SPEEDING UP CONDITIONAL SIMULATION: USING SEQUENTIAL GAUSSIAN SIMULATION WITH RESIDUAL SUBSTITUTION

Alejandro Cáceres

Geoinnova Consultores, Chile

Xavier Emery

Department of Mining Engineering, University of Chile, Chile

ALGES Laboratory, Advanced Mining Technology Center, University of Chile, Chile

Marcelo Godoy

Golder Associates, Chile

ABSTRACT

Sequential Gaussian simulation (SGSIM) and turning bands simulation (TBSIM) are widely used to generate realisations of mineral grades and to evaluate mineral resources. SGSIM relies on the recursive application of Bayes theorem and is designed as a direct conditional method. TBSIM is based on a stereological device that generates non-conditional realisations, which are subsequently conditioned through residual substitution kriging. SGSIM uses a random path to visit the nodes targeted for simulation. In practice, for each realisation the path has to be changed in order to avoid artefacts or artificial correlation between realisations. This is clearly the case when conditional simulation is performed. However, for non-conditional simulation with a proper implementation, the resulting realisations do not exhibit such artificial behaviour.

In this context, the proposed algorithm (SGSIM-RS) uses the sequential approach to generate non-conditional realisations, with a single controlled path. Then, a conditioning step is performed using residual substitution as it is done in the TBSIM approach. The advantage of using a single path is the possibility to generate many realisations at the same time and to condition them in a single kriging step. The CPU time reduction is considerable. For example, the time to create n realisations is equivalent to the time of one sequential simulation plus one conditioning kriging. The counterpart is the memory needed for storing the non-conditional realisations all together. However, this requirement is less demanding with actual hardware and operative systems. The proposed algorithm is presented through a case of study and its performance is compared to the traditional SGSIM and TBSIM approaches.

INTRODUCTION

Running time and management of large files involved in simulation studies could be discouraging factors for incorporating conditional simulation as a daily basis practice in the mining industry and

become critical when large simulation grids are used [1]. The most widespread algorithms for Gaussian conditional simulation are sequential Gaussian (SGSIM) [2, 3, 4, 5, 6] and turning bands (TBSIM) [7, 8] simulation, with available implementations in opensource projects and comercial software.

Both algorithms rest upon the multi-Gaussian model and the homoscedasticity property of the Gaussian distribution, together with the orthogonality of simple kriging. TBSIM directly uses those properties by separating the problem in two steps: simulating first a non-conditional Gaussian random field ($Y(x)$) and then conditioning to the data using the residual substitution (RS) approach [7]. RS is applicable to convert non-conditional simulations into conditional ones. In contrast, the SGSIM algorithm [3], making use of screening effects, search strategy, node migration, visiting sequence and multiple grids, among other considerations, directly derives a conditional distribution at each target location x, from which a simulated value is drawn as follows:

$$Ycs(x) = Y_{sk}(x) + \sigma_{sk}(x) * U \quad (1)$$

where $Y_{sk}(x)$ is the simple kriging estimate of $Y(x)$ calculated from the original data and previously simulated nodes, $\sigma_{sk}(x)$ is the associated kriging standard deviation, and U is an independent Gaussian random variable.

An attractive feature of SGSIM is its ability to directly provide conditional simulations, avoiding the two-step approach used in TBSIM [3]. However, the cost of this remarkable feature is the requirement of whole re-simulation if new data are added or removed, while TBSIM can be updated by just adding or removing the data in the conditioning step.

A method that allows faster simulation and can easily manage the update of new drilling campaigns or removing certain data would be beneficial to practitioners in the mining industry. This paper presents such a method that uses the sequential Gaussian algorithm for generating non-conditional simulations and the residual substitution approach for conditioning to sample data.

SEQUENTIAL SIMULATION ALGORITHMS

Traditional Sequential Gaussian Simulation (SGSIM)

Sequential Gaussian simulation uses a random visiting order for the nodes targeted for simulation. This visiting sequence, often called “random path”, is changed from realisation to realisation in order to avoid artificial correlation or similarity between realisations. This is clearly valid when conditional simulation is performed, because for every node the same conditioning data locations and original data values are used to determine the local distribution of the value to simulate. The implementation of the sequential Gaussian approach has therefore two sources of randomness: a theoretical one related to the Gaussian value U used in equation 1 and the random visiting sequence as an implementation aspect.

