Hierarchical Indicator Simulation

J. Ortiz C.[1] *and C. V. Deutsch*[1]

ABSTRACT

Indicator simulation allows specification of different continuity of high and low values through the use of grade thresholds. The sequential implementation of indicator simulation has been widely used, but the problem of order relation deviations remains a significant drawback. A new procedure for "Hierarchical Indicator Simulation" is introduced that avoids order relations and offers greater flexibility than the conventional sequential approach. The new implementation simulates all locations at the largest threshold and then considers decreasing thresholds in turn. All nodes with grades that are above the previous threshold are excluded from the domain of the current threshold and therefore, they are not used as conditioning data. The probability of being above or below the threshold is calculated using simple indicator kriging using all conditioning data in the domain, thus the covariance structure is preserved. The theoretical framework of this new approach is fully developed. Practical implementation details are discussed. Application to a copper deposit shows the advantages and drawbacks of the method. Finally, the possibility of incorporating multiple-point statistics is discussed.

INTRODUCTION

Stochastic simulation aims to reproduce the spatial patterns observed from samples. The better the reproduction of these features, the better the decisions that will be made based on the numerical models. The characterization of the spatial continuity of the variables is often done using two-point statistics such as the variogram or covariance function.

Gaussian simulation methods such as turning bands and sequential Gaussian simulation rely on the multivariate Gaussian assumption, that is, all bivariate, trivariate, and higher statistics are deemed to be multivariate Gaussian. If the assumption is correct, then these methods fully characterize the multivariate distribution of the variable with only two-point statistics required from the data. However, variables in the Earth Sciences are not multivariate Gaussian. A transformation is applied to enforce univariate Gaussianity, although bivariate and higher order distributions cannot be enforced through transformation and are usually not even checked. Any departure of the real data from the multivariate Gaussian model leads to numerical models that may be unsuitable for decision-making.

The non-parametric formalism of indicators was introduced in 1982 by A. G. Journel. It provides a richer approach that allows us to model variables that depart noticeably from the bivariate Gaussian assumption (Journel 1982; Journel, 1983). Sequential indicator simulation avoids the need of a bi-Gaussian assumption. It permits the random variable to have different spatial continuity for high and low values; the Gaussian formalism requires the spatial continuity to be symmetric with respect to the median. The reproduction of the global histogram is ensured by the reproduction of the proportions at every threshold, therefore there is no need to assume any parametric shape of the conditional distributions. The simulated values are drawn from these conditional distributions.

[1] University of Alberta, Edmonton, Alberta, CANADA

One of the advantages of the indicator formalism is that they can incorporate higher order statistics than the two-point covariance function. This could allow reproduction of more complex features, such as curvilinear patterns and large range connectivity. Simulated annealing and iterative techniques have also proven to be efficient in this task (Deutsch and Gringarten, 2000). One of the main difficulties with multiple-point statistics is their inference. In current applications, these statistics are retrieved from training images taken from outcrops, conceptual geological models at different scales or realizations of some other algorithm such as object-based modeling (Strebelle and Journel, 2000).

We propose the use of "runs" (Figure 1), that is, sets of adjacent samples (composites) with grades higher than a given threshold (Mood, 1940). In mining, the frequency of runs can be easily obtained from drillhole data. They can be calculated for different grade cutoffs. In this context, indicator techniques appear as the obvious choice to incorporate these multiple-point statistics.

The inference of runs in the horizontal direction may require the assumption of some anisotropy ratio from geological interpretation of the mineralization. Incorporating runs will improve the reproduction of large range connectivity, but will not help in getting curvilinear patterns since it is by definition a linear measure of connectivity.

The use of runs is compatible with the indicator approach, but requires a framework different than the one provided by sequential methods. We present a hierarchical implementation of indicator simulation that will allow the subsequent incorporation of the runs to improve the resulting realizations, and therefore lead to better decisions based on the numerical models generated.

We start with a short review on indicator simulation and then propose our hierarchical framework for simulating thresholds from the highest to the lowest. This approach will permit the subsequent incorporation of runs as a multiple-point statistic inferred from the data. We conclude with a practical application of this new implementation of indicator simulation.

REVIEW ON INDICATOR SIMULATION

The basic idea behind indicator techniques is to code the data as probabilities and then estimate the conditional distributions at unsampled locations directly from them. Then a simulated value can be drawn directly from the conditional distribution. Previously simulated nodes are included in the conditioning information to obtain the correct covariance between simulated values as well as between simulated values and the original sample data.

