Mineral resource modelling of variables with inequality constraints: a case study of an iron ore deposit

S. Abulkhair and N. Madani

Nazarbayev University, Kazakhstan

In multivariate geostatistics, it is common to have different types of complexities between variables of interest. In this context, an inequality constraint is an example of complex bivariate relationships. Unfortunately, traditional co-kriging and co-simulation algorithms cannot reproduce this type of bivariate complexity, leading to the overestimation of disturbing elements. This paper proposes a new algorithm based on a hierarchical sequential Gaussian co-simulation framework, integrated with inverse transform sampling, to model inequality constraints between variables. First, the proposed methodology's validity was evaluated by applying it to a real case study from an iron deposit, with an inequality constraint between iron and aluminum oxide. Then the simulated results were compared with a conventional hierarchical co-simulation algorithm to investigate the effect of inverse transform sampling on the quality of the co-simulation. The results showed that the proposed algorithm can reproduce an inequality constraint between variables.

INTRODUCTION

Geostatistics aims to produce unbiased resource models using information from exploration boreholes and geophysical data (Goovaerts, 1997; Pyrcz and Deutsch, 2014). The accuracy of the produced models is usually investigated through histogram and variogram validation. However, the reproduction of spatial correlations between co-regionalised variables is another crucial aspect addressed by multivariate geostatistics (Wackernagel, 2003). Among others, Gaussian-based co-simulation algorithms, such as sequential Gaussian co-simulation (SGCOSIM) (Verly, 1993) and turning bands co-simulation (TBCOSIM) (Emery, 2008), are widely-used techniques. Moreover, various factorisation-based methods have been designed to avoid the tedious fitting of cross variograms (Davis and Greenes, 1983; Desbarats and Dimitrakopoulos, 2000; Leuangthong and Deutsch, 2003; Emery and Ortiz, 2012; Barnett *et al.*, 2014). However, one of the limitations of many co-simulation and factorisation algorithms is the reproduction of an inequality constraint. This is a type of bivariate complexity, which is characterised by linear inequation. Several authors have tried to incorporate factorisation-based approaches in the modelling of multivariate datasets, with inequality constraints between variables. For example, Emery (2012) proposed transforming variables into Gaussian random fields using stepwise conditional transformation (Leuangthong and Deutsch, 2003), followed by co-simulation of transformed variables. Alternatively, projection pursuit multivariate transform (Barnett *et al.*, 2014) can be applied to variables changed into ratios (Arcari Bassani *et al.*, 2018). Another method involves changing data into variables free of inequality constraints using minimum/maximum autocorrelation factors (Desbarats and Dimitrakopoulos, 2000), and then simulating them independently (Abildin, 2019). Nevertheless, factorisation-based methods have their limitations, including poor performance in the case of heterotopic sampling, and unidentical marginal distributions (Madani and Abulkhair, 2020).

A new geostatistical algorithm is proposed in this paper, which was inspired by hierarchical cosimulation with an acceptance–rejection method (Madani and Abulkhair, 2020). Considering that the acceptance–rejection technique can be time-consuming depending on the dataset, the method was replaced with inverse transform sampling. As a result, the simulation speed was considerably faster, because inverse transform sampling resimulates rejected values in one iteration. Furthermore, this algorithm can potentially work with partially heterotopic datasets. For example, Abulkhair and Madani (2021) integrated heterotopic moving neighbourhood configurations, namely single and multiple searching strategies, in the proposed algorithm to replace collocated and multicollocated co-kriging. The objectives of this paper are as follows: 1) Provide a methodology of conventional hierarchical cosimulation (Almeida and Journel, 1994), and integrate inverse transform sampling into the second simulation. 2) Present a real case study from an iron deposit with an inequality constraint between iron and aluminum oxide. 3) Compare the proposed and conventional hierarchical co-simulations based on the reproduction of histograms, variograms, and bivariate relationships