If non-conditional simulation is performed with SGSIM, the very first nodes (for which there is no conditioning data) are simulated from a Gaussian distribution without any covariance or spatial consideration. Then, simulated values are available and the procedure goes on using these first simulated values as conditioning data, i.e., non-conditional simulation in SGS becomes a simulation conditional to these first nodes.

By construction, the non-conditional values simulated at the first nodes are independent from realisation to realisation, so the use of a changing visiting sequence for each realisation can be avoided. Therefore several non-conditional realisations can be generated in a single execution of the sequential algorithm using an unique visiting sequence. This approach has already been suggested [9] and the use of deterministic or modified visiting sequences been explored [10, 11, 12, 13].

Proposed Methodology (SGSIM-RS)

The global procedure to get conditional realisations using the SGSIM-RS approach is:

- Normal score transformation of the raw data into Gaussian values $Y(x_i), i = 1 \dots n$.
- Variogram analysis of the normal scores data, defining a covariance model $C_Y(h)$ (or, equivalently, a variogram model)
- Non-conditional Gaussian simulation using a single-path sequential approach. Get the residuals between the Gaussian data values and every realisation at the data locations:

$$R^k(x_i) = Y(x_i) - Y_{NC}^k(x_i) \tag{2}$$

  where $Y_{NC}^k(x)$ is the *k*-th non-conditional realisation of the Gaussian random field $Y(x)$.
- Estimation of the residuals over the domain of interest by simple kriging (*sk*), using the covariance model $C_Y(h)$. Because the weigths are the same in all the realisations, the residual estimates are simultaneously obtained in a single kriging run.
- Addition of the residual estimates and the corresponding non-conditional realisation to generate the conditional realisation $Y_C^k(x)$ over the domain:

$$Y_C^k(x) = Y_{NC}^k(x) + R_{sk}^k(x) \tag{3}$$

- Back-transformation of the conditional Gaussian realisations to the original values.

The non-conditional Gaussian simulation step is detailed below:

- Define a single visiting sequence, which can be achieved using a low discrepancy sequence [7], a regular sequence or just a random sequence with a multiple grid approach [13].
- With the LU algorithm, generate several (n) realisations for the first thousand nodes of the visiting sequence. These simulated values, which will be used as conditioning data for the subsequent nodes, account for the covariance model and are by construction independent from one realisation to another.
- Continue with the sequential approach and generate n non-conditional realisations ($\vec{Y}_{NC}$) in each visited node x:

$$\vec{Y}_{NC}(x) = \vec{Y}_{sk}(x) + \sigma_{sk}(x) * \vec{U} \tag{4}$$

  where $\vec{Y}_{sk}(x)$ stands for the vector of simple kriging estimates for the n realisations given the previously simulated nodes, $\sigma_{sk}(x)$ for the simple kriging standard deviation, and $\vec{U}$ for an independent Gaussian random vector. This way, all the realisations can be generated simultaneously.

As an example, for a domain of 100 × 100 nodes, one hundred non-conditional realisations are generated using an isotropic spherical variogram of range 20. Figure 1 (left) presents the probability intervals of the simulated values at each node as a function of the sequence order: it is remarkable that the intervals are almost the same for all the nodes. To give another look at this feature, the conditional variance is calculated at each node and plotted as a function of the sequence order (Figure 1, right) without exhibiting any trend or artificial pattern.

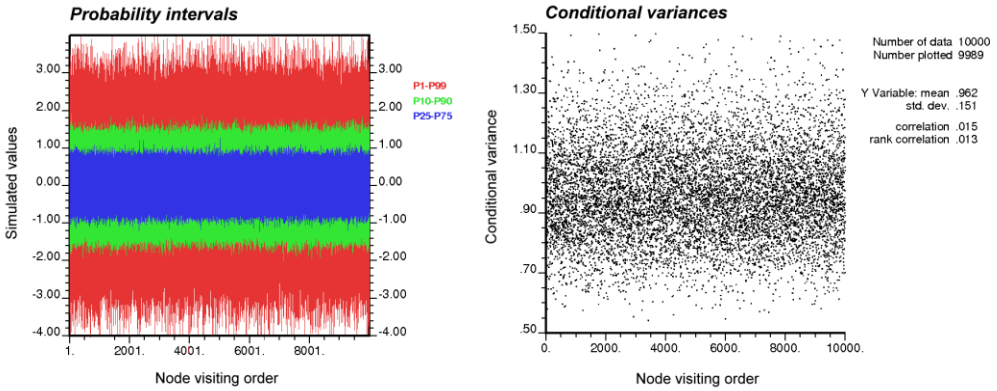

Figure 1: Probability intervals (left) and conditional variances (right) as a function of the visiting order

