Simulation of Multiple Variables with Combined SGS/LU for Correlation Matrix Reproduction

Olena Babak, David F. Machuca-Mory and Clayton V. Deutsch

Multivariate simulation is a longstanding problem in geostatistics. Fitting a model of coregionalization to many variables is intractable; however, the matrix of collocated correlation coefficients is often well informed. Performing a matrix simulation with LU decomposition of the correlation matrix at each step of sequential simulation is implemented in some software. The target correlation matrix is not reproduced because of conditioning to local data. A correction procedure is developed to calculate a modified correlation matrix that leads to reproduction of the target correlation matrix. The theoretical and practical aspects of this correction are developed.

Introduction

When dealing with several attributes in spatial modeling it is desirable to reproduce the collocated correlation among them. Reproduction of the correlation between multiple variables can be achieved by using a variants of cokriging (Deutsch, 2002; Wackernagel, 2003). Sequential Gaussian Simulation (SGS) can be extended to simultaneous modeling of several random variables. Consider modeling N variables.

Sequential Gaussian Simulation with Simple Cokriging can be applied to multivariate simulation (Journel and Huijbregts, 1978; Chiles and Delfiner, 1999). A significant problem with using a full model of coregionalizaiton is fitting the $N(N+1)/2$ direct and cross variograms; this becomes intractable with more than four variables. Another possibility would be to apply Sequential Simulation with either Collocated Cokriging (Xu, *et al*., 1992; Almeida and Journel, 1994) or Intrinsic Collocated Cokriging (Babak and Deutsch, 2008). These algorithms, however, are designed for a single secondary variable.

Multivariate sequential Gaussian simulation with correlated residuals (Davis, 1987) is another option; it is based on independent modeling of each random variable. The data related to each variable is used to calculate the mean and variance of the local conditional distribution for that variable using simple kriging; no inference of the joint model for spatial continuity is needed. Then, the drawing from the N conditional distributions is performed with correlation. This has many desirable features including reproduction of the target mean and variance as well as target variogram models by simulated realizations. Although the correlation matrix between the random variables is used for the residuals, it is not reproduced by the final simulated values. The influence of conditioning causes the lack of conditioning.

A novel approach for finding the "correct" correlation matrix for residuals is developed. The correct or modified correlation matrix leads to reproduction of the correct correlation matrix between the variables of interest. The proposed correction scheme is applied in several examples where significant improvement in the reproduction of correlation matrices over conventional method is shown.

Recall of SGS with Correlated Residuals

SGS with correlated residuals proceeds in normal scores or Gaussian units. Locations are visited in a random order to avoid artifacts. At each location, simple kriging is performed to find the mean $m_{SK,i}(\mathbf{u})$ and variance $\sigma^2_{SK,i}(\mathbf{u})$ of the local conditional distributions for all N variables. a vector of correlated standard normal residuals $\mathbf{R}(\mathbf{u})^T = [R_1(\mathbf{u}),\ldots,R_N(\mathbf{u})]^T$, such that

$$\rho(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = \rho_{ij}^{\text{target}}, \quad \forall i, j = 1,\ldots,N, \quad \forall k \tag{1}$$

where $\boldsymbol{\rho}^{\text{target}}$ is the target global correlation matrix between random variables. Note that the residuals $\mathbf{R}(\mathbf{u})^T = [R_1(\mathbf{u}),\ldots,R_N(\mathbf{u})]^T$ are drawn independently from location to location, that is,

$$\rho(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = 0, \quad \forall i, j = 1,\ldots,N, \quad \forall k \neq l; \tag{2}$$

The vector of simulated values $\mathbf{Z}(\mathbf{u})^T = [Z_1(\mathbf{u}),\ldots,Z_N(\mathbf{u})]^T$ for each of the N random variables at the unsampled location $\mathbf{u}$ as

$$\mathbf{Z}(\mathbf{u}) = \mathbf{m}_{SK}(\mathbf{u}) + \boldsymbol{\sigma}_{SK}(\mathbf{u})\mathbf{R}(\mathbf{u}) \tag{3}$$

where $\mathbf{m}_{SK}(\mathbf{u})^T = [m_{SK,1}(\mathbf{u}),\ldots,m_{SK,N}(\mathbf{u})]^T$ is the vector of Simple Kriging means; matrix $\boldsymbol{\sigma}_{SK}(\mathbf{u})$ is given by

$$\boldsymbol{\sigma}_{SK}(\mathbf{u}) = \begin{bmatrix} \sigma_{SK,1}(\mathbf{u}) & 0 & \cdots & 0 \\ 0 & \sigma_{SK,2}(\mathbf{u}) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma_{SK,N}(\mathbf{u}) \end{bmatrix}. \tag{4}$$

where values on the diagonal are equal to the square root of the Simple Kriging variances $\sigma^2_{SK,i}(\mathbf{u})$ obtained for N variables. The simulated values are added as data for the simulation of subsequent nodes. Multiple equally-probable realizations can be created by changing the random number seed. The shape of the conditional distributions is Gaussian ensuring that the simulated realizations for all variables will be standard normal. The variogram reproduction is ensured by using not original data and previously simulated nodes.

Problem of Sequential Gaussian Simulation with Correlated Residuals

An unfortunate feature of multivariate SGS with correlated residuals is that the target correlation matrix between multiple random variables is not reproduced. Thus, results may have a systematic bias. To show this, let us calculate the correlation-covariance between two simulated values for variables Z_i and Z_j, respectively, at arbitrary location $\mathbf{u}$ in the study domain, $i, j \in \{1,\ldots,N\}$. These simulated values $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$ in sequential Gaussian simulation are given by

$$\begin{aligned} Z_i(\mathbf{u}) &= m_{SK.i}(\mathbf{u}) + \sigma_{SK,i}(\mathbf{u})R_i(\mathbf{u}), \\ Z_j(\mathbf{u}) &= m_{SK,j}(\mathbf{u}) + \sigma_{SK,j}(\mathbf{u})R_j(\mathbf{u}), \end{aligned} \tag{5}$$

thus, due to independence between residual vector $[R_i(\mathbf{u})\ R_j(\mathbf{u})]^T$ and vector of Simple Kriging means $[m_{SK,i}(\mathbf{u})\ m_{SK,j}(\mathbf{u})]^T$,

$$\begin{aligned} \rho(Z_i(\mathbf{u}),Z_j(\mathbf{u})) &= Cov(Z_i(\mathbf{u}),Z_j(\mathbf{u})) \\ &= Cov(m_{SK,i}(\mathbf{u}),m_{SK,j}(\mathbf{u})) + \sigma_{SK,i}(\mathbf{u})\sigma_{SK,j}(\mathbf{u})Cov(R_i(\mathbf{u}),R_j(\mathbf{u})) \\ &= Cov(m_{SK,i}(\mathbf{u}),m_{SK,j}(\mathbf{u})) + \sigma_{SK,i}(\mathbf{u})\sigma_{SK,j}(\mathbf{u})\rho_{ij}^{\text{target}}. \end{aligned} \tag{6}$$

Therefore, the global correlation between variables Z_i and Z_j is equal to

$$\overline{\rho(Z_i(\mathbf{u}),Z_j(\mathbf{u}))} = \overline{Cov(m_{SK,i}(\mathbf{u}),m_{SK,j}(\mathbf{u}))} + \overline{\sigma_{SK,i}(\mathbf{u})\sigma_{SK,j}(\mathbf{u})}\rho_{ij}^{\text{target}}. \tag{7}$$

Since the Simple Kriging mean values, $m_{SK,i}(\mathbf{u})$ and $m_{SK,i}(\mathbf{u})$ at any location of the study domain $\mathbf{u}$ are calculated based on the conditioning data and previously simulated nodes, the covariance between variables Z_i and Z_j at any location $\mathbf{u}$ within study domain is a linear function of the correlation coefficient $\rho_{ij}^{\text{target}}$. Thus, we can rewrite Equation (7) for the global correlation between random variables Z_i and Z_j as follows:

$$\overline{\rho(Z_i(\mathbf{u}),Z_j(\mathbf{u}))} = a_{ij} + b_{ij}\rho_{ij}^{\text{target}}, \tag{8}$$

where a_{ij} and b_{ij}, $i, j = 1,...,N$ are constants calculated based on the Simple Kriging weights (and variances). Thus, we see that the only instance in which the target correlation-covariance matrix of the random variables to be simulated, $\boldsymbol{\rho}^{\text{target}}$, is reproduced, that is,

$$\overline{\rho(Z_i(\mathbf{u}), Z_j(\mathbf{u}))} = \rho_{ij}^{\text{target}}, \quad i, j = 1, \ldots, n, \tag{9}$$

is when

$$a_{ij} = 0, \quad b_{ij} = 1, \quad i, j = 1, \ldots, n. \tag{10}$$

Obviously, this event is highly unlikely. Thus, a correction to the multivariate Sequential Gaussian Simulation approach with correlated residuals must be made to reproduce the target correlation between variables Z_i and Z_j at lag 0.