METHODOLOGY

The proposed algorithm is based on hierarchical sequential Gaussian co-simulation (Almeida and Journel, 1994), which simulates the primary variable first, and then simulates each of the remaining variables, conditional on the previous simulations using simple collocated co-kriging. In this study, this algorithm was adapted for bivariate datasets with an inequality constraint between variables. To do so, the secondary variable could be resimulated in multiple iterations using an acceptance–rejection method (Madani and Abulkhair, 2020), or in one iteration using inverse transform sampling (Abulkhair and Madani, 2021). However, it is important to note that these algorithms resimulate only those values that lie outside an inequality constraint. The steps of the proposed hierarchical co-simulation integrated with inverse transform sampling are as follows:

1. The transformation of primary Y_A and secondary Y_B variables into their respective normal scores Z_A and Z_B. The primary variable should be the most prominent in the deposit, and comes first in the hierarchical order.

2. The definition of the simulation path (either regular or random) so that each target location x_0 is visited only once.

3. The determination of the Gaussian conditional cumulative distribution function (ccdf) at each node x_0 in order to obtain global statistical parameters of the primary variable. Simple co-kriging is used as an estimator for the co-simulation of n realisations for each transformed variable:

$$Z_A^n(x_0) = Z_{A_{SCK}}(x_0) + \sqrt{\sigma^2_{A\,SCK}(x_0)} \cdot U^n \qquad [1]$$

   where $Z_{A_{SCK}}$ is a simple co-kriging estimator, $\sigma^2_{A\,SCK}$ is its estimation variance, and U^n is an independent random value.

   Step 3 is applied to all grid nodes so that the primary variable will be available in the entire region.

4. Global statistical parameters of the secondary variable are obtained by determining the Gaussian ccdf at each node x_0. In this step, the simple collocated co-kriging SCCK is used for co-simulation, taking into account simulated results of the primary variable in addition to the hard data:

$$Z_B^n(x_0) = Z_{B_{SCCK}}(x_0) + \sqrt{\sigma^2_{B\,SCCK}(x_0)} \cdot U^n \qquad [2]$$

   where $Z_{B_{SCCK}}$ is a simple collocated co-kriging estimator and $\sigma^2_{B\,SCCK}$ is its estimation variance.

   Step 4 is looped to simulate the secondary variable for all grid nodes.

5. After step 4 is looped until all grid nodes are simulated, inverse transform sampling is used to generate random numbers within thresholds from an inequation. This step proceeds as follows:

   5.1. The back-transformation of simulated values $Z_A^n(x_0)$ (primary variable) to the original scale $Y_A^n(x_0)$.

   5.2. The determination of minimum and maximum thresholds according to the linear inequation between primary and secondary variables. Linear inequation can be negative, $Y_B \leq aY_A + b$, or positive, $Y_B \geq aY_A + b$, and thresholds are obtained in the following way:

$$\begin{cases} Y_{B\,min}^{n} = \min(Y_B) \;\; and \;\; Y_{B\,max}^{n} = aY_A^n(x_0) + b \;\; if \;\; inequation \; is \; negative \\ Y_{B\,min}^{n} = aY_A^n(x_0) + b \;\; and \;\; Y_{B\,max}^{n} = \max(Y_B) \;\; if \;\; inequation \; is \; positive \end{cases} \quad [3]$$

   where a is a slope, b is an intercept, $\min(Y_B)$ and $\max(Y_B)$ are the minimum and maximum values of the secondary variable in the original dataset.

   5.3. The transformation of $Y_{B\,min}^{n}$ and $Y_{B\,max}^{n}$ to normal scores $Z_{B\,min}^{n}$ and $Z_{B\,max}^{n}$ to identify simulated values of the secondary variable that lie outside this interval. Store identified values as m realisations.

   5.4. The implementation of inverse transform sampling to generate random numbers V^m within truncated thresholds $[Z_{B\,min}^{m}, Z_{B\,max}^{m}]$, for all values identified in Step 5.3:

$$V^m = F^{-1}(F(Z_{B\,min}^{m}) + \left(F(Z_{B\,max}^{m}) - F(Z_{B\,min}^{m})\right) \cdot U^m \quad [4]$$

   where F is the conditioned cumulative distribution function and F^{-1} is the quantile function.