APPLICATION TO A MINING DATASET

Presentation of the Case Study

The area under study is part of the Río Blanco – Los Bronces porphyry copper deposit [14], a breccia complex located in the Chilean central Andes. A set of 2376 diamond drill hole samples, located in a 400 × 600 × 130 m^3 volume, are available with information on total copper grades. Figure 2 presents the available data coloured by copper content.

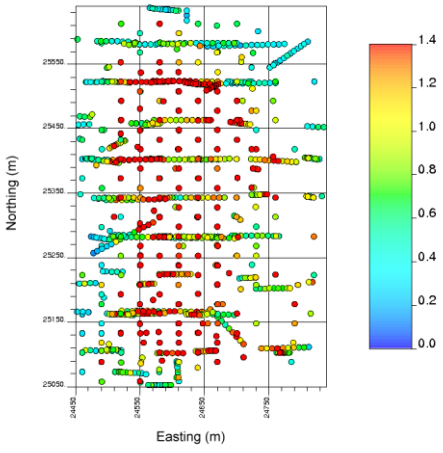

Figure 2: Location map of copper grade data

Simulation Approaches

Three different algorithms are compared by simulating the copper grades over a 2D regular grid of 390 × 600 nodes with a 1 × 1 m spacing:

- *Sequential Gaussian simulation (SGSIM):* this approach is performed using the SGSIM program of the GSLIB package [3], using multiple grids and migration of data to nodes.
- *Sequential Gaussian simulation with residual substitution (SGSIM-RS*): a single path is used with multiple grids and a uniformly random ordering of the nodes of each grid.
- *Turning bands simulation (TBSIM):* this approach is performed in the ISATIS software, using 1000 turning lines.

The same search radius (250 m) and number of conditioning data (16) are considered for each method. Simple kriging is used for non-conditional simulation and conditioning step in SGSIM-RS, and in the conditioning steps of TBSIM and SGSIM. There is no loss of data in the migration to the nodes in SGSIM and SGSIM-RS, so the effective datasets are the same in each method. The same normal score transformation, back-transformation and isotropic variogram model (Table 1) are used in each method to avoid differences due to implementation.

Table 1: Isotropic variogram model for transformed copper grades

Structure	Range (m)	Sill contribution
Nugget	-	0.12
Spherical	112	0.7
Exponential	416	0.18

Comparison Between Methods

In the following subsections, the resulting realisations are compared in several ways. First of all, the visual inspection of several realisations of SGSIM-RS does not indicate the presence of any artefact or strange pattern. Even more, it is impossible to distinguish if the realisation comes from SGSIM, SGSIM-RS or TBSIM (Figure 3).

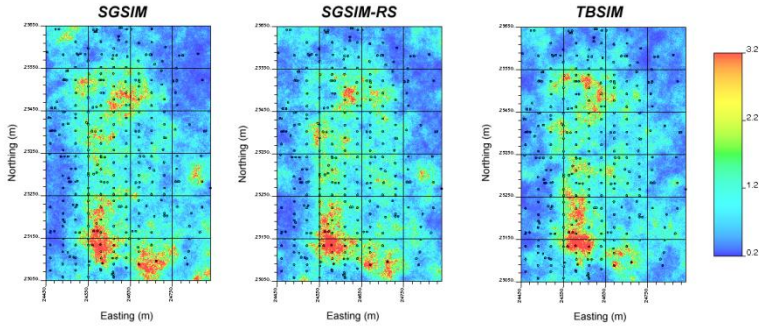

Figure 3: Examples of realisations

Basic statistics: Figure 4 (left) shows the distributions of the average copper grade per realisation for each algorithm. These distributions are close to each other, although that of SGSIM-RS presents slightly higher values. The variance per realisation is also indicated in Figure 4 (right): SGSIM-RS and SGSIM have similar distributions, while TBSIM shows a slightly wider range of variances.