Correction to the Multivariate Sequential Gaussian with Correlated Residuals

Let us assume that residuals $\mathbf{R}(\mathbf{u})^T = [R_1(\mathbf{u}), \ldots, R_N(\mathbf{u})]^T$ in multivariate sequential Gaussian simulation have the following structure. The residuals $\mathbf{R}(\mathbf{u})^T = [R_1(\mathbf{u}), \ldots, R_N(\mathbf{u})]^T$ are, as before, independent from location to location, that is,

$$\rho(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = 0, \quad \forall i, j = 1,..., N, \quad \forall k \neq l; \tag{11}$$

but correlated at the same location with some correlation-covariance matrix $\boldsymbol{\rho}$, that is,

$$\rho(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = \rho_{ij}, \quad \forall i, j = 1,..., N, \quad \forall k. \tag{12}$$

Then the global correlation matrix for residuals, $\boldsymbol{\rho}$, to be used in multivariate Sequential Gaussian simulation to reproduce the target correlation-covariance matrix of the random variables to be simulated, $\boldsymbol{\rho}^{\text{target}}$ following the same procedure as before. In particular, it can be shown that if $Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = \rho_{ij}$, then the global correlation between random variables Z_i and Z_j is equal to

$$\rho_{ij}^{\text{target}} = a_{ij} + b_{ij}\rho_{ij}, \tag{13}$$

where a_{ij} and b_{ij}, $i, j = 1,...,N$ are constants calculated based on the Simple Kriging weights (and variances). Note that if the multivariate Sequential Gausssian Simulation is unconditional, then constant a_{ij} in equation (13) is equal to 0 and

$$\rho_{ij}^{\text{target}} = b_{ij}\rho_{ij}. \tag{14}$$

In order for the multivariate Sequential Gaussian Simulation to honor the correlation matrix between random variables at lag 0, $\boldsymbol{\rho}^{\text{target}}$, it is required, at each step of the simulation, to generate residuals with the following correlation structure:

$$\rho(R_i(\mathbf{u}), R_j(\mathbf{u})) = \rho_{ij} = \begin{cases} \dfrac{\rho_{ij}^{\text{target}} - a_{ij}}{b_{ij}}, & if \left| \dfrac{\rho_{ij}^{\text{target}} - a_{ij}}{b_{ij}} \right| \leq 1, \quad \forall (i \neq j) = 1,..., N; \\ 1 \ \text{ or } \ -1, & \text{otherwise}; \end{cases} \tag{15}$$

$$\rho(R_i(\mathbf{u}), R_i(\mathbf{u})) = \rho_{ii} = 1, \quad \forall i = 1,..., N.$$

The constants a_{ij} and b_{ij}, required for the calculation of the correlation matrix , that will be used in multivariate SGS to reproduce the target correlation matrix $\boldsymbol{\rho}^{\text{target}}$, can be found as follows.

1. Independent ($\rho_{ij} = 0, \quad i \neq j = 1, \ldots, N$) multivariate SGS realizations are generated to find the coefficients a_{ij}'s as a correlation between realizations of the variables Z_i, Z_j, $i, j = 1,..., N$. Due to ergodic fluctuations, we expect only a minor change in the resultant correlation coefficients between

different random variable realizations. Therefore the value of each constant a_{ij}, $(i \neq j) = 1,...,N$, is taken as the average coefficient of correlation between realizations for variables Z_i and Z_j.

2. Perfectly dependent ($R_i(\mathbf{u}) = r, \quad i = 1,\ldots,N$) multivariate SGS realizations are generated in order to find the b_{ij} values as the average difference between the correlation of random variables Z_i and Z_j obtained in fully dependent simulation and the constants a_{ij} from (1), $(i \neq j) = 1,...,N$.

This approach is theoretically valid; however, the target correlation $\boldsymbol{\rho}^{\text{target}}$ may not be positive definite. This arises due to the fact that the coefficients of correlation for some pairs of variables are required to be significantly increased (due to b_{ij} being small). Therefore, if the matrix (15) is not positive definite, a positive definiteness correction to this matrix must be applied before it can be used in multivariate SGS.

Software Implementation

A program `sgsim_lu_matrix` was prepared based on program `usgsim` of GSLIB group (Deutsch and Journel, 1998) to calculate the matrix of correlations between residuals that reproduces the correlation between variables at lag 0. The parameter file for program `sgsim_lu_matrix` is presented below:

```
START OF MAIN:
10                              -number of realizations to generate
2                               -number of variables being simulated
69069                           -random number seed
256   0.5    1.0                -nx,xmn,xsiz
256   0.5    1.0                -ny,ymn,ysiz
1     0.5    1.0                -nz,zmn,zsiz
sgsim.out                       -file for output
12                              -number of data to use per variable
1                               -multiple grid search (0=no, 1=yes)
0     3                         -octant search (0=no, 1=yes), number
32.0   32.0   32.0              -maximum search radii (hmax,hmin,vert)
0.0    0.0    0.0               -angles for search ellipsoid
2                               -number of variograms
1                               -variable
2    0.0                        -nst, nugget effect
3    0.9  0.0   0.0   0.0       -it,cc,ang1,ang2,ang3
         32.0  32.0  10.0       -a_hmax, a_hmin, a_vert
1    0.1  0.0   0.0   0.0       -it,cc,ang1,ang2,ang3
         16.0  16.0  10.0       -a_hmax, a_hmin, a_vert
2                               -variable
2    0.0                        -nst, nugget effect
3    0.1  0.0   0.0   0.0       -it,cc,ang1,ang2,ang3
         32.0  32.0  10.0       -a_hmax, a_hmin, a_vert
1    0.9  0.0   0.0   0.0       -it,cc,ang1,ang2,ang3
         16.0  16.0  10.0       -a_hmax, a_hmin, a_vert
nodata                          -file with data
1  2  0  0                      -columns for X,Y,Z,wt
3  4                            -columns for variables
-1.0e21      1.0e21             -trimming limits
1                               -transform the data (0=no, 1=yes)
2                               -number of min/max values
1  0.0   15.0                 - variable, min, max
2  0.0   15.0                 - variable, min, max
0                               -number of reference distributions
0.5  -0.3                         -correlation coefficients: 0 with 1,2,.
     0.4                        -correlation coefficients: 1 with   2,.
```

A run of the program `sgsim_lu_matrix` creates an output file with five columns. First two columns give the *a*'s and *b*'s coefficients from (13); third column repeats the values of the target correlations between variables; column 4 presents the correlations between residuals calculated using Equation (15); column (5) presents a column of final correlations between residuals to be used in multivariate simulation, these coefficients are corrected to make the matrix of residual correlations positive-definite (if no correction is required, then result in column 5 is the same as in column 4). Because the correlation matrices and matrices of *a*'s and *b*'s coefficients are symmetric, the values in each column of the output file inform only on the upper triangular part of each matrix. Thus, if *n* variables are simulated, then the output file would contain $n(n+1)/2$ correlation values in each column.

Unconditional Multivariate Example 1

Consider the two standard normal random variables Z_1 and Z_2 with the following variograms characterizing their spatial continuity:

$$\begin{aligned}\gamma_{Z_1}(\mathbf{h}) &= 0.1 \cdot Sph_{16}(\mathbf{h}) + 0.9 \cdot Gaus_{32}(\mathbf{h}) \\ \gamma_{Z_2}(\mathbf{h}) &= 0.9 \cdot Sph_{16}(\mathbf{h}) + 0.1 \cdot Gaus_{32}(\mathbf{h})\end{aligned}, \tag{16}$$

The coefficient of correlation between the two random variables is assumed to be 0.5.

Our goal is to apply multivariate Sequential Gaussian Simulation to generate 100 realizations of the two random variables using their respective variogram models, that is, $\gamma_{Z_1}(\mathbf{h})$ and $\gamma_{Z_2}(\mathbf{h})$, so that the correlation at lag 0 between Z_1 and Z_2 is 0.5. As a first step we must find the coefficient of correlation ρ_{12} to be used in multivariate SGS to reproduce the target correlation $\rho_{12}^{\text{target}}$ of 0.5. For such purpose a set of fully dependent and fully independent multivariate SGS realizations must be generated first to find coefficients b_{12} and a_{12} in (15). Note that because the simulation is unconditional, the coefficient a_{12} should be equal to 0 within acceptable ergodic fluctuation.