   5.5. The resimulation of the values found in Step 5.3 using random numbers V^m generated through inverse transform sampling:

$$Z_B^m(x_0) = Z_{B_{SCCK}}(x_0) + \sqrt{\sigma_{B_{SCCK}}^2(x_0)} \cdot V^m \quad [5]$$

   5.6. Loop until all identified values are re-simulated.

6. Back-transformation of normal score simulated values Z_A^n and Z_B^n to the original scale of primary Y_A^n and secondary Y_B^n variables.

Different co-kriging configurations can be implemented in both steps of the hierarchical co-simulation. This study focused on simple and collocated co-kriging, as used in the original conventional algorithm.

RESULTS

Case study

The proposed algorithm was applied to a real case study of an iron deposit, which showed a strong inequality constraint between iron and aluminum oxide. The original data consisted of 608 sample points from exploration boreholes, and both iron and aluminum oxide grades were used as hard conditioning data (see Figure 1). However, there was a particular bias in the sampling strategy, as there were zones where boreholes are not available. Therefore, cell-declustering (Deutsch, 1989) was applied to resolve this issue using 800 m × 800 m × 80 m cell dimensions. The statistical parameters after declustering are presented in Table 1.

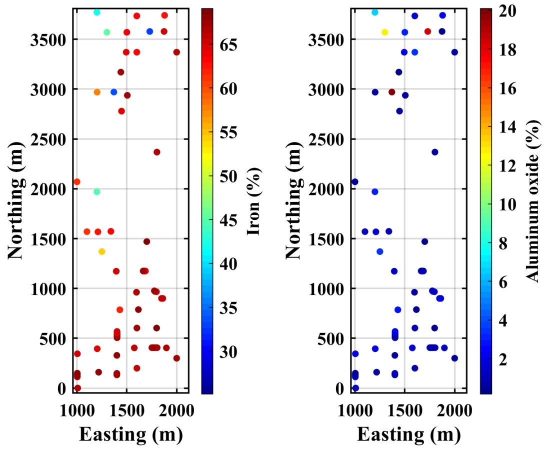


Figure 1. Location maps of iron and aluminum oxide grades.

Table 1. Declustered statistical parameters of the iron dataset

Parameter	Value (%)
Mean iron grade	63.77
Mean aluminum oxide grade	1.87
Standard deviation of iron	7.14
Standard deviation of aluminum oxide	3.70
Correlation coefficient	-0.81

Figure 2 shows a scatter plot between the primary variable, iron, and the secondary variable, aluminum oxide. Each variable's marginal distribution demonstrated that the deposit, in general, has a high iron content and low aluminum oxide grade. There was a sharp inequality constraint, which should be reproduced to obtain an accurate model of the deposit. The inequality constraint was characterised by an inequation with a slope $a = -0.89$ and an intercept $b = 62$:

$$Aluminum\ oxide = -0.89 \cdot Iron + 62 \qquad [6]$$

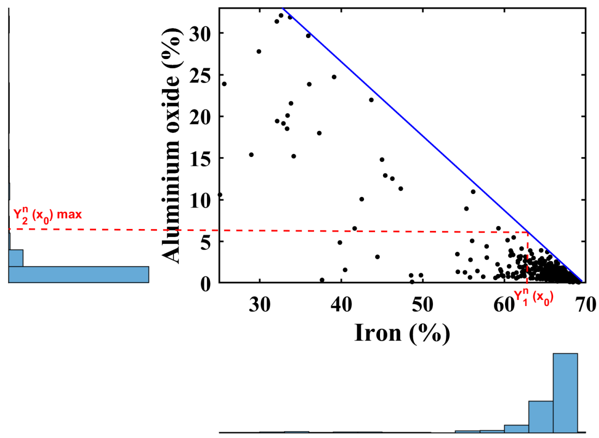


Figure 2. Scatter plot of iron and aluminum oxide, with their corresponding marginal distributions. Blue line: inequality constraint; red dashed line: example of obtaining truncated thresholds for inverse transform sampling.