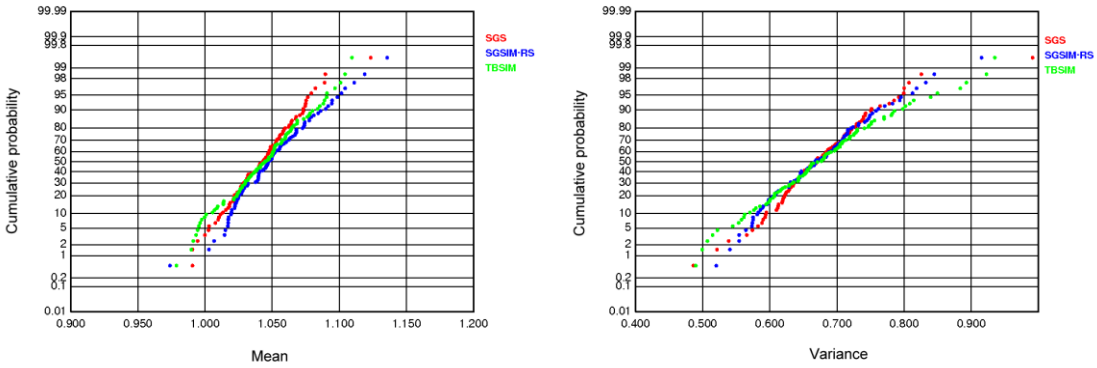

Figure 4: Distribution of the average (left) and variance (right) of simulated copper grades per realisation (right)

Grade-tonnage curves: the tonnage and grades above a set of cutoff grades are calculated on each realisation. The expected curves are presented in Figure 5 (left), showing virtually the same values for all the algorithms. In counterpart, Figure 5 (right) presents the widths of the 80% confidence intervals for the grades and tonnages by cutoff. It is seen that the turning bands algorithm presents a higher variability in grades for almost every cutoff, whereas SGSIM-RS and SGSIM exhibit a similar behaviour for this measure.

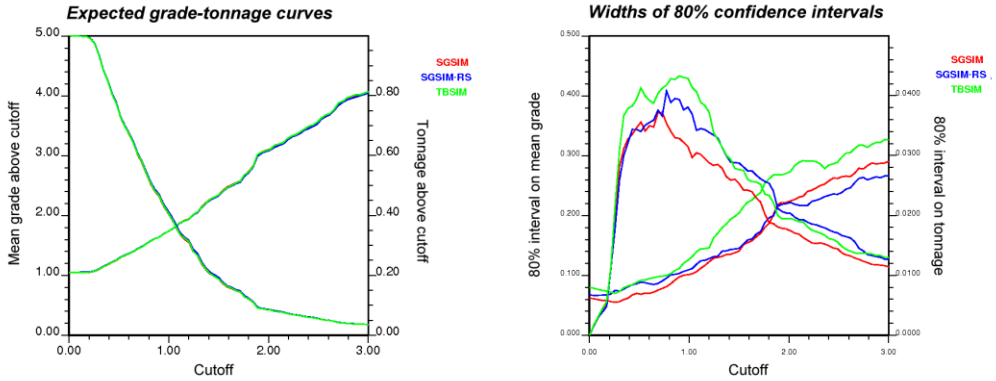

Figure 5: Expected grade-tonnage curves (left) and widths of 80% confidence intervals for grade-tonnage curves (right)

Local expectation and uncertainty measures: Figure 6 (left) shows the distributions of the conditional expectation (mean of realisations) calculated at each node. The curves are almost identical for the three algorithms. The conditional variance distributions (Figure 6, right) are close for SGSIM and SGSIM-RS, whereas TBSIM shows a 5% of the nodes with higher values.

Figure 7 displays the conditional coefficient of variation. The high and low zones are located in the same parts of the domain for all the algorithms. However, extreme values are notorious in the outer part of the deposit for SGSIM-RS and TBSIM, which is caused by the lack of conditioning data in this sector.