Figure 1 shows the histograms of the coefficients b_{12} and a_{12} obtained in 100 fully dependent and fully independent, respectively, multivariate SGS realizations. It can be clearly seen from Figure 1 that coefficient a_{12} is virtually zero (-0.006), as expected. While the value of coefficient b_{12} is 0.746. This implies that a correlation coefficient between residuals for generation of Z_1 and Z_2 is given by

$$\rho_{12} = \frac{\rho_{12}^{\text{target}}}{b_{12}} = \frac{0.5}{0.746} = 0.670 \tag{17}$$

needs to be applied in the multivariate SGS in order to reproduce the target correlation of 0.5. Figure 2 shows the distribution of the correlation coefficients between the two random variables under study obtained by multivariate Sequential Gaussian Simulation with residual's correlation coefficient equal to 0.670. Figure 2 also shows the distribution of the correlation coefficients between realizations for Z_1 and Z_2 obtained by multivariate SGS with residual's correlation coefficient equal to 0.5 (target correlation); which is the conventional approach.

From Figure 2 we can clearly see that the corrected correlation works perfectly for this example; the target correlation is nicely reproduced. The same cannot be said for the conventional approach. Multivariate SGS in this case results in a correlation of 0.367 25% below the target) when the conventional approach is used.

Figure 3 shows the variogram reproduction for Z_1 and Z_2 obtained in multivariate Sequential Gaussian Simulation with correlation coefficient between residuals fixed at 0.670. As expected, both variograms are reproduced within ergodic fluctuation. Correlation between residuals has no impact whatsoever on the variogram structure. This is because residuals are independent from location to location. The proof of variogram/covariance reproduction in multivariate sequential simulation that reproduces correlation at lag 0 is a straightforward extension of the proof for variogram/covariance reproduction in the univariate SGS (Goovaerts, 1997).

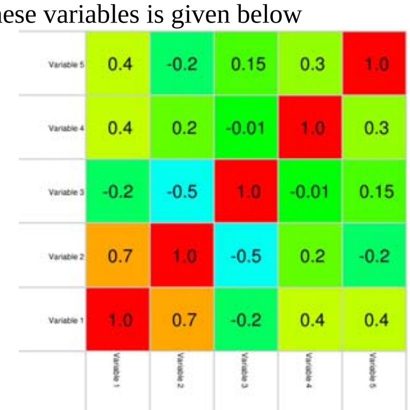
It is also worth noting that the average largest achievable correlation in multivariate SGS between two random variables Z_1 and Z_2 with spatial continuity characterized by $\gamma_{Z_1}(\mathbf{h})$ and $\gamma_{Z_2}(\mathbf{h})$ given by (11) is 0.746 (see Figure 1). Note that the correlation coefficient of 0.746 results in an unfeasible Linear Model of Correlation. So, in a sense, the multivariate correlated SGS approach can allow an increased flexibility.

Unconditional Multivariate Example 2

Let us now consider five standard normal random variables $Z_i, \quad i = 1,\ldots,5,$ with the following variograms characterizing their spatial continuity

$$\begin{aligned}
\gamma_{Z_1}(\mathbf{h}) &= 0.1 \cdot Sph_{16}(\mathbf{h}) + 0.9 \cdot Gaus_{32}(\mathbf{h}) \\
\gamma_{Z_2}(\mathbf{h}) &= 0.5 \cdot Exp_{20}(\mathbf{h}) + 0.5 \cdot Sph_{40}(\mathbf{h}) \\
\gamma_{Z_3}(\mathbf{h}) &= 0.3 \cdot Exp_{5}(\mathbf{h}) + 0.7 \cdot Sph_{12}(\mathbf{h}) \\
\gamma_{Z_4}(\mathbf{h}) &= 0.9 \cdot Sph_{16}(\mathbf{h}) + 0.1 \cdot Gaus_{32}(\mathbf{h}) \\
\gamma_{Z_5}(\mathbf{h}) &= 0.5 \cdot Sph_{16}(\mathbf{h}) + 0.5 \cdot Gaus_{32}(\mathbf{h})
\end{aligned}, \tag{18}$$

The correlation matrix between these variables is given below



Now let us consider multivariate SGS for generation of the random variables that honor the correlation matrix between variables at lag 0. We split our example into 2 parts:

1. multivariate SGS for the variables $Z_i, \quad i = 1,2,3;$
2. multivariate SGS for all five variables, $Z_i, \quad i = 1,\ldots,5;$

and compare results of the conventional multivariate SGS with corrected multivariate SGS proposed in this thesis.

To find the coefficients of correlation ρ_{12}, ρ_{13} and ρ_{23} to be used in multivariate SGS to reproduce the target correlations of 0.7, -0.2 and -0.5, respectively, fully dependent and fully independent multivariate SGS realizations are generated first. These realizations are used to obtain the coefficients b_{12}, b_{13}, b_{23}, a_{12}, a_{13} and a_{23} in (15). Because the simulation is unconditional, the coefficients $a_{ij}, (i \neq j) = 1,\ldots,3,$ should be equal to 0 within acceptable ergodic fluctuation.

Figure 4 shows the distributions of coefficients b_{ij} and $a_{ij}, (i \neq j) = 1,\ldots,3,$ obtained by 100 fully dependent and fully independent, respectively, multivariate SGS realizations. From Figure 4 we see that the coefficients a_{ij}'s are, as expected, virtually zero; the coefficients b_{ij}'s are given below

$$b_{12} = 0.857; \quad b_{13} = 0.476; \quad b_{23} = 0.835. \tag{19}$$

This implies that the following correlation coefficients

$$\begin{aligned}
\rho_{12} &= \frac{0.7}{0.857} = 0.817; \\
\rho_{13} &= \frac{-0.2}{0.476} = -0.420;
\end{aligned} \tag{20}$$

$$\rho_{23} = \frac{-0.5}{0.835} = -0.599.$$

need to be applied in the multivariate SGS in order to reproduce the target correlations of 0.7, -0.2 and -0.5, respectively, between random variables Z_1, Z_2 and Z_3.

Figure 5 shows the distribution of the correlation coefficients ρ_{12}, ρ_{13} and ρ_{23} obtained by newly proposed multivariate SGS. For comparison, Figure 5 also shows the distribution of the correlation coefficients that would be obtained by conventional approach. From Figure 5 one can clearly note that the approach proposed in this paper results in almost perfect reproduction of the target correlations. The largest absolute mismatch in the correlation coefficients is 0.004. The same cannot be said about the conventional approach. The largest absolute mismatch in the correlation coefficients in the conventional approach is more than 0.1, which is quite significant.

Figure 6 shows the variogram reproduction for all three variables obtained in the multivariate SGS with corrected correlation matrix. All variograms are aceptably reproduced within ergodic fluctuation. The same procedure as before is applied to find the coefficients $a_{ij}, b_{ij},\ (i \neq j) = 1, \ldots, 5,$ in (15). The following is a summary of the results for the $b_{Y_{ij}}$'s coefficients (simulation is unconditional => a_{ij}'s coefficients are zero)

$$\begin{aligned} &b_{12} = 0.857; \quad b_{13} = 0.476; \quad b_{14} = 0.742; \quad b_{15} = 0.873; \quad b_{23} = 0.834; \\ &b_{24} = 0.949; \quad b_{25} = 0.999; \quad b_{34} = 0.895; \quad b_{35} = 0.818; \quad b_{45} = 0.943; \end{aligned} \tag{21}$$

This implies that the following correlation coefficients:

$$\begin{aligned} &\rho_{12} = \frac{0.7}{0.857} = 0.817; \quad \rho_{13} = \frac{-0.2}{0.476} = -0.420; \quad \rho_{14} = \frac{0.4}{0.742} = 0.539; \\ &\rho_{15} = \frac{0.4}{0.873} = 0.458; \quad \rho_{23} = \frac{-0.5}{0.834} = -0.600; \quad \rho_{24} = \frac{0.2}{0.949} = 0.211; \\ &\rho_{25} = \frac{-0.2}{0.999} = -0.2; \quad \rho_{34} = \frac{-0.01}{0.895} = 0.011; \quad \rho_{35} = \frac{0.15}{0.818} = 0.183; \\ &\rho_{45} = \frac{0.3}{0.943} = 0.318. \end{aligned} \tag{22}$$

Note that the corrected correlation coefficients corresponding to correlation between first three random variables are exactly the same as before. This observation once again confirms the linear relationship in the correlation and independence of the solutions for the correlation coefficients. The combined matrix of the correlation coefficients, however, is not positive semi-definite. The smallest eigenvalue of the corrected correlation matrix is -0.0434. Thus, a correction must be applied. There are two possible choices for the correction:

1. Standardize the off-diagonal elements in the correlation matrix by (1 – smallest eigenvalue). In our case the off-diagonal elements should be standardized by 1.0434. This is a minor correction.
2. Correct only rows and columns corresponding to the variables making the correlation matrix negative definite. This correction may be feasible if these variables were less important than other. In our case correlation matrix calculated based on the first four random variables was positive definite, after adding fifth variable it became negative definite. To make it positive definite we need to multiply the correlations of the firth variable with all others by 0.84. In our example we consider all the variables equally important; thus, we will not consider this type of correction in our work.