The next step was to model direct and cross-variograms using variables transformed into normal scores. After checking the multi-Gaussianity assumption, it was concluded that this dataset was suitable to employ a co-simulation algorithm. It was decided to use an omnidirectional variogram, because no anisotropy was detected in the vertical and horizontal directions. As a result, theoretical variograms were manually fitted to obtain a two-structured linear model of co-regionalisation (see Figure 3).

$$\begin{pmatrix} \gamma_{Fe} & \gamma_{Fe/Al_2O_3} \\ \gamma_{Al_2O_3/Fe} & \gamma_{Al_2O_3} \end{pmatrix} = \begin{pmatrix} 0.68 & -0.49 \\ -0.49 & 0.69 \end{pmatrix} Sph(45m, 45m, 45m) + \begin{pmatrix} 0.23 & -0.25 \\ -0.25 & 0.34 \end{pmatrix} Sph(225m, 225m, 225m) \quad [7]$$

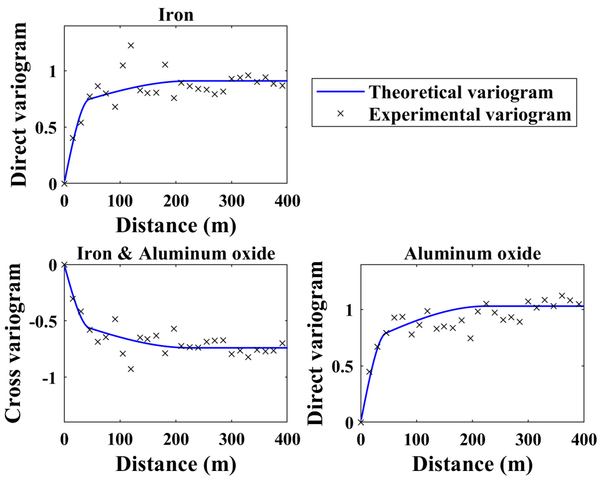


Figure 3. Direct and cross variograms of iron and aluminum oxide normal scores.

Co-simulation results

The proposed hierarchical co-simulation algorithm was assessed by comparing it to the conventional co-simulation algorithm. The fairness of this comparison is justified because the only difference between algorithms is the integration of inverse transform sampling. Using the block dimension of 15 m × 15 m × 15 m, 100 realisations of iron and aluminum oxide grades were simulated by both the proposed and conventional algorithms. E-type and standard deviation maps of aluminum oxide show that the proposed algorithm prohibits the overestimation of the secondary variable (see Figure 4).

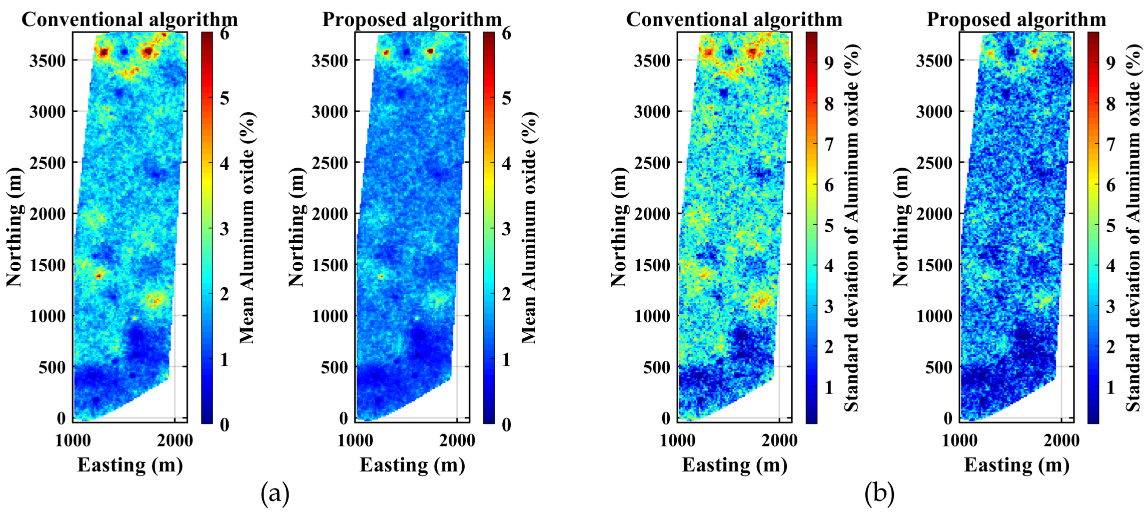


Figure 4. E-type (a) and standard deviation (b) maps of aluminum oxide over 100 realisations from conventional and proposed co-simulation algorithms.