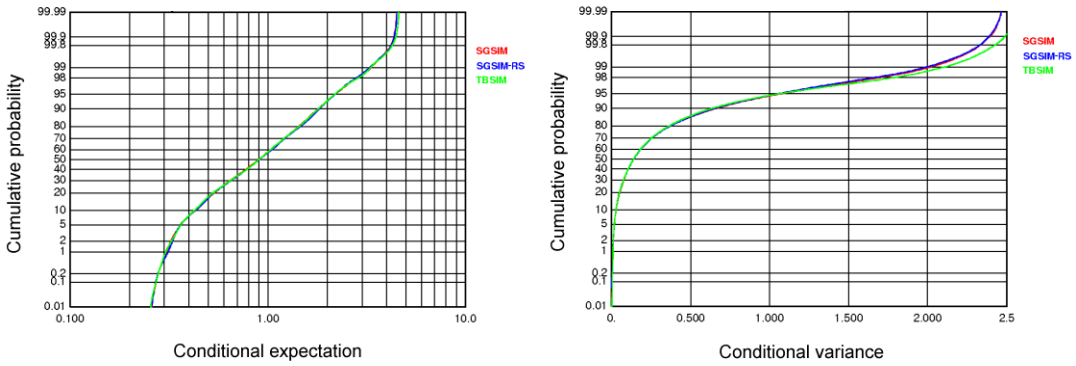

Figure 6: Distributions of the conditional expectation (left) and conditional variance (right)

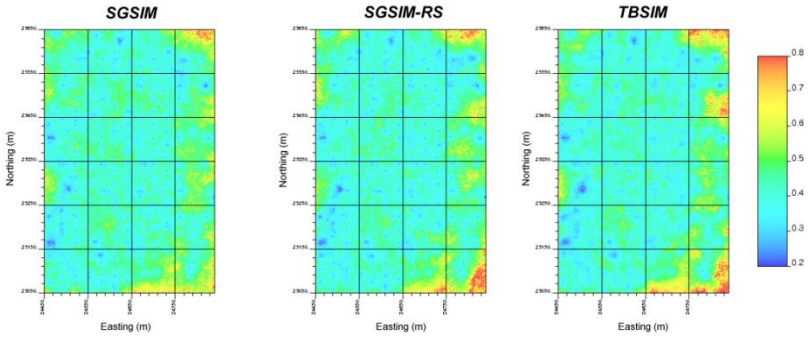

Figure 7: Conditional coefficients of variation

Performance Evaluation

Given a set of 20,893 blast hole data in which the true copper grades are known, a validation model is created by ordinary kriging on 5 × 5 m blocks in the bench where the copper grade realisations have been generated. Every realisation is regularised to this block size and compared against the validation model.

Error distribution: The percentual mean error (PME) is defined as:

$$PME(k) = \frac{1}{nb}\sum_{i=1}^{nb} 100 * \frac{[S^k(i) - R(i)]}{R(i)} \quad (5)$$

where $S^k(i)$ is the simulated grade at block i, $R(i)$ is the grade of the validation model and nb is the number of blocks where both models are defined. Figure 8 (left) presents the distributions for the three algorithms. The differences are marginal, although TBSIM performs slightly better than the other two algorithms.

Grade correlation distribution: The linear correlation coefficient between every realisation and the validation model is calculated. Figure 8 (right) shows a close correlation distribution between the algorithms.

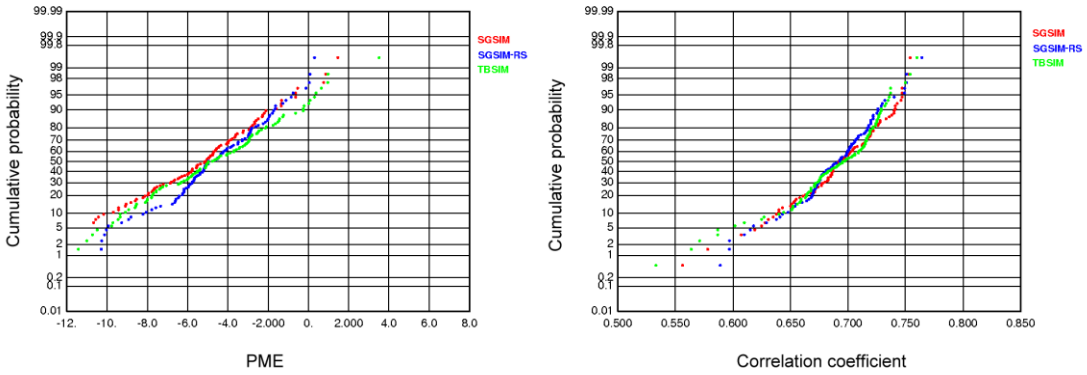

Figure 8: PME distribution (right) and correlation distribution (left)

Accuracy plot: Given an interval of probability p in the distribution of simulated block grades, the effective percentage of blocks of the validation model that fall within the interval is calculated. The closest this percentage to p, the better is the model in terms of accuracy and precision [15]. Figure 9 (left) displays the accuracy plots, showing a close performance for the three algorithms.