The following matrix shows the input correlation matrix to the multivariate SGS (the correlation matrix was made positive definite via correction option 1)

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0000 & 0.7829 & -0.4028 & 0.5167 & 0.4391 \\ 0.7829 & 1.0000 & -0.5748 & 0.2020 & -0.1919 \\ -0.4028 & -0.5748 & 1.0000 & -0.0107 & 0.1757 \\ 0.5167 & 0.2020 & -0.0107 & 1.0000 & 0.3048 \\ 0.4391 & -0.1919 & 0.1757 & 0.3048 & 1.0000 \end{bmatrix} \qquad (23)$$

Figure 7 shows the correlation matrix between variables reproduced by the newly proposed multivariate SGS. For comparison, Figure 7 also shows the reproduced correlation matrix in the conventional approach. The mismatch in the results for correlation obtained by the two approaches to multivariate SGS are shown in Figure 8. Note that the maximum absolute mismatch in the correlations obtained using multivariate sequential simulation that reproduces correlation at lag zero is 0.028, while in the conventional approach it is 0.107. The slight increase in the mismatch in the reproduction of the target correlation coefficients obtained via the new approach is connected to the correction of the input correlation matrix to make it positive definite.

Example of Conditional Multivariate SGS

Let us consider the following small example. Figure 9 shows the locations of 20 primary data in the study domain of size 100 by 100 units; the primary data distribution, the crossplot between primary data and collocated secondary data and the distribution of the secondary data collocated to primary. All data are in Gaussian units.

The following linear model of coregionalization describes the joint spatial continuity of the primary and secondary data:

$$\begin{aligned} \gamma_{YY}(\mathbf{h}) &= 0.3 \cdot Exp_{\substack{a_1=10 \\ a_2=20}}(\mathbf{h}) + 0.7 \cdot Sph_{\substack{a_1=20 \\ a_2=40}}(\mathbf{h}) \\ \gamma_{YZ}(\mathbf{h}) &= 0.45 \cdot Exp_{\substack{a_1=10 \\ a_2=20}}(\mathbf{h}) + 0.35 \cdot Sph_{\substack{a_1=20 \\ a_2=40}}(\mathbf{h}), \\ \gamma_{ZZ}(\mathbf{h}) &= 0.8 \cdot Exp_{\substack{a_1=10 \\ a_2=20}}(\mathbf{h}) + 0.2 \cdot Sph_{\substack{a_1=20 \\ a_2=40}}(\mathbf{h}) \end{aligned} \qquad (24)$$

Now, let us consider modeling the primary and secondary random variables via multivariate Sequential Gaussian Simulation approach honoring the correlation between random variables at lag distance 0.

The coefficients b_{12} and a_{12} can be calculated to be equal to

$$a_{12} = 0.087 \quad \text{and} \quad b_{12} = 0.842,$$

thus the coefficient of correlation between residuals of

$$\rho_{12} = \frac{\rho_{12}^{\text{target}} - a_{12}}{b_{12}} = \frac{0.8 - 0.087}{0.842} = 0.847 \qquad (25)$$

needs to be applied in the multivariate SGS to reproduce the target correlation of 0.8.

Figure 10 shows the distribution of the coefficients of correlation between the primary and secondary random variables obtained by multivariate Sequential Gaussian Simulation with the residual's coefficient of correlation given in (25). Note from Figure 10 that the target correlation is acceptably reproduced.

Figure 11 shows the variogram reproduction for the primary and secondary random variables obtained in multivariate Sequential Gaussian Simulation with correlation coefficient between residuals fixed at 0.847. Note that variogram reproduction is acceptable within ergodic fluctuation. Also it is interesting to note that for the considered example the crossvariogram between variables given in (24) is very nicely reproduced within ergodic fluctuations, see Figure 12.

Local Correlation

With a reasonable practical effort, the correction proposed for multivariate SGS can be localized. That is, multivariate unconditional SGS with locally varying correlated residuals $\mathbf{R}(\mathbf{u})^T = [R_1(\mathbf{u}),\ldots,R_N(\mathbf{u})]^T$ can be developed. The procedure for finding the prescribed locally varying correlation matrix $\boldsymbol{\rho}(\mathbf{u})$ of the residuals to reproduce the target correlation-covariance matrix for the random variables to be simulated, $\boldsymbol{\rho}^{\text{target}}$, is as follows.

The residuals $\mathbf{R}(\mathbf{u})^T = [R_1(\mathbf{u}),\ldots,R_N(\mathbf{u})]^T$ are assumed to be independent from location to location, that is,

$$\rho(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = 0, \quad \forall i, j = 1,..., N, \quad \forall k \neq l; \tag{26}$$

but correlated at the same location with locally varying correlation-covariance matrix $\boldsymbol{\rho}(\mathbf{u})$, that is,

$$\rho(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_k)) = \rho_{ij}(\mathbf{u}), \quad \forall i, j = 1,..., N, \quad \forall k. \tag{27}$$

Let us now calculate the correlation-covariance between two simulated values for variables Z_i and Z_j, respectively, at an arbitrary location $\mathbf{u}$ in the study domain, $i, j \in \{1,\ldots,N\}$. These simulated values $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$ in sequential Gaussian simulation are given by

$$\begin{aligned} Z_i(\mathbf{u}) &= m_{SK,i}(\mathbf{u}) + \sigma_{SK,i}(\mathbf{u})R_i(\mathbf{u}), \\ Z_j(\mathbf{u}) &= m_{SK,j}(\mathbf{u}) + \sigma_{SK,j}(\mathbf{u})R_j(\mathbf{u}), \end{aligned} \tag{28}$$

then due to independence between the vector of residuals $[R_i(\mathbf{u})\ R_j(\mathbf{u})]^T$ and the vector of Simple Kriging means $[m_{SK,i}(\mathbf{u})\ m_{SK,j}(\mathbf{u})]^T$,

$$\begin{aligned} \rho(Z_i(\mathbf{u}), Z_j(\mathbf{u})) &= Cov(Z_i(\mathbf{u}), Z_j(\mathbf{u})) \\ &= Cov(m_{SK,i}(\mathbf{u}), m_{SK,j}(\mathbf{u})) + \sigma_{SK,i}(\mathbf{u})\sigma_{SK,j}(\mathbf{u})Cov(R_i(\mathbf{u}), R_j(\mathbf{u})) \\ &= Cov(m_{SK,i}(\mathbf{u}), m_{SK,j}(\mathbf{u})) + \sigma_{SK,i}(\mathbf{u})\sigma_{SK,j}(\mathbf{u})\rho_{ij}(\mathbf{u}). \end{aligned} \tag{29}$$

Note that

$$\begin{aligned} m_{SK,i}(\mathbf{u}) &= \sum_{k=1}^{n_i(\mathbf{u})} \lambda_{SK,k}^{i} Z_i(\mathbf{u}_k), \\ m_{SK,j}(\mathbf{u}) &= \sum_{l=1}^{n_j(\mathbf{u})} \lambda_{SK,k}^{j} Z_j(\mathbf{u}_l), \end{aligned} \tag{30}$$

where $Z_i(\mathbf{u}_k),\ \ k = 1,\ldots,n_i(\mathbf{u})$, and $Z_j(\mathbf{u}_k),\ \ l = 1,\ldots,n_j(\mathbf{u})$ denote the $n_i(\mathbf{u})$ and $n_j(\mathbf{u})$ closest simulated nodes for variables Z_i and Z_j, respectively; $\lambda_{SK,k}^{i}$, $k = 1,\ldots,n_i(\mathbf{u})$, and $\lambda_{SK,l}^{i}$, $l = 1,\ldots,n_j(\mathbf{u})$, denote the Simple Kriging weights obtained for location $\mathbf{u}$ when estimating variables Z_i and Z_j, respectively.