The general statistical parameters of the produced realisations proved that inverse transform sampling helps to produce a higher correlation coefficient between variables (see Table 2). However, there was also an underestimation of aluminum oxide's mean and standard deviation values with the proposed algorithm. One of the main limitations of the proposed algorithm, as discussed by Madani and

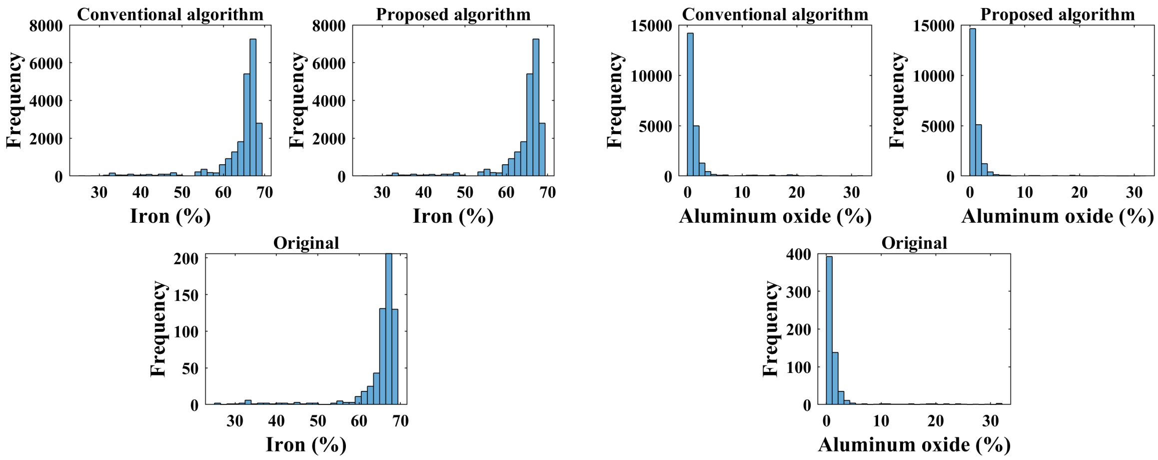
Abulkhair (2020), is the reproduction of variance of the secondary variable. The reason for this is still unknown at this point. Furthermore, the resimulation of the secondary variable using inverse transform sampling always decreases the mean and variance if the inequation is negative. One possible reason for this could be the biased sampling pattern in the original dataset, where most of the samples had high iron and low aluminum oxide grades. Nevertheless, conventional co-simulation overestimates aluminum oxide and produces results way above the inequality constraint, however, the simulated mean is still lower than the original.

Table 2. Reproduction of the mean, standard deviation and correlation coefficient by the conventional and proposed cosimulation algorithms

Parameter	Conventional	Proposed	Original
Mean: iron (%)	63.99	63.99	63.77
Mean: aluminum oxide (%)	1.95	1.50	1.87
Standard deviation: iron (%)	6.68	6.68	7.14
Standard deviation: aluminum oxide (%)	3.89	2.75	3.70
Correlation coefficient	–0.61	–0.7	–0.81

Statistical validation

Validation of geostatistical models is based on two significant factors: histogram and variogram reproduction. Histogram validation illustrates how well the model reproduces the marginal distribution of the original dataset. At the same time, variogram validation shows the reproduction of local statistics or spatial continuity. Figure 5 compares histograms from the first realisation simulated by the conventional and proposed algorithms with the original dataset. The histograms of both iron and aluminum oxide were well reproduced, and integration of inverse transform sampling did not affect the shape of the marginal distribution.



(a) (b)

Figure 5. Histogram reproduction of iron (a) and aluminum oxide (b) simulated by the conventional and proposed co-simulation algorithms.