Destination mismatch: The percentage of correctly classified blocks (mill/dump) for several cutoff grades is calculated for each realisation. The expected percentages for each algorithm and cutoff are indicated in Figure 9 (right). SGSIM-RS presents a sligthly lower performance than the other two algorithms.

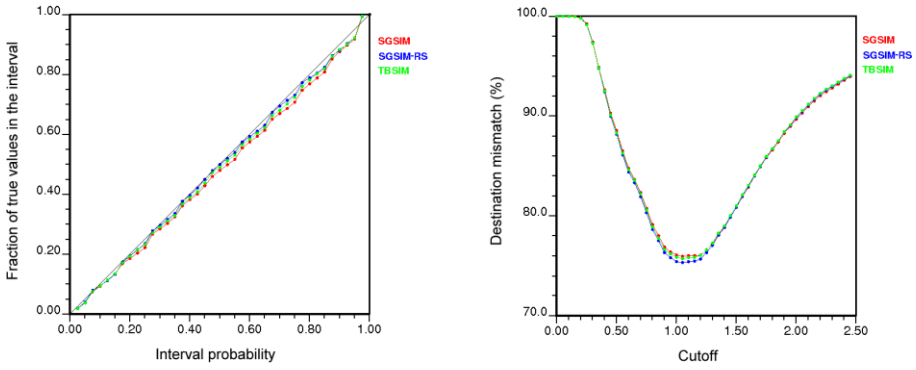

Figure 9: Accuracy plot (left) and destination match percentage (right)

CPU time by algorithm: each method has been executed in a 2.16 GHZ Core2Duo computer under the same conditions. The gain of CPU time of SGSIM-RS against the other methods is evident: the time for SGSIM-RS to write out 100 conditional realisations (280 s) is 20% that of SGSIM (1430 s) and 4.3% that of TBSIM (6,500 s).

The time reduction of SGSIM-RS increases with the number of realisations to generate. Obviously this reduction is restricted by the amount of available RAM memory. For example, the memory required to store a floating array of 8,000,000 elements, equivalent to a simulation grid of 200 × 200 × 200 nodes, is about of 30 mega bytes. Therefore with 3 available Gb, one hundred realisations can be done in a single execution of SGSIM-RS. The higher elapsed time of TBSIM is probably related

to the search strategy used in conditioning kriging (no migration of data to grid nodes; no use of spiral search as in sequential algorithms).

DISCUSSION

About the Use of a Single Visiting Sequence

Even for non-conditional simulation, the choice of the visiting sequence is not simple. It is clear that the use of a regular ordering, for example by cycling along the grid axes, can create artefacts that are likely to increase with higher-dimensional domains. There is therefore an open space of research to design visiting sequences that consider aspects such as the variogram model, data locations and geometry of the domain, in order to ensure the best reproduction of model statistics.

Potential Improvements of SGSIM-RS

In the non-conditional step, it is possible to use the LU algorithm instead of estimating the local distributions by kriging [16]. LU simulation can generate several realisations at each node, so the principles of the algorithm remain equal. However, the LU simulation of groups of nodes [17] can considerably increase the speed of the already fast proposed approach.

Another improvement is the possibility to use ordinary kriging in the conditioning step, when the mean value of the Gaussian random field is deemed uncertain (cases of local stationarity, when the mean varies slowly in space), while still using of simple kriging to construct the non-conditional realisations. In general, ordinary kriging is not recommended in SGSIM because of the design of the kriging neighbourhood (the screening effect is often partial for nodes on dense simulation grids), yielding a poor reproduction of the model statistics [3]. In the same vein, SGSIM-RS allows for the separate simulation of the nugget effect in the non-conditional step, which is of interest because the nugget effect is also often responsible for a poor screening effect.