The Simple Kriging means given in (30) can be rewritten because the simulation is unconditional:

$$\begin{aligned} m_{SK,i}(\mathbf{u}) &= \sum_{k=1}^{N_i(\mathbf{u})} \mu_k^i R_i(\mathbf{u}_k), \\ m_{SK,j}(\mathbf{u}) &= \sum_{l=1}^{N_j(\mathbf{u})} \mu_k^j R_j(\mathbf{u}_l), \end{aligned} \tag{31}$$

where $R_i(\mathbf{u}_k),\ \ k = 1,\ldots,N_i(\mathbf{u})$, and $R_j(\mathbf{u}_k),\ \ l = 1,\ldots,N_j(\mathbf{u})$ denote the Gaussian residuals generated for calculation of the $n_i(\mathbf{u})$ and $n_j(\mathbf{u})$ closest simulated nodes to an estimation location $\mathbf{u}$ for variables Z_i and Z_j, respectively; μ_k^i, $k = 1,\ldots,N_i(\mathbf{u})$, and μ_l^i, $l = 1,\ldots,N_j(\mathbf{u})$, denote the weights given to these residuals. Then,

$$
Cov(m_{SK,i}(\mathbf{u}), m_{SK,j}(\mathbf{u})) = Cov\left(\sum_{k=1}^{N_i(\mathbf{u})} \mu_k^i R_i(\mathbf{u}_k), \sum_{l=1}^{N_j(\mathbf{u})} \mu_k^j R_j(\mathbf{u}_l) \right)
$$

$$
= \sum_{k=1}^{N_i(\mathbf{u})} \sum_{l=1}^{N_j(\mathbf{u})} \mu_k^i \mu_k^j Cov(R_i(\mathbf{u}_k), R_j(\mathbf{u}_l)) = \sum_{s=1}^{N(\mathbf{u})} \tilde{\mu}_s^j \tilde{\mu}_s^i \rho_{ij}(\mathbf{u}_s), \tag{32}
$$

where $N(\mathbf{u})$ denotes the number of location with residuals common to both random variables; $\tilde{\mu}_s^i, \tilde{\mu}_s^i$ denote the residual weights assigned to location with residuals common to both random variables; and $\rho_{ij}(\mathbf{u}_s)$ denotes the correlation between residuals at location with residuals common to both random variables $s = 1, \ldots, N(\mathbf{u})$. Moreover, because we aim at

$$
\rho(Z_i(\mathbf{u}), Z_j(\mathbf{u})) = \rho_{ij}^{\mathbf{target}}(\mathbf{u}), \tag{33}
$$

the following equality must hold:

$$
\rho_{ij}^{\mathbf{target}}(\mathbf{u}) = \sum_{s=1}^{N(\mathbf{u})} \tilde{\mu}_s^j \tilde{\mu}_s^i \rho_{ij}(\mathbf{u}_s) + \sigma_{SK,i}(\mathbf{u}) \sigma_{SK,j}(\mathbf{u}) \rho_{ij}(\mathbf{u}). \tag{34}
$$

thus, in order for the multivariate unconditional Sequential Gaussian Simulation to honor the locally varying correlation matrix between random variables at lag 0, $\boldsymbol{\rho}^{\text{target}}(\mathbf{u})$, the residuals with the following correlation structure need to be generated locally for each simulation location **u**:

$$
\rho_{ij}(\mathbf{u}) = \begin{cases} \dfrac{\rho_{ij}^{\text{target}}(\mathbf{u}) - \sum_{s=1}^{N(\mathbf{u})} \tilde{\mu}_s^j \tilde{\mu}_s^i \rho_{ij}(\mathbf{u}_s)}{\sigma_{SK,i}(\mathbf{u}) \sigma_{SK,j}(\mathbf{u})}, & if \left| \dfrac{\rho_{ij}^{\text{target}}(\mathbf{u}) - \sum_{s=1}^{N(\mathbf{u})} \tilde{\mu}_s^j \tilde{\mu}_s^i \rho_{ij}(\mathbf{u}_s)}{\sigma_{SK,i}(\mathbf{u}) \sigma_{SK,j}(\mathbf{u})} \right| \le 1, \\ 1 \quad \text{or} \quad -1, & \text{otherwise;} \end{cases} \tag{35}
$$

$$
\rho_{ii}(\mathbf{u}) = 1, \quad \forall i = 1,...,N.
$$

for any $(i \neq j) = 1,..., N$.

This correction is not necessarily positive definite. Therefore if matrix (35) is not positive definite, a positive definiteness correction to this matrix must be applied first at the location of non-positive-definiteness, then it can be used in multivariate SGS. The only situation where matrix (35) is known to be positive definite at any estimation location **u** is in the case of multivariate unconditional Gaussian simulation with only two random variables.

Conclusions

Multivariate Sequential Gaussian Simulation honoring correlation between variables at lag 0 represents a neat alternative to Sequential Gaussian Simulation with Intrinsic Collocated Cokriging and Sequential Gaussian Simulation with Simple Cokriging in the case when secondary data is not exhaustively sampled or no secondary information is available.

This is because multivariate Sequential Gaussian Simulation approach with correlated residuals is simple; it ensures reproduction of all target statistics, that is, mean variance and target direct variograms. Moreover, despite the multivariate SGS honoring the correlation between the variables at lag distance 0 is not designed to reproduce the crossvariograms/crosscovariances between variables, we have observed through many examples that multivariate SGS usually results in quite good reproduction of the cross variograms, while it does not require (on the contrary to Sequential Gaussian Simulation with Simple Cokriging) the joint model of covariances to be input to the simulation.

A possible drawback of the approach is the possible non- positive definiteness of the correlation matrices for the residuals to be used in multivariate Sequential Gaussian Simulation to honor correlation between variables at lag 0. This must be corrected as part of the algorithm.

References

Almeida, A., Journel, A., 1994. Joint simulation of multiple variables with a Markov-type coregionalization model. Mathematical Geology 26(5), 565–588.

Babak, O. and Deutsch, C.V., 2008. An Intrinsic Model of Coregionalization that Solves Variance Inflation in Collocated Cokriging. Accepted to Computers & Geosciences.

Chilès, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty, 1st edn., Wiley, New York, NY, 695 pp.

Davis, M. W., Production of conditional simulations via the LU triangular decomposition of the covariance matrix, Mathematical Geology, 19(2), 1987, 99-107.

Deutsch, C.V., 2002. Geostatistical Reservoir Modeling, 1st edn., Oxford University Press, New York, NY, 376 pp.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation, 1st edn., Oxford University Press, New York, NY, 483 pp.

Journel A.G., Huijbregts, Ch.J., 1978. Mining Geostatistics, 1st edn., Academic Press, London, 600 pp.

Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction with Applications, 3rd edn., Springer, New York, NY, 387 pp.

Xu, W., Tran, T. T., Srivastava, R. M., Journel, A. G. 1992. Integrating seismic data in reservoir modeling: The Collocated Cokriging alternative. 67th Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Washington, DC, SPE paper 24742, pp. 833–842.

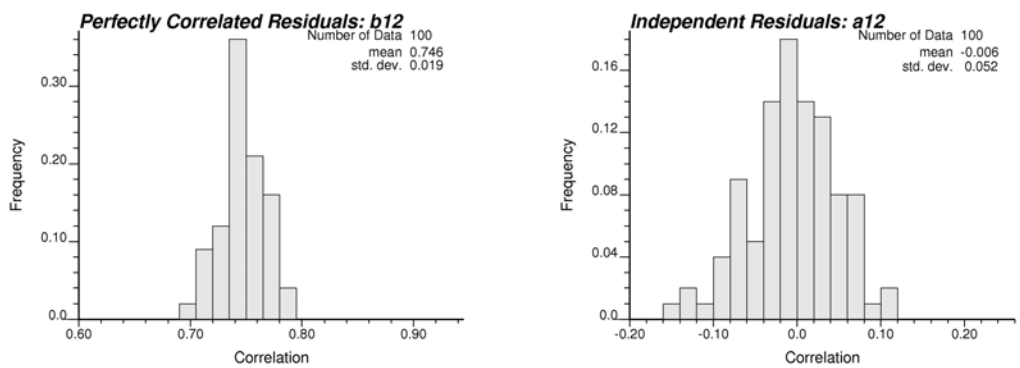


Figure 1: The histogram of the coefficients b_{12} (left) and a_{12} (right) obtained in 100 fully dependent and fully independent, respectively, multivariate Sequential Gaussian Simulations.