However, considering the skewness of aluminum oxide's marginal distribution, histograms from Figure 5b are difficult to compare. Therefore, distributions of realisations produced by both algorithms were compared to the original distribution of aluminum oxide through a quantile-quantile (Q–Q) plot (see Figure 6). The logarithmic scale was used to demonstrate Q–Q plots, because of the scarcity of data in higher quantiles and most original aluminum oxide samples having a grade of 0–4 %. As a result, realisations fit the diagonal lines up to the grade of 4%, meaning that data distributions were correctly reproduced. The difference between algorithms is evident in the upper quantiles, which can be explained by the scarcity of data.

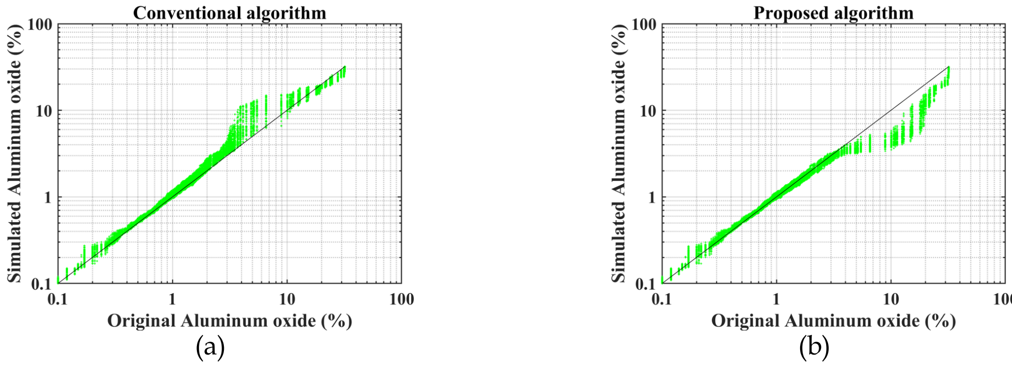


Figure 6. Quantile–quantile plots between original and simulated aluminum oxide grades for the conventional (a) and proposed (b) algorithms. Green points: realisations; black line: identity line.

The spatial correlation structure can be assessed by modelling direct, and cross variograms of each realisation and comparing the results with experimental variograms (see Figure 7). Both algorithms showed promising results in terms of local statistics reproduction. The variability was higher in direct variograms of aluminum oxide from conventional simulations, which has also been demonstrated in global statistics. On the other hand, cross variograms were better reproduced by the proposed algorithm.

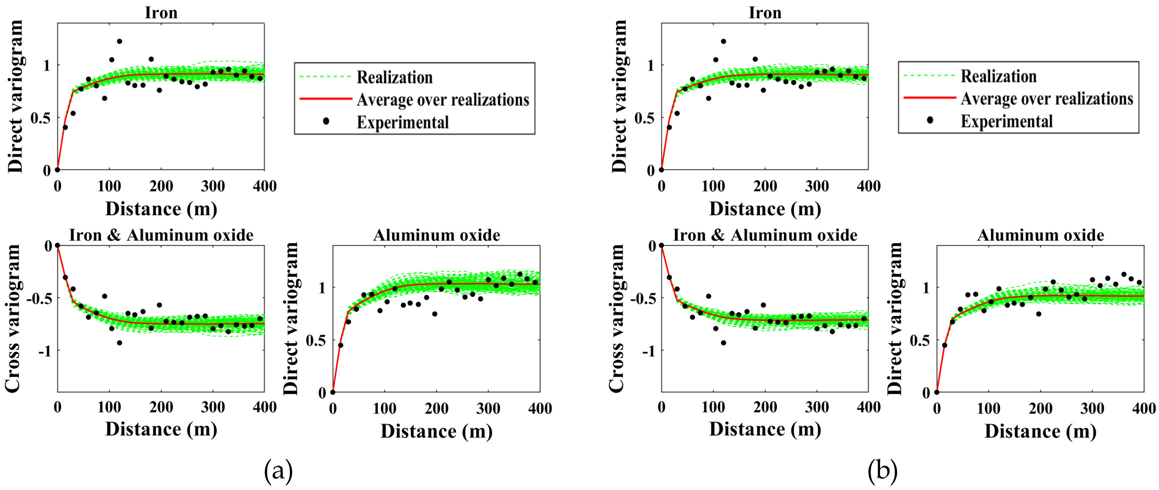


Figure 7. Reproduction of standardised direct and cross variograms produced by the conventional (a) and proposed (b) co-simulation algorithms.