Data from different sources, at different supports and of different precision can be simultaneously considered (Goovaerts, 1997; Deutsch and Journel, 1998). If K thresholds are used, the estimation of the cdfs can be done by:

- Kriging the indicator values at a given threshold (Solow, 1984). This disregards the information at different thresholds than the one being estimated. It requires modeling of K direct indicator variograms.
- Performing median indicator kriging. The mosaic model is assumed, that is, all indicator variograms and cross variograms are proportional to a common variogram model, or equivalently, all correlograms are equal (Journel, 1984).
- Cokriging the indicators at different threshold. Although this is theoretically better, practice has shown that it requires considerably more inference effort and brings little improvement (Goovaerts, 1994; Goovaerts, 1997). It requires modeling of $K \cdot (K+1)/2$ direct and cross indicator variograms.
- Cokriging the indicators at the threshold being estimated with the uniform transform of the data, that is, with their standardized rank ordering. This technique is known as probability kriging (Sullivan, 1984; Verly and Sullivan, 1985). This method requires the inference and modeling of $2 \cdot K + 1$ direct and cross variograms.

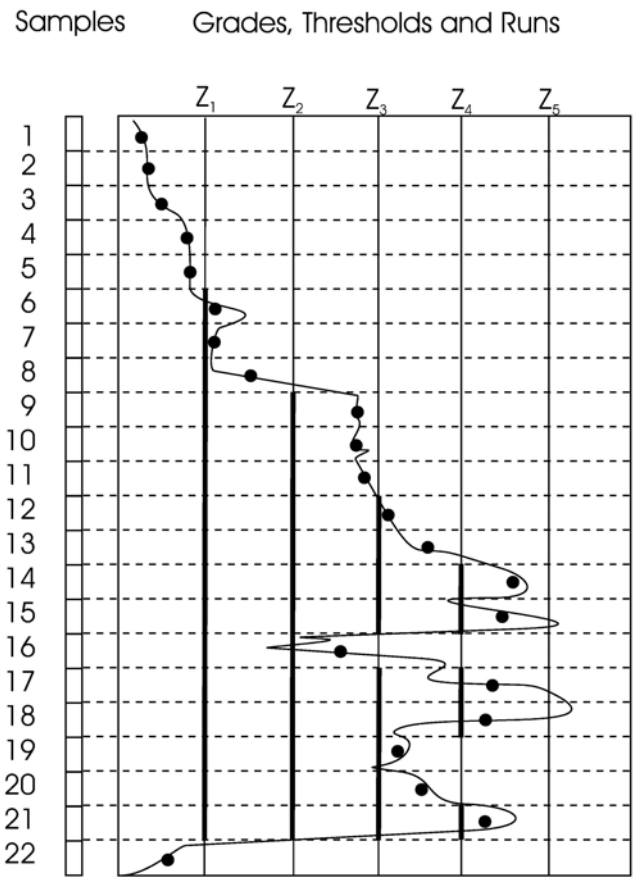

Figure 1: Schematic showing the calculation of frequency of runs in a drillhole with 22 samples. The solid line represents the actual grade, the black dots are the sample values. Runs are presented as thick solid lines under each threshold. For threshold z_1, there is one run of length 16; z_2 has one run of length 13; z_3 presents two runs of length 4 and 5 respectively; for z_4 there are two runs of length 2 and one of length 1; finally, z_5 does not present any run.

- Simple kriging with local prior means. A secondary variable is used to calculate locally the mean of the primary. Then simple kriging using this mean is performed using the indicator data of the primary variable (Deutsch and Journel, 1998).
- Soft Cokriging. This corresponds to cokriging the indicator transform of a primary variable, using also the indicator transform of a secondary variable (Goovaerts, 1997).
- Cokriging the indicators at the threshold of interest using only one secondary datum: the colocated one. This avoids the demanding modeling of cross indicator covariances (Zhu and Journel, 1993).

Several important advantages are derived from this basic idea of directly estimating the probabilities:

- The spatial correlation at different thresholds can be specified differently,

- Secondary information can be coded in the same probability units, which gives a flexibility to integrate such information,
- Change of support can be performed, and
- Recoverable reserves of blocks can be calculated.