Advantages of SGSIM-RS

SGSIM-RS inherits from the residual substitution and conditioning kriging the ability to update the realisations to new data: there is no need to recalculate the non-conditional simulations to add or to remove data. Also SGSIM-RS does not suffer from information losses due the migration procedure of SGSIM: if two or more data are associated with the same node, all the samples can be kept because they have different residuals, which can be used in the conditioning step.

CONCLUSIONS

A simple sequential algorithm to obtain conditional Gaussian simulations has been presented. It allows for generating realisations faster than the traditional SGSIM and TBSIM algorithms. The comparisons and validation results indicate no significant differences in terms of accuracy and reproduction of statistics between the proposed method and the traditional ones. However CPU time reduction allows for generating more realisations in the same amount of time, hence resulting in a better evaluation of transfer functions and uncertainty measures.

ACKNOWLEDGEMENTS

Thanks are due to R. Riquelme and S. Khosrowshahi for their comments on an earlier version of this paper.

REFERENCES

Godoy, M. (2003) *The Effective Management of Geological Risk in Long-Term Production Scheduling of Open Pit Mines*. PhD thesis, The University of Queensland, 256 pp. [1]

Alabert, F. (1987) *Stochastic Imaging of Spatial Distributions Using Hard and Soft Information*. Master's thesis, Stanford University, Department of Applied Earth Sciences, 197 pp. [2]

Deutsch, C.V., & Journel, A.G. (1998) *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 369 pp. [3]

Gómez–Hernández, J.J., & Journel, AG. (1993) *Joint Sequential Simulation of Multigaussian Fields*. In: Soares A (ed.) Geostatistics Troia'92. Kluwer Academic, Dordrecht, pp. 85-94. [4]

Journel, A.G. (1994) *Modeling Uncertainty: Some Conceptual Thoughts*. In: Dimitrakopoulos, R. (ed.), Geostatistics for the Next Century. Kluwer Academic, Dordrecht, pp. 30-43. [5]

Ripley, B.D. (1987) *Stochastic Simulation*. Wiley, New York, 237 pp. [6]

Chilès, J.P., & Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 695 pp. [7]

Matheron, G. (1973) *The Intrinsic Random Functions and Their Applications*. Advances in Applied Probability 5, pp. 439-468. [8]

Goovaerts, P. (1997) *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 491 pp. [9]

Isaaks, E.H. (1990) *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data*. PhD thesis, Stanford University, 213 pp. [10]

Richmond, A.J. (1998) *Multi-Scale Ore Texture Modelling for Mining Applications*. Master's thesis, The University of Queensland, 190 pp. [11]

Richmond, A.J., & Dimitrakopoulos, R. (2001) *Evolution of a Simulation: Implications for Implementation*. In: Kleingeld, W.J., and Krige, D.G. (eds.) Geostats 2000 Cape Town. Geostatistical Association of Southern Africa, Johannesburg, pp. 134-144. [12]

Tran, T.T. (1994) *Improving Variogram Reproduction on Dense Simulation Grids*. Computers and Geosciences 20(7), pp. 1161-1168. [13]

Serrano, L., Vargas, R., Stambuk, V., Aguilar, C., Galeb, M., Holmgren, C., Contreras, A., Godoy, S., Vela, I., Skewes, M.A., & Stern, C.R. (1996) *The Late Miocene to Early Pliocene Río Blanco-Los Bronces Copper Deposit, Central Chilean Andes*. In: Camus, F., Sillitoe, R.H., and Petersen, R. (eds.) Andean Copper Deposits: New Discoveries, Mineralizations, Styles and Metallogeny. Society of Economic Geologists, Special Publication no. 5, Littleton, Colorado, pp. 119-130. [14]

Deutsch, C.V. (1997) *Direct Assessment of Local Accuracy and Precision*. In: Baafi, E.Y, Schofield, N.A. (eds.) Geostatistics Wollongong'96. Kluwer Academic, Dordrecht, pp. 115-125. [15]

Davis, M.W. (1987) *Production of Conditional Simulations Via the LU Triangular Decomposition of the Covariance Matrix*. Mathematical Geology 19(2), pp. 91-98. [16]

Dimitrakopoulos, R., & Luo, X. (2004) *Generalized Sequential Gaussian Simulation on Group Size v and Screen-Effect Approximations for Large Field Simulations*. Mathematical Geology 36(5), pp. 567-591. [17]