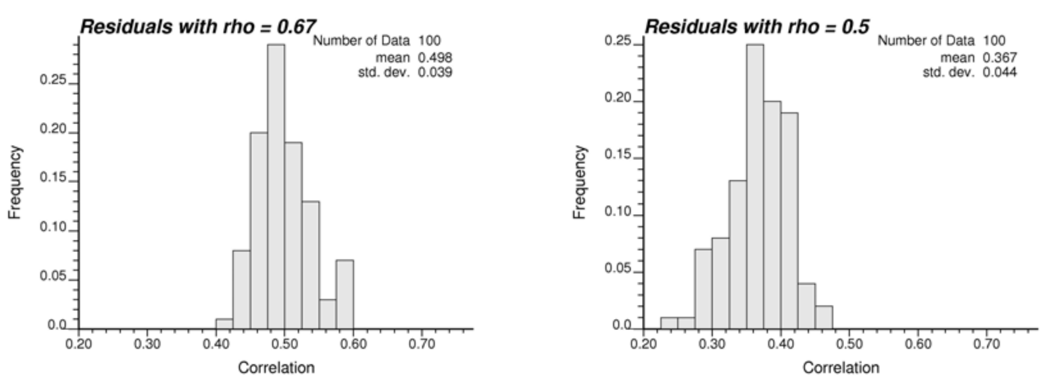


Figure 2: Distribution of the correlation coefficients between Z_1 and Z_2 obtained by multivariate Sequential Gaussian Simulation with residual's correlation coefficient equal to 0.670 (left); Distribution of the correlation coefficients between Z_1 and Z_2 obtained by conventional approach (right).

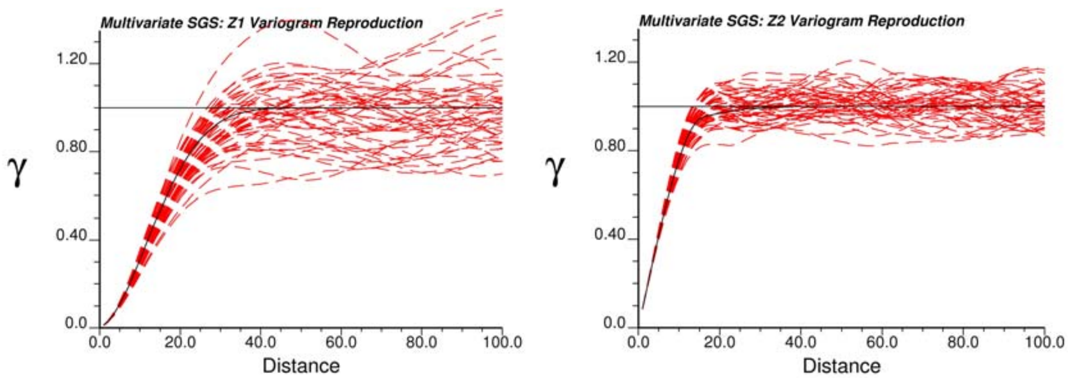


Figure 3: The variogram reproduction for Z_1 (left) and Z_2 (right) obtained in multivariate Sequential Gaussian Simulation with correlation coefficient between residuals fixed at 0.670.

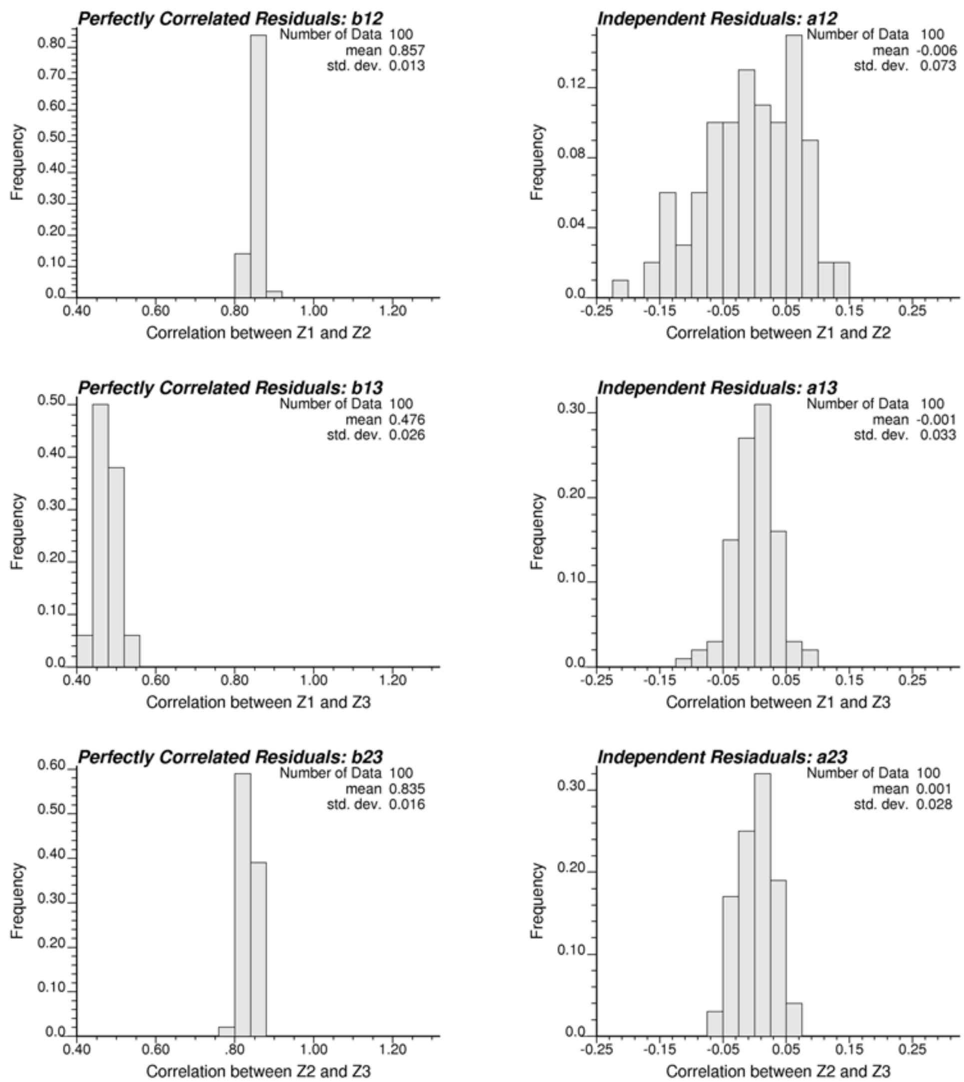


Figure 4: Distributions of coefficients b_{12} (top left), b_{13} (middle left), b_{23} (bottom left), a_{12} (top right), a_{13} (middle right), and a_{23} (middle bottom),obtained by 100 fully dependent and fully independent, respectively, multivariate Sequential Gaussian Simulations.