The implementation of indicator methods is not always straightforward:

- The coding of soft data as if they were hard data is useful, but secondary information cannot be used as primary, even though the coding is the same. Some model of coregionalization has to be used.
- The use of data at different support is also a difficult task, since the correlation between the variables changes at different supports.

Drawbacks of Indicator Methods

Two are the main problems of indicator techniques: order relation deviations and the interpolation between and extrapolation beyond the discrete estimated probabilities at thresholds $z_k, k = 1,...,K$:

- **Order relation deviations.** The estimated probabilities generated through indicator kriging must satisfy the conditions of a cumulative distribution: they have to be non-decreasing between 0 and 1 (Journel, 1984; Goovaerts, 1997; Deutsch and Journel, 1998). The kriged indicator value can lie outside the interval [0,1] because the kriged estimate is non-convex. Lack of data in some classes and differences in the variogram models from one threshold to the next are important factors that could lead to a non-increasing function. The *a posteriori* forward and downward correction of the ccdfs works well in general, as documented by Deutsch and Journel. Although more difficult in its implementation, constraining the kriging system, so that it satisfies the order relations by construction is also a solution (Goovaerts 1997).
- **Interpolation and extrapolation of the conditional distribution.** Since the number of data is limited, the distribution of local uncertainty is discretized using only a few thresholds. The continuous ccdf at every location $\mathbf{u}$ is then represented by a set of points $[i(\mathbf{u}; z_k)]^*$ with $k = 1,...,K$, that lie in [0,1]. It is therefore necessary to interpolate the values between thresholds, and extrapolate the values beyond the smallest and largest values. It may be sufficient to interpolate linearly between the indicator values at thresholds z_{k-1} and z_k. When extrapolating the tails, a minimum and maximum possible values should be considered and the extrapolation should not be done linearly, since this would imply an unrealistic uniform distribution between the minimum value and z_1, and between z_K and the maximum value. Power and hyperbolic models are sometimes used to extrapolate the ccdfs beyond the lower and higher indicator values. Another possibility is to consider the global cdf and scale it to extrapolate the tails of the ccdfs. There is no satisfactory solution.

Indicator Simulation

Indicator simulation uses the ccdf obtained through indicator kriging to draw a simulated value by Monte Carlo simulation. It is important to emphasize that the conditioning data used to get the ccdf consists of actual data and previously simulated values within the search neighborhood. In this way, the covariance is reproduced.

The sequential simulation approach for a grid of N odes proceeds as follows:

- Step 1: Randomly pick a node.
- Step 2: Search for nearby data and previously simulated nodes.

- Step 3: Perform indicator kriging at each threshold to build the ccdf.
- Step 4: Draw by Monte Carlo simulation, a value from that conditional distribution.
- Step 5: Go to Step 1.

The conditioning information increases from n data to $n+N-1$ data and previously simulated nodes. The bigger the number of conditioning data, the larger will be the kriging system. This problem is overcome by using a search neighborhood and maximum number of data within that radius.

Variogram reproduction depends on different factors such as the size of the search neighborhood and the importance of high order correlation compared with two-point statistics.

Generalization of Indicator Kriging

A brief presentation of generalized indicator approach follows (Guardiano and Srivastava,1993).

Consider N dependent events $A_j, j=1,...,N$. They can be sequentially estimated or simulated using the following expression that comes from a repeated application of Bayes postulate:

$$\begin{aligned}\text{Prob}\{A_j, j=1,...N\} = \text{Prob}\{A_N \mid A_j, j=1,...,N-1\}\cdot \text{Prob}\{A_{N-1} \mid A_j,,=1,...,N-2\}\cdot ...\cdot \\ \text{Prob}\{A_2 \mid A_1\}\cdot \text{Prob}\{A_1\}\end{aligned}$$

This relation is general and exact (Journel, 1983). Two implementation problems arise:

- Inference of the N conditional probabilities $\text{Prob}\{A_i \mid A_j, j=1,...,i-1\}$ with $i=1,...,N$.
- The size of the conditioning information increases from n to $n+N-1$, i.e. the kriging system to be solved becomes too big.

