A Short Note on Within Class Modeling in Indicator Simulation

Julián M. Ortiz, Chad Neufeld, and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil & Environmental Engineering
University of Alberta

Sequential indicator simulation (SIS) is a powerful alternative to Gaussian based methods for simulating continuous random variables; however, there are two common problems: (1) transition probabilities and the ordering of the classes are not reproduced, causing an unrealistic patchy look to the realizations, and (2) the drawing within classes is random, hence problems of a too high nugget effect and unrealistic abrupt changes when a high value within a class is drawn close to a low value of a different lower class.

The first problem can be partially solved by using a full cokriging of the indicators to estimate the local conditional probabilities. Papers 401 and 402 in this report address this problem to some extent. This short note focuses on the second problem. We propose a solution to improve the resolution of SIS within classes, which requires considering the distance to the edge between classes. A distance measure could be used as a secondary variable to inject this continuity across the boundaries. This secondary variable could even add some fuzziness to the class boundaries, given that kriging may extrapolate beyond the minimum and maximum values defined by the data, in this case, the distance to the edge of the class.

Introduction

Sequential indicator simulation (SIS) is commonly used to generate stochastic rock type models. The approach is flexible in its ability to capture hard and soft data and locally varying means. SIS is also used for continuous variables where it provides an alternative to the widely used multivariate Gaussian methods such as sequential Gaussian simulation (SGS), simulation based on matrix decomposition, and turning bands (Journel, 1983; Journel and Isaaks, 1984, Gómez Hernández and Srivastava, 1990).

The main advantages of SIS are the integration of secondary variables and the ability to consider different correlation functions at different thresholds (Alabert, 1987). Notwithstanding these advantages, practitioners avoid its use due to the unrealistic look of the realizations obtained by this method. Realizations look patchy because the transition probabilities between thresholds are not considered in the estimation of the indicators.

The estimation of the conditional distributions is usually done by approximating the conditional expectation of the indicator by its simple kriging estimate. Consider the indicator coding of the continuous values of Z:

$$i(\mathbf{u_\alpha}; z_k) = \text{Prob}\{z(\mathbf{u_\alpha}) \leq z_k\} \qquad \forall k = 1,...,K$$

Where z_k, $k=1,...,K$ are predefined thresholds (normally K is between 7 to 10) and $\mathbf{u}_\alpha$ is a location. The conditional expectation of the indicator value is estimated by:

$$E\{I(\mathbf{u}_\alpha;z_k)\,|\,(n)\} \approx [i(\mathbf{u}_\alpha;z_k)]^*_{SK} = \sum_{\alpha=1}^{n} \lambda_\alpha^{SK}(\mathbf{u};z_k)\cdot i(\mathbf{u};z_k) + \left[1-\sum_{\alpha=1}^{n}\lambda_\alpha^{SK}(\mathbf{u};z_k)\right]\cdot F(z_k)$$

where the weights are calculated by solving the well known simple kriging system:

$$\sum_{\beta=1}^{n} \lambda_\beta^{SK}(\mathbf{u};z_k)\cdot C_I(\mathbf{u}_\beta-\mathbf{u}_\alpha;z_k) = C_I(\mathbf{u}-\mathbf{u}_\alpha;z_k) \qquad \forall\alpha=1,\ldots,n$$

SIS consists of implementing this inference step in a sequential fashion with Monte Carlo Simulation (MCS) implemented at each step from the conditional distributions. This approach provides realizations with an unrealistic look (**Figure 1**). This could be partially solved by considering a full indicator cokriging to estimate the indicator values at unsampled locations or by accounting for the transition probabilities between the classes defined by the thresholds (Goovaerts, 1994; Carle and Fogg, 1996).

A second problem is the lack of correlation within classes due to the random drawing, once the class has been defined. This can be seen clearly, by looking at the simulated values within a particular class defined by a lower and upper threshold. As shown in the example, the variogram of points within that class is a pure nugget effect, since values have been drawn considering independent uniform random numbers between the cumulative distribution function values for the upper and lower thresholds (**Figure 2**).

This lack of correlation carries an additional problem. We can have the highest value of a class close to a boundary between that class and a lower class, having the lowest value of that lower class close to the boundary too (**Figure 3**). This causes a problem of consistency in the simulated model, which we would like to avoid.

Proposed Methodology

Our idea to address the problem of within class resolution is to consider a distance from each node to the closest edge of the class. This distance and the closest class type could be used to introduce correlation in the probabilities when drawing values within the class.