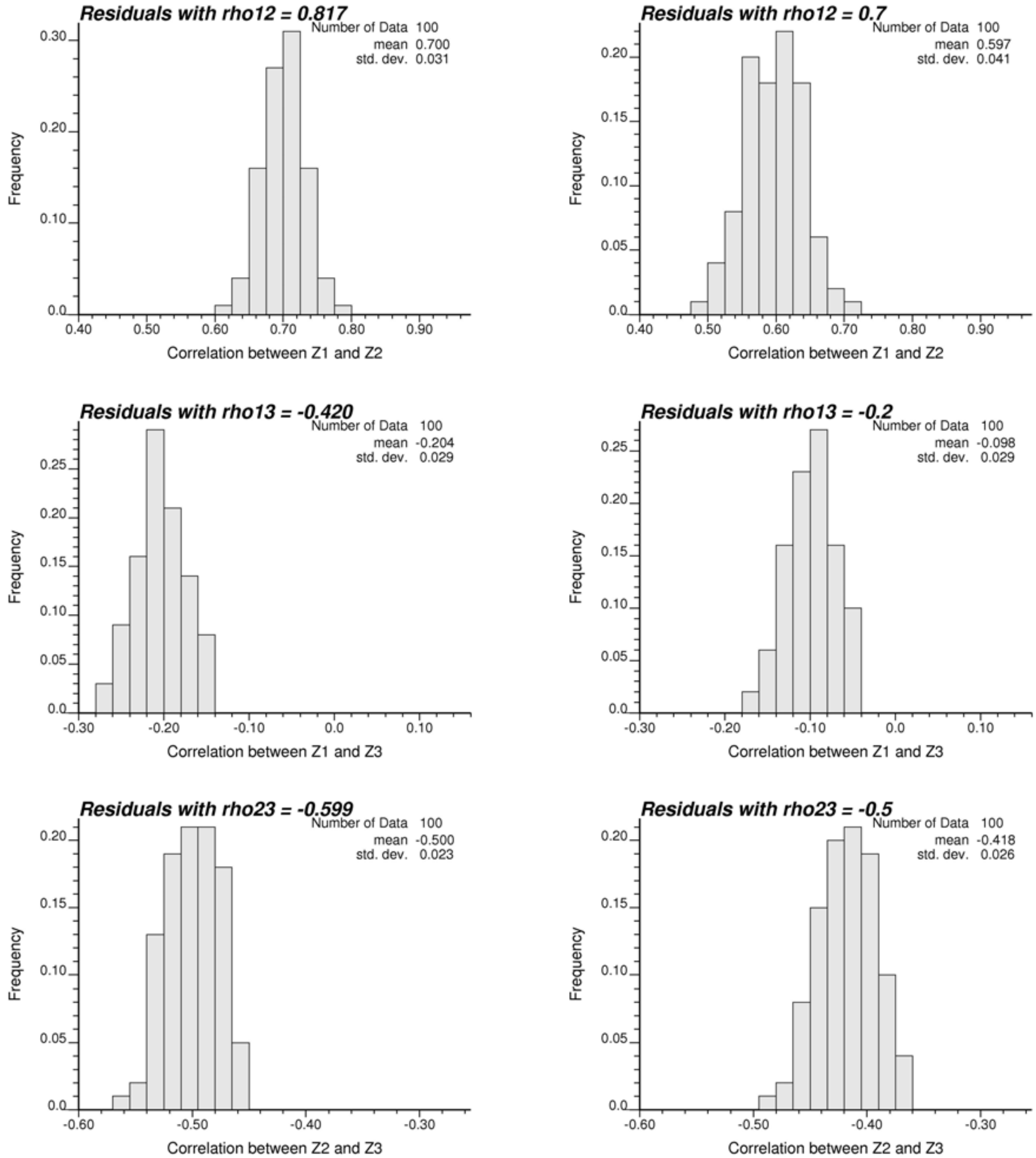


Figure 5: Distribution of the correlation coefficients ρ_{12}, ρ_{13} and ρ_{23} obtained by the newly proposed corrected multivariate SGS (left) and by conventional approach (right).

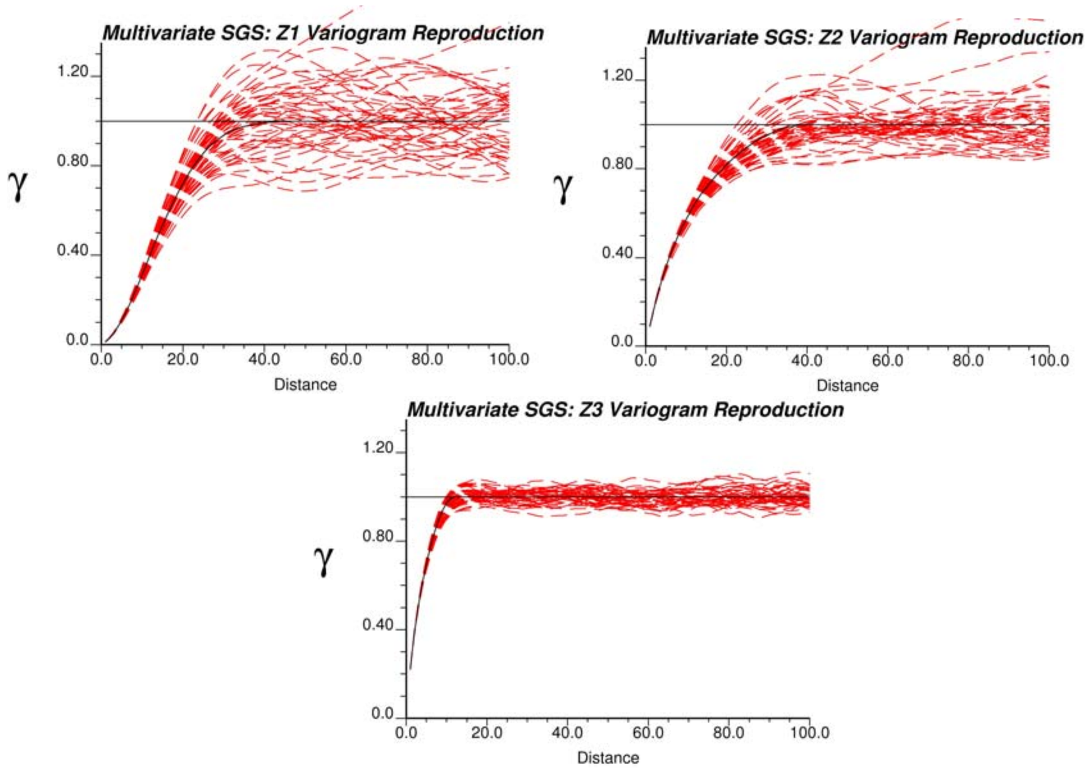


Figure 6: Variogram reproduction for $Z_1, Z_2,$ and Z_3 obtained in the multivariate SGS with corrected correlation matrix in (15).

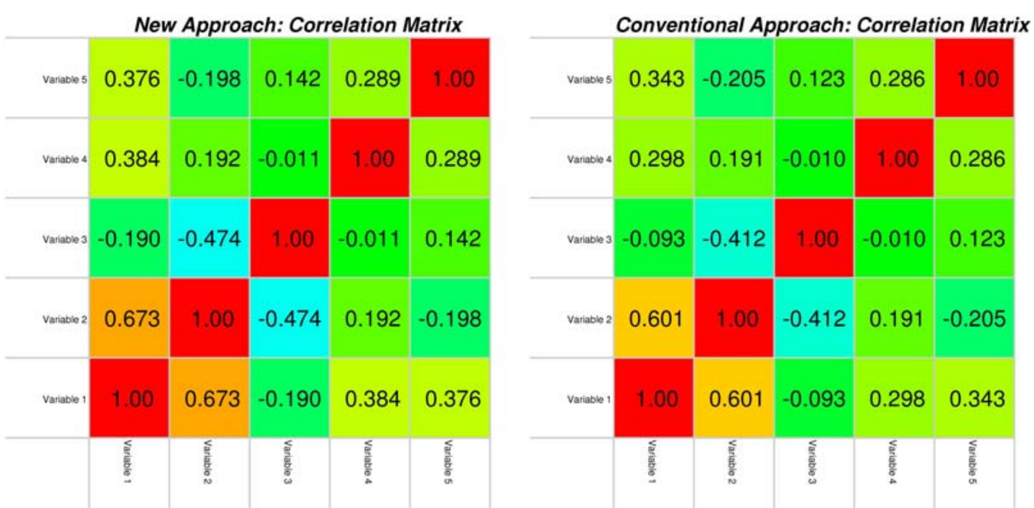


Figure 7: Correlation matrix between Z_1, Z_2, Z_3, Z_4, and Z_5 reproduced by the newly proposed multivariate SGS (left) and by conventional approach (right).

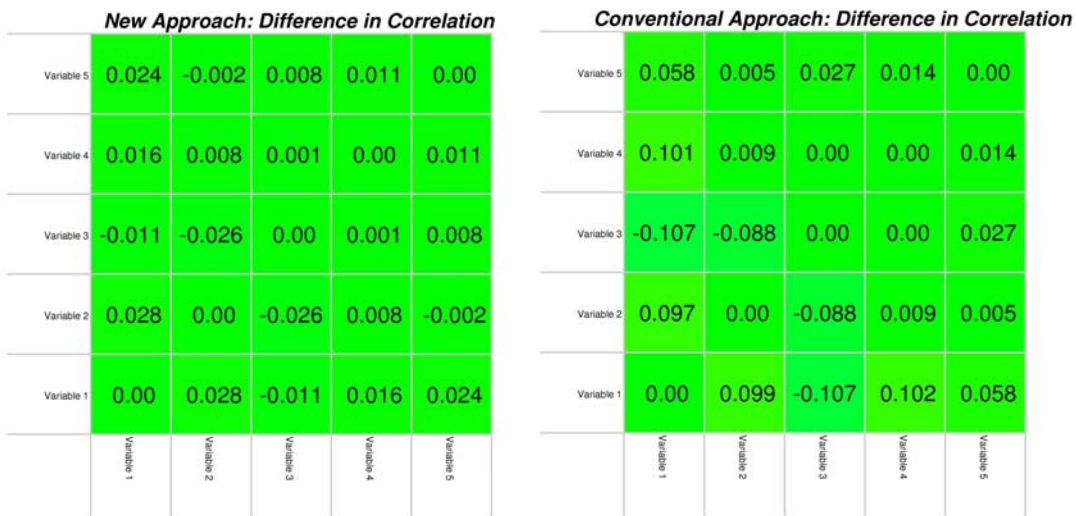


Figure 8: The mismatch in the reproduced correlation matrix between Z_1, Z_2, Z_3, and Z_4 obtained by the newly proposed multivariate SGS (left) and by conventional approach (right).