Due to the difficult inference of those conditional probabilities, some approximations are made to implement sequential indicator simulation. The conditional probability $F(u;z_k \mid (n))$ is interpreted as the conditional expectation of an indicator random variable $I(u;z_k)$ given the n data. The conditional probability should then be obtained by cokriging using the $n\cdot K$ indicators. It is assumed that the indicators for the same threshold are more relevant than the indicators for other thresholds. The calculation of cross-correlations is also considered too demanding. Thus the probability is approximated by:

$$\text{Prob}\{Z(u)\le z_k \mid Z(u_\alpha)=z(u_\alpha), \alpha\in(n)\} = E\{I(u;z_k) \mid I(u_\alpha;z_k)=i(u_\alpha;z_k), \alpha\in(n)\}$$

All cross-correlation between indicators at different thresholds are then ignored. This first approximation is made because, in general, the improvement in the resulting conditional distributions does not justify increasing the work required (Goovaerts, 1994).

The conditional expectation can be written as a function of the conditioning information in the following manner:

$$\begin{aligned}E\{I(u;z_k) \mid I(u_\alpha;z_k)=i(u_\alpha;z_k), \alpha\in(n)\} &= \varphi\{i(u_\alpha;z_k), \alpha\in(n)\} \\ &= a_0 + \sum_{\alpha\in(n)} a_1(\alpha)\cdot i(u_\alpha;z_k) + \sum_{\alpha\in(n)}\sum_{\substack{\alpha'\in(n)\\ \alpha\neq\alpha'}} a_2(\alpha,\alpha')\cdot i(u_\alpha;z_k)\cdot i(u_{\alpha'};z_k) \\ &\quad + ... + a_n\cdot\prod_{\alpha\in(n)} i(u_\alpha;z_k)\end{aligned}$$

The classical application of indicators considers only the use of univariate and bivariate statistics (the sample histogram and the covariance or variogram function), since the inference of higher

order is in general difficult. This is done by retaining only the first $n+1$ terms of the previous expansion; however the use of higher order statistics could be considered, through the use of extended normal equations (Guardiano and Srivastava, 1992). This approach was originally based on multiple-point statistics inferred from training images. The use of runs to estimate high order covariances would avoid the need of training images, by inferring the multiple-point covariances directly.

HIERARCHICAL INDICATOR SIMULATION

The idea of proceeding hierarchically was originally proposed by Michel Dagbert (Dagbert, 1990), under the name "Nested Indicator Approach". The thresholds are considered one at a time, and only the subset of data, which has a value above the previous cutoff, is used to *estimate* the conditional probability at the current threshold. The domain from which data are used to estimate the conditional distribution at subsequent threshold decreases and order relations are avoided by construction. These conditional distributions were used for estimating block grades and calculating recoverable reserves, but they were not used to generate realizations of the random function. We propose a simulation method that combines Dagbert's original idea with the well-known sequential framework, to simulate one threshold at a time, starting at the highest.

Procedure

The algorithm starts by doing a binary sequential indicator simulation at the highest threshold. If the value obtained is one, i.e. the node has a grade less than or equal to the cutoff, then a simulated value is not assigned to that node yet, and the node will belong to the domain for the next cutoff. However, if the value is found to be above the current threshold, i.e. it has a value of zero in the binary simulation, then an actual simulated value is drawn between the current (highest) cutoff and a maximum value defined by the user. For the next cutoff, a reduced domain is available. Again, within that domain a binary sequential indicator simulation is performed, and the values above the threshold are taken out the domain for the next cutoff.

The algorithm proceeds starting at the highest threshold and going down:

1. The domain corresponds to all uninformed nodes.
2. Pick a node in the domain randomly.
3. Calculate the simple indicator kriging estimate at the current threshold given the nearby data and previously simulated nodes.
4. Draw a random number and assign a one if it is lower or equal than the simple indicator kriging estimate of the probability at that threshold, and a zero otherwise.
5. Go to 2 until all nodes in the domain have been visited.
6. If the value is above the threshold (a value of zero was assigned in the binary simulation) eliminate the node of the domain for the next threshold.
7. If the value is below the threshold include it in the domain for the next threshold.

At the end of the simulation, each node has been assigned to one of the $K+1$ classes, if K thresholds were used. Then at each node a simulated value is drawn into the corresponding class at which it belongs. Some shape has to be assumed for interpolating between thresholds and extrapolating to a maximum and minimum value.

Since we are basically applying sequential indicator simulation in restricted domains, the algorithm ensures the reproduction of the covariance function within that domain. The fact that the nodes higher than the previous threshold are not considered as conditioning data generates some distortion on the covariance reproduction, since the nodes in the domain are drawn uncorrelated with the nodes kept out of the domain.