Finally, the validation of the bivariate relationship between iron and aluminum oxide is shown through scatter plots (see Figure 8). It can be seen that the conventional co-simulation algorithm failed to reproduce the bivariate relationship. While individual statistics of each variable showed promising results, only the proposed algorithm was able to reproduce an inequality constraint without significant loss in terms of other statistical validations. Inability to reproduce inequality constraints can lead to overestimation of disturbing elements, negatively affecting mine planning results.

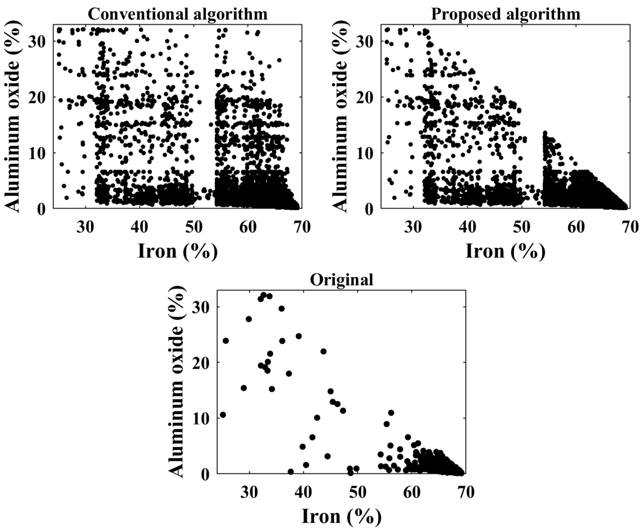


Figure 8. Reproduction of scatter plots between iron and aluminum oxide by the conventional and proposed co-simulation algorithms.

CONCLUSION

An updated geostatistical algorithm based on hierarchical co-simulation is presented in this paper, which integrates inverse transform sampling to model bivariate datasets with inequality constraints. The benefits of this type of algorithm were demonstrated through comparison with conventional hierarchical co-simulation. For this purpose, the conventional and proposed co-simulation algorithms were applied to a case study of an iron deposit, with an inequality constraint between iron and aluminum oxide. Compared to conventional co-simulation, the proposed algorithm was able to reproduce an inequality constraint between iron and aluminum oxide. Nevertheless, one of the many concerns before conducting this study was the effect of inverse transform sampling on the marginal distribution of the secondary variable. This study showed that the shape of the marginal distribution is not significantly affected by the resimulation of the secondary variable within truncated thresholds. Furthermore, simulated direct and cross-variograms fit in well with the experimental variogram points.

However, both algorithms considerably underestimated aluminum oxide in terms of mean and standard deviation. One possible reason for this is the nature of this dataset, which had quite a biased sampling pattern, with the majority of samples coming from zones with high iron and low aluminum oxide grades. Moreover, conventional co-simulation methodologies reproduce bivariate relationships only, based on the cross-correlation structure. Therefore, highly skewed marginal distributions, and moderate correlation coefficients can result in simulated points being spread out even more. Alternatively, a similar case study with a high correlation coefficient (–0.95) and more uniform marginal distributions showed far better simulation results, without severe underestimation of the secondary variable (Madani and Abulkhair, 2020).

The proposed algorithm can be significantly improved in several ways: 1) Multiple and multicollocated searching strategies can be implemented as far better options than single, isotopic, and collocated searches (Madani and Emery, 2019). 2) The algorithm can be extended to multivariate cases with more than two variables, so that each variable can be simulated hierarchically conditional to previous simulations. 3) Inverse transform sampling can also be used with non-linear inequations, instead of only linear inequality constraints. Nevertheless, the proposed methodology still needs to be further researched to improve the reproduction of basic statistical parameters (i.e., mean and variance).