To illustrate the idea, let us consider a realization where only two classes for a continuous variable with uniform distribution in [0,1] have been defined. The continuous variable can be coded as a categorical variable, and the distance to the edge of the class can be calculated (**Figure 4**). The distance to the nearest edge can be used as a secondary variable to draw from the within class distribution in order to get a better resolution of SIS within the classes and a more consistent numerical model of the continuous variable of interest. The main difficulty is to find what the correlation of the distance measure is to draw a value.

One approach to determine this correlation would be to infer the variogram of the distance measure using as a reference a Gaussian simulation. We could have a hybrid method that uses the classes defined by SIS (with all its advantages: different variograms for different thresholds, integration of secondary information) and the distance correlation from a truncated Gaussian simulation.

Another possibility to account for this distance to the edge would be to calculate the cross-correlation between the value within a particular class and its distance value. Simulated values within that class could be drawn by considering this coregionalization. This approach may even allow changing the position of the edges, since kriging is a non-convex estimator that allows generating values outside the range of the data. The borders between classes may become fuzzy and change from what was originally defined by the simulation.

Conclusions and future work

One significant problem of sequential indicator simulation is that values simulated within each class defined by the thresholds occur randomly, generating inconsistencies due to the possibility of simulating a high value very close to a low value, without capturing the proper transition between high and low values that a continuous variable should show.

We propose investigating the use of a secondary variable related to the distance of the node to the edge with the other classes defined by the indicator coding of the data, in order to impose some degree of correlation and consistency within the classes. This methodology could potentially improve the performance of numerical models generated with sequential indicator simulation.

Future work is required to determine the best approach to impose the within class correlation. Considering the correlation found in a Gaussian random field is one possibility that will be explored.

References

- Alabert, F.G., 1987, *Stochastic Imaging of Spatial Distributions Using Hard and Soft Information*, Master's Thesis, Stanford University, Stanford, CA.
- Carle, S. F., and Fogg, G. E., 1996, *Transition Probability-Based Indicator Geostatistics*, Mathematical Geology, Vol. 28, No. 4, pp 453-476.
- Goovaerts, P., 1994, *Comparative Performance of Indicator Algorithms for Modeling Conditional Probability Distribution Functions*, Mathematical Geology, Vol. 26, No. 3, pp 385-410.
- Gómez-Hernández, J.J. and Srivastava, R.M., 1990, *ISIM3D: An ANSI-C Three Dimensional Multiple Indicator Conditional Simulation Program*, Computers & Geosciences, Vol. 16, No. 4, pp. 395-410.
- Journel, A.G., 1983. *Nonparametric estimation of spatial distribution*. Mathematical Geology, Vol. 15, No. 3, pp. 445-468.
- Journel, A.G. and Isaaks, E.H., 1984. *Conditional indicator simulation: Application to a Saskatchewan uranium deposit.* Mathematical Geology, Vol. 16, No. 7, pp. 685-718.

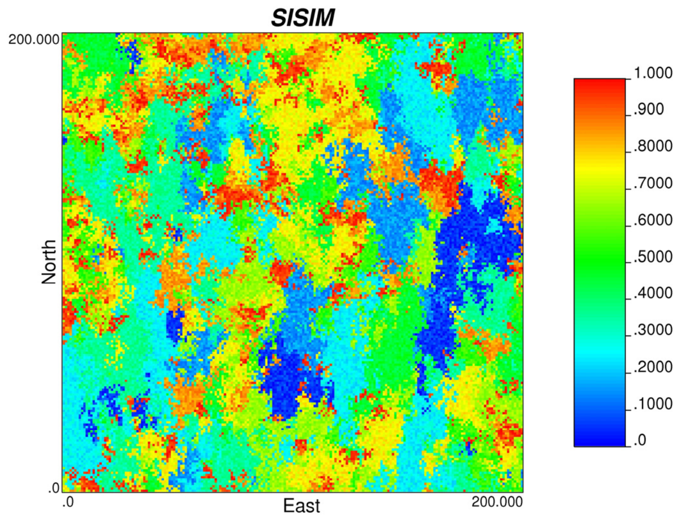

Figure 1: Typical sequential indicator simulation realization. The continuous simulated values are shown and the discontinuity between the grades across the boundary defined by two classes appears quite clearly.

