to solve the cokriging systems (Equations [3] and [6]) requires the inference of the direct and cross-covariances or cross-variograms between the variables. The linear model of coregionalization (LMC) assumes that these covariances or variograms are linear combinations of elementary covariances or variograms (basic nested structures). In this model, the set of $p \times p$ direct and cross-variograms $\Gamma_{ij}(h)$ is defined as:

$$\Gamma_{ij}(h) = \sum_{s=1}^{S} b_{ij} g_s(h)$$

[10]

where, for each $s = 1 \ldots S$, $(b_{ij})_{i,j=1,p}$ is a $p \times p$ real-valued, symmetric, positive semi-definite matrix (coregionalization matrix) and $g_s(h)$ is a stationary variogram model. The positive definiteness of the coregionalization matrices ensures the mathematical consistency of the LMC (Journel and Huijbregts, 1978; Goovaerts, 1997; Wackernagel, 2003; Marcotte, 2012). In the case of two variables, namely target and auxiliary variables, the coregionalization matrix $(b_{ij})_{i,j=1,2}$ is positive semi-definite if the following inequalities are verified:

$$b_{11}^T \geq 0 \quad b_{22}^T \geq 0$$

$$b_{12}^T b_{22}^T - b_{12}^T b_{22}^T \geq 0 \quad \rightarrow \quad |b_{12}^T| \leq \sqrt{b_{11}^T b_{22}^T}$$

[11]

[12]

In practice, a LMC can be fitted to a set of experimental direct and cross-variograms with the recourse to semi-automated algorithms (Goulard and Voltz, 1992; Emery, 2010) that aim to minimize a weighted sum of the squared differences between experimental and modelled variograms. Note that there are other models for inferring the spatial correlation structure of co-regionalized variables, e.g., Markov models (Almeida and Journel, 1994; Journel, 1999), which are more restrictive than the linear model of coregionalization used in this work.

CASE STUDY

Presentation and Geostatistical Modelling of the Data-set

The data-set is composed of 710 samples obtained from blast-holes at a Chilean copper mine. Two geometallurgical variables, iron (Fe) and copper concentration (CuCo) grades, have been assayed at all the sample locations. In order to preserve the confidentiality of the data-set, the original variables have been multiplied by a constant scale factor.

For co-simulation, the following steps are necessary.

- The first step is to infer representative distributions of the underlying variables, in order to avoid any impact of a possible preferential sampling. Cell declustering (Deutsch and Journel, 1998) is such a technique that assigns weights to each data point based on a division of the region into sub-cells, so that isolated data locations receive more importance than clustered data locations (Goovaerts, 1997). Table 1 shows the statistical parameters of the data after declustering, which are suitable for comparing the conditional simulations characteristics and original input data (Rondon, 2012).
Table I. Declustered statistics of geometallurgical variables (Chilean copper deposit).

<table>
<thead>
<tr>
<th>Geometallurgical variable</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe (%)</td>
<td>1.468</td>
<td>1.687</td>
<td>0.016</td>
<td>10.784</td>
</tr>
<tr>
<td>CuCo (%)</td>
<td>25.113</td>
<td>10.351</td>
<td>17.483</td>
<td>32.805</td>
</tr>
</tbody>
</table>

- The second step is to transform the data into standard normal scores, i.e., data with a Gaussian distribution with mean 0 and variance 1. In this study, we follow the Gaussian anamorphosis methodology presented by Rivoirard (1994). Table II presents the correlation coefficients after (upper diagonal) and before (lower diagonal) the transformation into normal scores, respectively. The correlation coefficient increases approximately 14% after transformation.

Table II. Correlation coefficients between Fe and CuCo before (lower diagonal) and after (upper diagonal) normal score transformation.

<table>
<thead>
<tr>
<th></th>
<th>Fe (%)</th>
<th>CuCo (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fe (%)</td>
<td>1</td>
<td>-0.807</td>
</tr>
<tr>
<td>CuCo</td>
<td>-0.713</td>
<td>1</td>
</tr>
</tbody>
</table>

- The separate transformation of each variable into normal scores does not guarantee that the transformed variables are jointly Gaussian. Accordingly, the third step is to examine such a hypothesis through the scatter plot between the transformed variables; this plot (Figure 1) has an elliptical shape, in agreement with the bivariate Gaussian assumption (Rivoirard, 1994). The check can be extended to the pairs of data of the same variable at different lags, which in practice can be tedious. Instead, a convenient check is to compare the experimental variograms against the experimental variograms of lower orders, namely madogram and rodogram. As for the scatter plot, this check is satisfactory, as the variograms follow the theoretical relationships that can be established under a bivariate Gaussian assumption (Emery, 2005) (Figure 2). In case that the bivariate normality hypothesis does not hold, a multivariate transformation (instead of a separate transformation of each variable) can be used in the previous step to convert the original data into normal scores, such as the projection pursuit multivariate transform (PPMT) (Barnett, Manchuk, and Deutsch, 2013) or the flow anamorphosis (van den Boogaart, Mueller, and Tolosana-Delgado, 2017).
The fourth step is to perform the variogram analysis of the normal score variables in order to model their spatial joint correlation structure. Omnidirectional experimental direct and cross-variograms are calculated, insofar as no clear anisotropy is detected in the data, then a linear model of coregionalization with two nested spherical structures is used to fit these variograms (Figure 3). The main diagonals of the coregionalization matrices give the fitted sills of the direct variograms and the off-diagonal terms give the fitted sill of the cross-variogram.