ACKNOWLEDGEMENTS

We are grateful to Nazarbayev University, Kazakhstan for funding this work via Faculty Development Competitive Research Grants for 2018–2020, under Contract No. 090118FD5336. The authors are also particularly thankful to the Geovariances Company for providing the iron dataset.

REFERENCES

Abildin, Y., Madani, N. and Topal, E. (2019). A Hybrid Approach for Joint Simulation of Geometallurgical Variables with Inequality Constraint. *Minerals*, 9(1), 24.

Abulkhair, S. and Madani, N. (2021). Assessing Heterotopic Searching Strategy in Hierarchical Cosimulation Framework for Modeling the Variables with Inequality Constraints. *Comptes Rendus. Géoscience*, 353(1), 115-134.

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

Arcari Bassani, M.A., Costa, J. and Deutsch, C.V. (2018). Multivariate geostatistical simulation with sum and fraction constraints. *Applied Earth Science*, 127(3), 83-93.

Barnett, R.M., Manchuk, J.G. and Deutsch, C.V. (2014). Projection Pursuit Multivariate Transform. *Mathematical Geosciences*, 46, 337-359.

Davis, B.M. and Greenes, K.A. (1983). Estimation using spatially distributed multivariate data: An example with coal quality. *Journal of the International Association for Mathematical Geology*, 15, 287-300.

Desbarats, A.J. and Dimitrakopoulos, R. (2000). Geostatistical Simulation of Regionalized Pore-Size Distributions Using Min/Max Autocorrelation Factors. *Mathematical Geology*, 32, 919-942.

Deutsch, C.V. (1989). DECLUS: a fortran 77 program for determining optimum spatial declustering weights. *Computers & Geosciences*, 15(3), 325-332.

Emery, X. (2008). A turning bands program for conditional co-simulation of cross-correlated Gaussian random fields. *Computers & Geosciences*, 34(12), 1850-1862.

Emery, X. (2012). Co-simulating Total and Soluble Copper Grades in an Oxide Ore Deposit. *Mathematical Geosciences*, 44, 27-46.

Emery, X. and Ortiz, J.M. (2012). Enhanced coregionalization analysis for simulating vector Gaussian random fields. *Computers & Geosciences*, 42, 126-135.

Goovaerts, P. (1997). *Geostatistics for Natural Resource Evaluation*. Oxford University Press, New York.

Leuangthong, O. and Deutsch, C.V. (2003). Stepwise Conditional Transformation for Simulation of Multiple Variables. *Mathematical Geology*, 35, 155-173.

Madani, N. and Abulkhair, S. (2020). A hierarchical cosimulation algorithm integrated with an acceptance–rejection method for the geostatistical modeling of variables with inequality constraints. *Stochastic Environmental Research and Risk Assessment*, 34, 1559-1589.

Madani, N. and Emery, X. (2019). A comparison of search strategies to design the cokriging neighborhood for predicting coregionalized variables. *Stochastic Environmental Research and Risk Assessment*, 33, 183–199.

Pyrcz, M.J. and Deutsch, C.V. (2014). *Geostatistical Reservoir Modeling*. Oxford University Press, New York.

Verly, G.W. (1993). Sequential Gaussian Cosimulation: A Simulation Method Integrating Several Types of Information. In A. Soares (Ed.), *Geostatistics Tróia '92. Quantitative Geology and Geostatistics*, vol 5 (pp. 543-554). Springer, Dordrecht.

Wackernagel, H. (2003). *Multivariate geostatistics*. Springer, Berlin.

Sultan Abulkhair

Research Assistant
Nazarbayev University

Sultan Abulkhair is currently a Research Assistant at the School of Mining and Geosciences, Nazarbayev University. His primary research areas include multivariate geostatistics, multiple-point statistics, stochastic mine planning and rapid updating of resource models. He obtained an MSc degree in Mining Engineering and graduated with honors from Nazarbayev University, Nur-Sultan, Kazakhstan. From 2015 to 2019, Mr. Abulkhair was a student at Kazakh-British Technical University, Almaty, Kazakhstan, where he received a BEng in Geology. Recently he was accepted as a PhD student at the University of Adelaide, Adelaide, Australia.