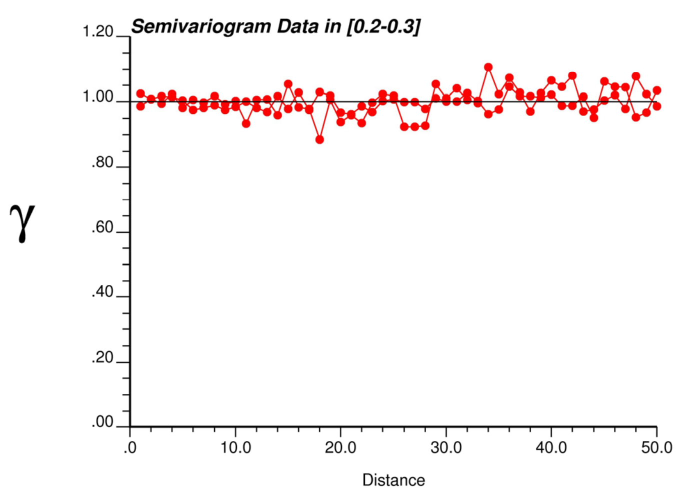

Figure 2: Variogram in the North-South and East-West directions considering only the values within the class defined by the thresholds 0.2 and 0.3.

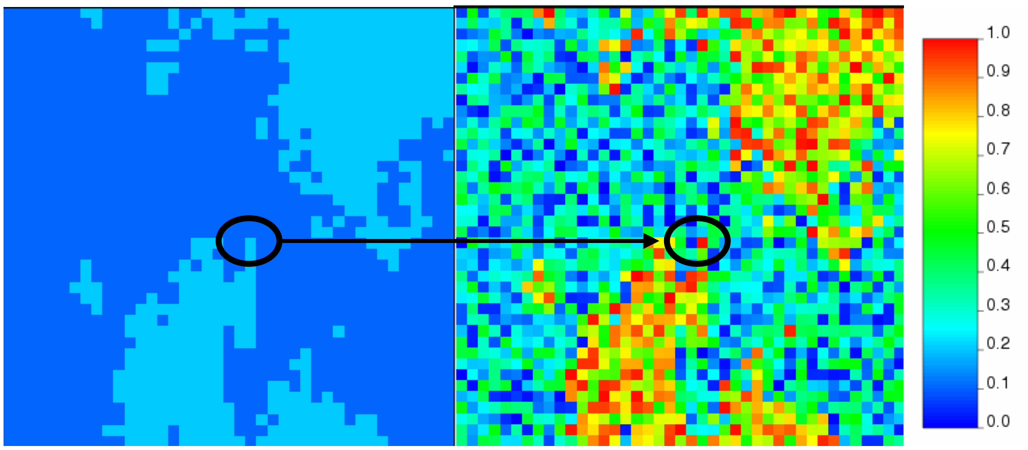

Figure 3: The problem of within class resolution. On the left, the simulated contact between two classes of continuous values. On the right, the actual simulated values can be seen. It is clear that, since SIS does not control the correlation within the class (and near the contact with the adjacent class), a high value of the higher valued class can be drawn close to a low value of the lower valued class.

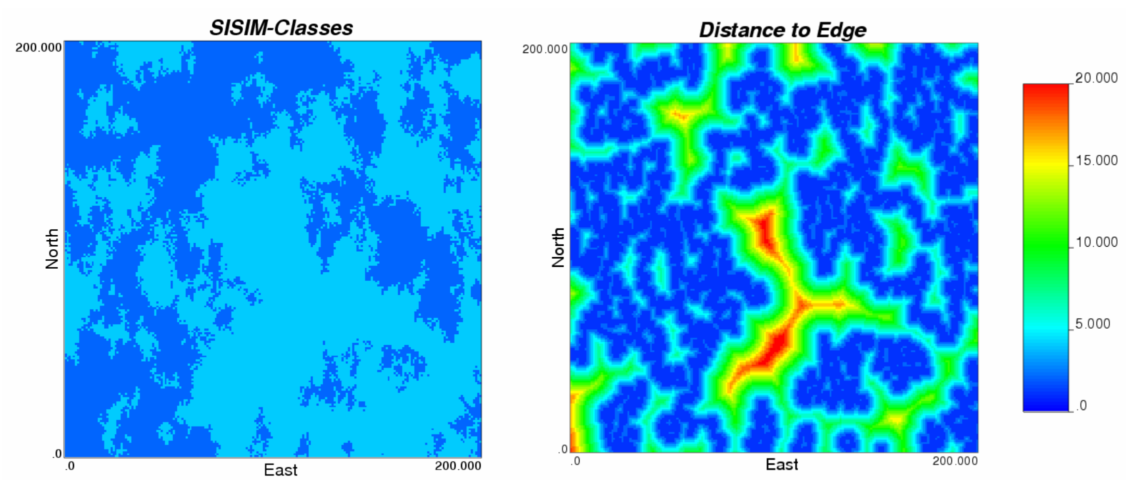

Figure 4: Left: Map showing the location of values that fall above (light blue) and below (blue) the median of the distribution. Right: Map showing the distance of each simulated node to the closest edge of its corresponding class.