\[
\begin{pmatrix}
Y_{Fe} & Y_{Fe/CuCo} \\
Y_{CuCo/Fe} & Y_{CuCo}
\end{pmatrix} = \begin{pmatrix}
0.184 & -0.059 \\
-0.059 & 0.270
\end{pmatrix} Sph(25m) + \begin{pmatrix}
0.990 & -0.888 \\
-0.888 & 0.835
\end{pmatrix} Sph(450m)
\]  

[13]

The fitted coregionalization matrices are symmetric and positive semi-definite.
Sequential Gaussian Co-simulation of CuCo and Fe

Once the theoretical direct and cross-variograms are derived, one is able to sequentially simulate the two variables. The first step is the simulation of the auxiliary variable (Fe) on a two-dimensional grid (representing a bench to be mined) with 125 × 30 nodes with a mesh size of 20 m × 20 m that covers the lateral extents of the region. The second step is then to simulate the target variable (CuCo), considering the information available on the grid nodes so obtained from the simulation of Fe. At this step, the two modified versions of cokriging discussed previously, collocated and multi-collocated cokriging, are applied, with a moving neighbourhood of horizontal radius 1000 m considering up to a maximum of 200 data points. Sequential co-simulation also requires the previously simulated nodes to be introduced in the neighbourhood of interest, which is set to 200 as well. Fifty realizations are generated. In both cases, a multiple-grid strategy with three nested grids is used, together with a random sequence to visit the nodes of each grid. The maps of the first realization for Fe and CuCo are illustrated in Figure 4 for both cases of co-simulation based on collocated and multi-collocated cokriging. The E-type map (average of the 50 realizations) is also calculated (Figure 5). The realizations reproduce the variability of Fe and CuCo at all spatial scales for a quasi-point support corresponding to the support of a blast-hole. They can be averaged to larger supports in order to emulate the blending of material to be sent to the processing plant, which allows studying the joint behaviour of Fe and CuCo at the processing plant scale, and not only at the laboratory (blast-hole data) scale.

(a) Direct variogram of Fe  
(b) Direct variogram of CuCo  
(c) Cross-variogram of Fe/CuCo

Figure 3. Direct and cross-variograms of the transformed (normal score) variables; crosses: experimental variograms; solid lines: theoretical fitted variograms.
Algorithm Validation
Checking of the realizations is necessary before decision-making (Leuangthong, McLennan, and Deutsch, 2004). The first check is to assess the reproduction of the marginal distributions (histograms) of the simulated variables, either in the original (back-transformed) or in the Gaussian scale. Figure 5 shows the normal probability plots of CuCo in the first realization in the cases of co-simulation based on collocated and multi-collocated cokriging. In general, except for the extreme lower tail (corresponding to the first centile of the distribution), the reproduction of the normal distribution is satisfactory.
A second check consists of comparing the theoretical correlation coefficient between Fe and CuCo, which can be obtained from the linear model of coregionalization, against the average experimental correlation coefficient derived from the realizations obtained with the collocated and multi-collocated cokriging approaches, respectively. Overall, the reproduction of correlation is acceptable for both cases, although the multi-collocated-based one is closer to the target correlation coefficient (Table III).

Table III. Correlation coefficients from 50 realizations obtained in both cases.

<table>
<thead>
<tr>
<th>Theoretical</th>
<th>Multi-collocated cokriging</th>
<th>Collocated cokriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.8294</td>
<td>-0.7366</td>
<td>-0.7239</td>
</tr>
</tbody>
</table>

Another check is to examine the reproduction of the spatial auto- and cross-correlation of the simulated variables before back-transformation. In this context, the madogram and rodogram (variograms of order 1 and 0.5, respectively) are computed over the realizations and compared with the corresponding theoretical model for measuring the spatial continuity of each variable. Concerning the joint spatial continuity, the cross-variograms are calculated and compared with the theoretical model (Figure 6). With both collocated and multi-collocated cokriging, the reproduction of the spatial continuity is good: on average over the realizations, the madogram, rodogram and cross-variograms are well reproduced.
Figure 6. Spatial continuity reproduction in terms of rodogram, madogram, and cross-variogram; for brevity, only the graphs of CuCo are displayed. Left: collocated cokriging; right: multi-collocated cokriging. Green: individual curves obtained from each realization; red: average curves over 50 realizations; blue: theoretical model.
CONCLUSIONS

The aim of this paper is to compare two implementations of sequential co-simulation for constructing block models that mimic the spatial distribution of geometallurgical variables. The case study shows that co-simulation built on either collocated or multi-collocated cokriging successfully reproduces the spatial correlation structure of the variables of interest, with a slight improvement for the multi-collocated implementation. This success can be explained by the large cokriging neighbourhood used to determine the successive conditional distributions (200 conditioning data points plus 200 previously simulated nodes) and by the particular coregionalization model of the geometallurgical variables, and may not be observed with other data or other neighbourhood sizes. The methodology proposed in this work is therefore of interest to validate the quality of the simulation.

ACKNOWLEDGEMENTS

The first author is grateful to Faculty Development Competitive Research Grants for 2018-2020 under contract no. 090118FD5336. The second author acknowledges funding of the Chilean Commission for Scientific and Technological Research (CONICYT), through grant CONICYT PIA Anillo ACT1407.

REFERENCES


Dr Madani received a PhD in Mining Engineering from University of Chile, Santiago, Chile. Currently, he is an Assistant Professor at Nazarbayev University, where he teaches and conducts research on Geostatistics (linear, nonlinear, and multivariate). Prior to this position, Dr Madani was an Assistant Researcher at the Advanced Mining Technology Centre (AMTC), University of Chile, from 2013 to 2016, involved in modelling and evaluation of orebodies. He has consulted to the mining/petroleum companies and conducted research projects in geostatistical modelling, together with providing a practical knowledge for interpreting data from sampling, experiments, and industrial tests.