Example

The algorithm is illustrated on Figure 2 for a simple example. Three thresholds have been considered and a one-dimensional grid of ten nodes is being simulated. Starting at the highest

threshold z_3, all ten nodes are simulated to be above or below that threshold. Nodes that have been assigned a zero are higher than z_3 and therefore simulated values are drawn between z_3 and a maximum value z_{max}, and with some extrapolation shape. All the nodes that have a grade higher than z_3 are not available at the following threshold z_2. The algorithm continues in this fashion up to the last threshold z_1. The nodes assigned a zero in the binary simulation are now drawn between z_1 and z_2 with some interpolation shape, and the values that are below z_1 are drawn as well between a minimum value z_{min} and z_1. This concludes the simulation.

Notice that sequential indicator simulation is used to realize the binary simulation within the domain. Therefore the values assigned as zero by this binary simulation are used as conditioning data for the following nodes at that same threshold. However, when all the nodes at the current threshold have been visited and the grid has been populated with ones and zeros, then the nodes with zeros are excluded from the domain for the next threshold and therefore they are not considered as conditioning values for calculating the conditional distributions at the lower thresholds.

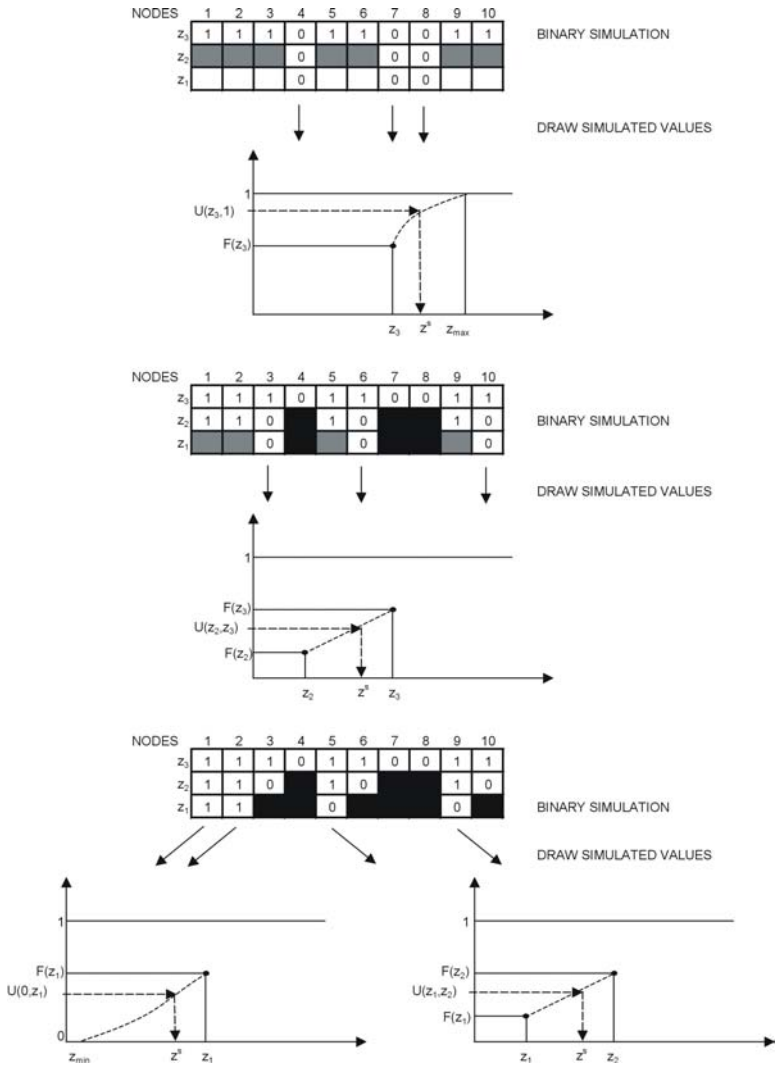

Figure 2: Illustration of the hierarchical indicator simulation method. The allowed domain is shown in gray for the following threshold. Nodes that have already been simulated above the previous threshold are frozen. They are represented as black boxes.

CASE STUDY

Drillhole data from a copper deposit are used to illustrate the methodology. The dataset has 869 ten-meter composites with copper grade, distributed on a 600 by 400 m^2 area and over 270 m in depth. The declustered histogram and location map of drillholes is presented in Figure 3.