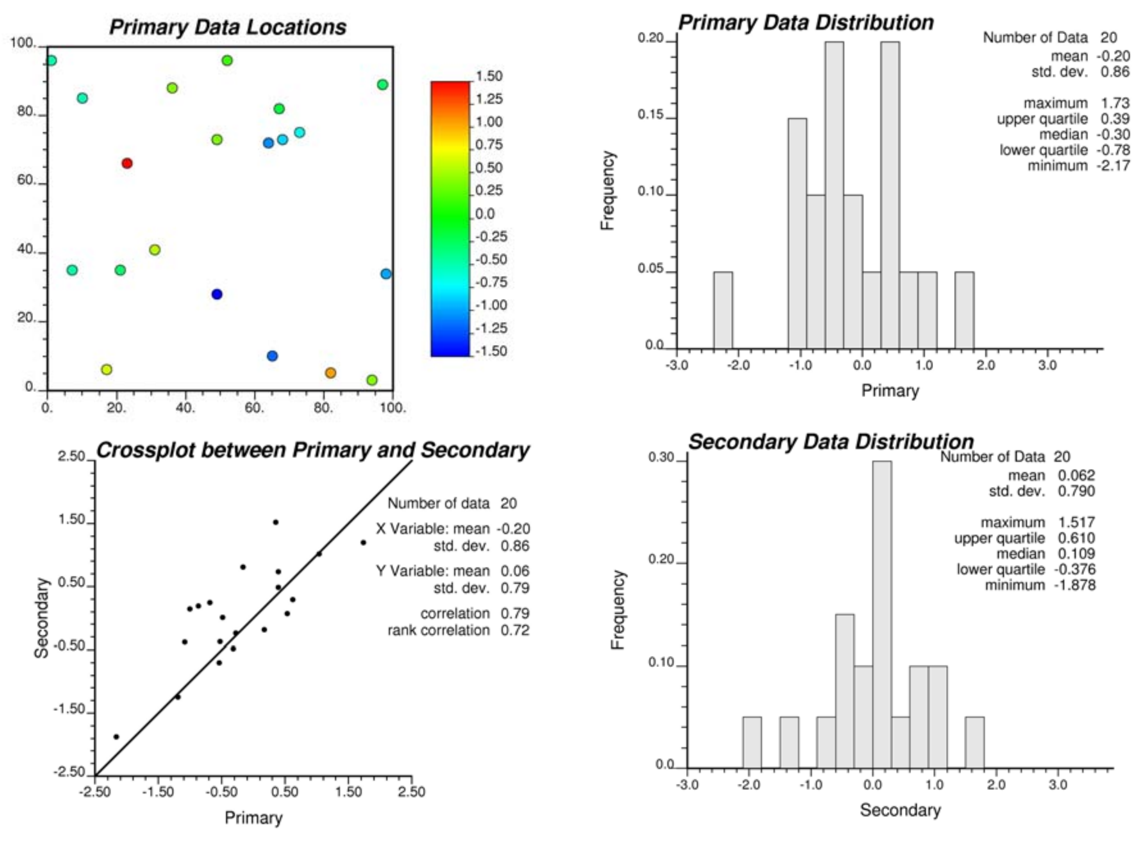


Figure 9: Locations of the 20 primary data (top left) and their distribution (top right); the crossplot between primary data and collocated secondary data (bottom left) and the distribution of the secondary data (bottom right). The data are in Gaussian units.

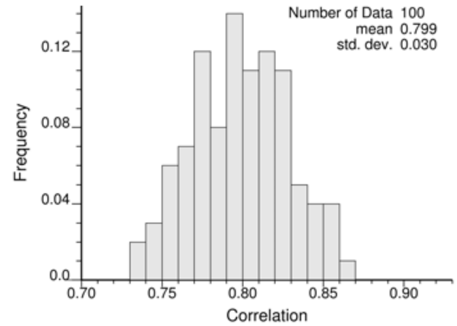


Figure 10: Distribution of the correlation coefficients between primary and secondary random variables obtained by multivariate Sequential Gaussian Simulation with residual correlation coefficient given in (22).

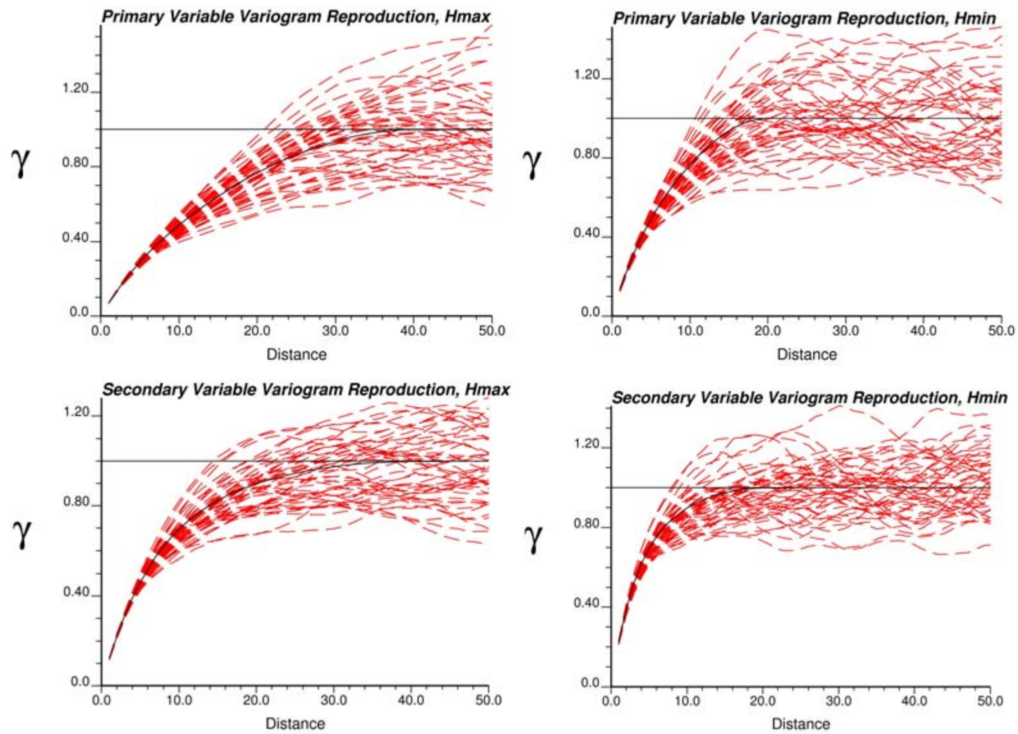


Figure 11: Variogram reproduction in the direction of major (left) and minor (right) continuity for primary (top) and secondary (bottom) random variables obtained in the multivariate SGS with correlation matrix given in (22).

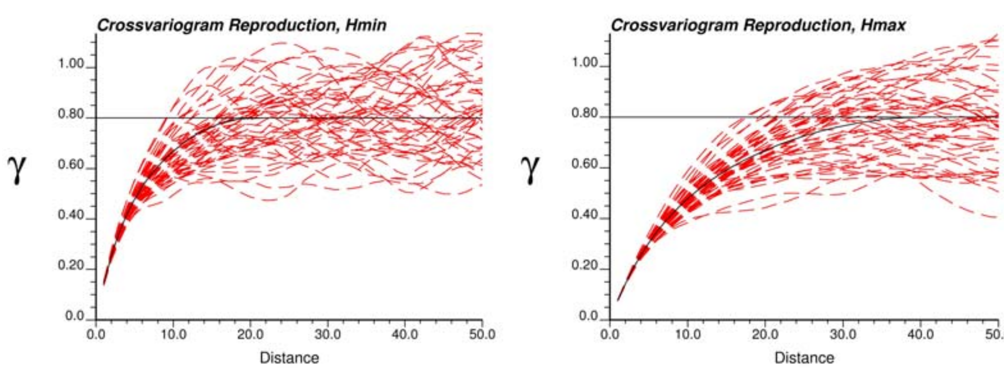


Figure 12: Crossvariogram reproduction in the direction of major (left) and minor (right) continuity obtained in the multivariate SGS with correlation matrix given in (22).