Five thresholds are used to discretize the conditional distributions. These are placed at regular quantiles. The mineralization is isotropic in the horizontal plane, but the continuity changes in the vertical direction. Three dimensional variogram models are fitted to the experimental variograms, as shown in Figure 4. The variogram models are summarized in Table 1.

Quantile	Nugget Effect	Model Type	Contribution	Range Horizontal	Range Vertical
10	0.20	Exponential	0.80	650	100
30	0.05	Exponential	0.95	380	50
50	0.10	Exponential	0.90	260	45
70	0.20	Exponential	0.80	130	30
90	0.30	Exponential	0.70	65	15

Table 1: Variogram models for five thresholds

Ten realizations are generated using Hierarchical Indicator Simulation. The model consists of as 120 by 80 by 27 cells with spacing of 5 m in the horizontal plane and 10 m in the vertical. This model is dense enough to calculate for example block averages for units of 20 by 20 by 10 m^3.
Maps showing some of the levels for one realization are presented in Figure 5. Variogram reproduction is shown in Figure 6. The variograms are well reproduced for high thresholds, but show a higher nugget effect and more spread for lower thresholds. This is due to the restricted domain utilized for simulating lower thresholds. The nodes that are excluded from this domain are not directly correlated with the nodes being simulated at that threshold, so there are more fluctuations in the variogram. Also, the nugget effect is increased for the same reason. Although this is a drawback of the method, we must emphasize that low values are often the less interesting, and therefore focus is put on reproducing the features of high values. Finally, we must mention that this method is being proposed as a framework for incorporating multiple-point statistics.

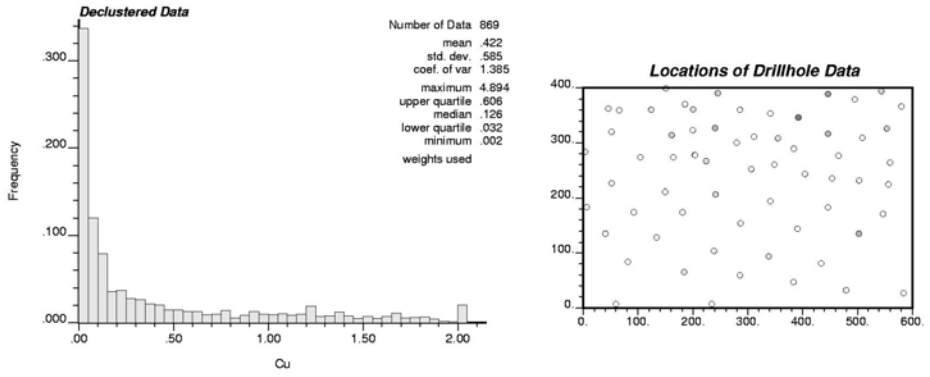

Figure 3: Declustered histogram and location map of drillholes – Copper deposit dataset. Drillholes are in a pseudo-regular grid with a spacing of approximately 70 m.

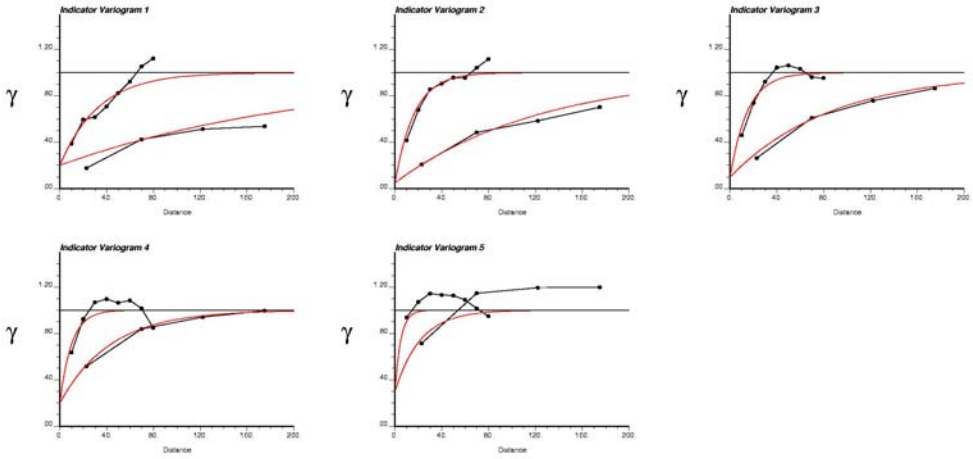

Figure 4: Indicator variogram models. Three dimensional anisotropic variogram models were fitted to the experimental variograms. As usual, high grades present a higher nugget effect and are less continuous than low grades.

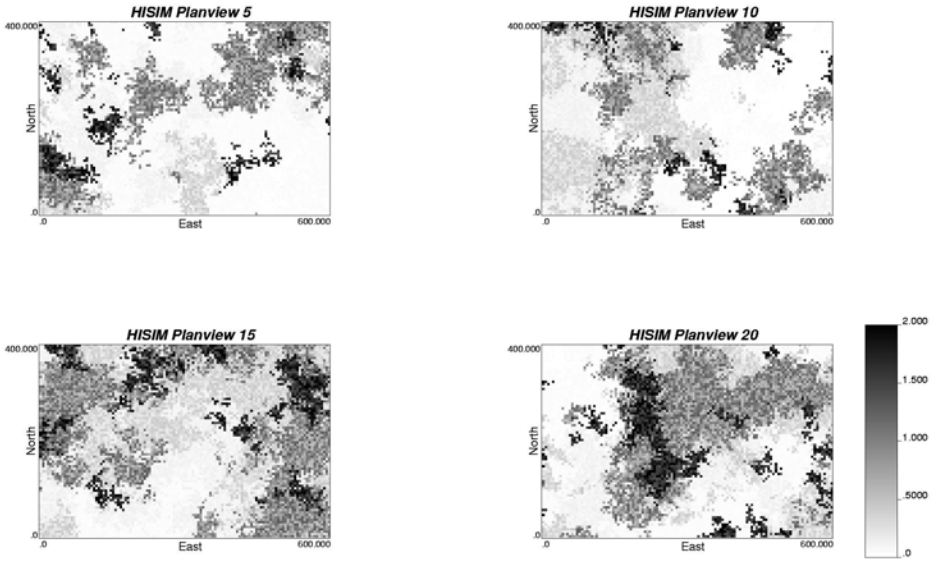

Figure 5: Four planviews of a particular realization. The continuity appears clearly different for high and low values.

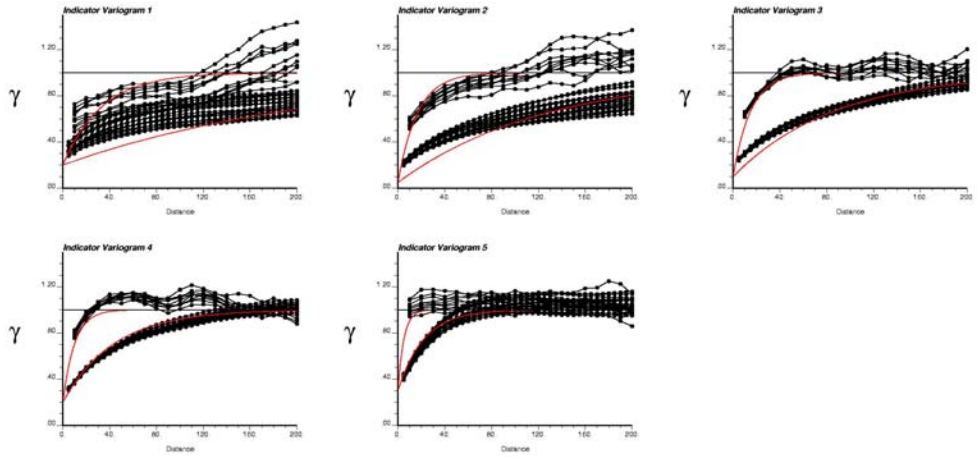

Figure 6: Indicator variogram reproduction. High thresholds show a very good reproduction of the variogram. However, for lower thresholds reproduction of the nugget effect is more difficult due to the restricted domain and the fact that nodes are simulated independently of the ones higher than the previous threshold.

COMMENT ON INCORPORING MULTIPLE-POINT STATISTICS

Considering runs when simulating if the node is above or below the current threshold will correct the smearing of the variogram. Runs can be seen as products of indicators: the data that form a run have grades larger than the cutoff, and therefore all their indicators are 0. When using runs to decide if we set a node below the threshold, we are basically deciding if we want to cut the run into two shorter ones or reduce its size by one. This fact may imply that variogram modeling will not be needed and the entire simulation could be based solely on the reproduction of the histogram of length of runs at that particular threshold. This would simplify geostatistical modeling considerably. The only decisions that remain are how many and which thresholds to use. Wrong decisions will lead to poor inference and unsatisfactory numerical models.

CONCLUSIONS

Hierarchical Indicator Simulation is a new implementation of indicator simulation based on Dagbert's original idea of estimating nested indicators. It fixes the problem of order relation deviations, which is one of the biggest drawbacks of sequential indicator simulation. The reduced domain may generate some problems with variogram reproduction. This algorithm is very suitable to incorporate multiple-point statistics in form of runs, that is, of sets of adjacent samples with grades higher than a given threshold. By incorporating these runs, the distortion on variogram reproduction will be overcome and furthermore, the numerical model will share more important features of the true underlying phenomenon: large range connectivity will be reproduced.

ACKNOWLEDGEMENTS

We are grateful tothe industry sponsors of the Centre for Computational Geostatistics at the University of Alberta and the Mining Department at the University of Chile for supporting this research.

REFERENCES

Dagbert, M., 1990. *Nested Indicator Approach for Ore Reserve Estimation in Highly Variable Mineralization*. In 92nd Annual General Meeting of CIM.

Deutsch, C. V. and E. Gringarten, 2000. *Accounting for Multiple-Point Continuity in Geostatistics Modeling.* In 6th International Geostatistics Congress, Cape Town, South Africa.

Deutsch, C. V. and A. G. Journel. 1998. *GSLIB: Geostatistical Software Library: and User's Guide, 2nd Ed*. New York: Oxford University Press.

Gómez-Hernández, J. J. and R. M. Srivastava, 1990. *ISIM3D: An ANSI-C Three Dimensional Multiple Indicator Conditional Simulation Program.* Computers & Geosciences, vol 16(4), pp 395-410.

Goovaerts, P., 1994. *Comparative Performance of Indicator Algorithms for Modeling Conditional Probability Distribution Functions.* Mathematical Geology, vol 26(3), pp 385-410.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation.* New York, Oxford University Press.

Guardiano, F. and M. Srivastava, 1993. *Multivariate Geostatistics: Beyond Bivariate Moments.* Geostatistics Troia '92, vol 1, ed. A. Soares, Kluwer Academic Publishers, pp 133-144.

Journel, A. G., 1982. *The Indicator Approach to Estimation of Spatial Distributions.* In Proceedings of the 17th International APCOM Symposium, Society of Mining Engineers, pp 793-806.

Journel, A. G., 1983. *Non-Parametric Estimation of Spatial Distribution*. Mathematical Geology, vol 15, pp 445-468.

Journel, A. G., 1984. *The Place of Non-Parametric Geostatistics.* Geostatistics for Natural Resources Characterization, Reidel, Dordrecht, Holland, vol 1, pp 307-355.

Journel, A. G. and F. Alabert, 1989. *Non-Gaussian Data Expansion in the Earth Sciences.* Terra Nova, vol 1, pp 123-134.

Journel, A. G. and D. Posa, 1990. *Characteristic Behavior and Order Relations for Indicator Variograms.* Mathematical Geology, vol 22(8), pp 1011-1025.

Mood, A. M., 1940. *The Distribution Theory of Runs.* Annals of Mathematical Statistics, vol 11, pp 367-392.

Solow, A. R., 1986. *Mapping by Simple Indicator Kriging.* Mathematical Geology, vol 18(3), pp 335-352.

Strebelle, S. and A. G. Journel, 2000. *Sequential Simulation Drawing Structures from Training Images.* In 6th International Geostatistics Congress, Cape Town, South Africa.

Sullivan, J., 1984. *Conditional Recovery Estimation Through Probability Kriging: Theory and Practice.* Geostatistics for Natural Resources Characterization, Reidel, Dordrecht, Holland, vol 1, pp 365-384.

Verly, G. and J. Sullivan, 1985. *MultiGaussian and Probability Krigings – Application to the Jerritt Canyon Deposit.* Mining Engineering, pp 568-574.

Zhu, H. and A. G. Journel, 1993. *Formatting and Integrating Soft Data: Stochastic Imaging via the Markov-Bayes Algorithm.* Geostatistics Troia '92, vol 1, ed. A. Soares, Kluwer Academic Publishers, pp 1-12.