Short Note on Cokriging in Sequential Indicator Simulation: The Adjacent cut-off Alternative.

Chad Neufeld, David F. Machuca Mory and Clayton V. Deutsch

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

A drawback of Sequential Indicator Simulation is the uncontrolled transitions between classes, which translates in the patchiness of high and low values areas in the resulting realizations. The full cokriging approach has been proposed to solve this disadvantage; all direct and cross indicator variograms would be used. This approach should introduce some order in the interclass transitions by including the interclass cross correlation information; however, the Linear Model of Corregionalization does not provide a satisfactory fitting for indicator cross variograms of extremely separated thresholds. The alternative proposed in this paper is to use only the corregionalization information of the two closest thresholds to the one that is been used for the conditional CDF estimation. This alternative has been implemented in the `SISIM_adj` *program. The implementation details, the results using synthetic and real data and the performance comparison of this alternative with the direct indicator simulation and full indicator simulation are shown in this paper.*

Introduction

Indicator based Kriging and simulation methods suffer several shortcomings and limitations, these have been documented by several authors (Chilès and Delfiner, 1999; Christakos, 2000; Emery and Ortiz, 2004). A particular unwarranted feature in the results of traditional Sequential Indicator Simulation (SIS) for continuous variables is the geologically unrealistic, disordered and uncontrolled transitions between classes of low and high values. This is caused by the non-parametric description of the cumulative distribution function (cdf) using the indicator cumulative probability values for several classes and the limitations of indicator direct variograms to provide information about the inter-class cross correlation.

The implementation of full co-kriging algorithm in Sequential Indicator Simulation has been suggested to account for the interclass correlation and, in this way, to control the interclass transitions and sequence. Nevertheless, the short distance continuity of indicator cross variograms becomes very pronounced as the difference between thresholds increases. In this case the linear model of corregionalization provides a very poor fit to experimental indicator cross variogram for widely separated cut-offs.

In view of this limitation, it is proposed the implementation of a cokriging system considering only the indicator cross variograms between the closest upper and lower threshold to the one which cdf value is being estimated.

Theoretical Framework

The indicator transform for a continuous variable is defined in relation to a threshold value, or cut-off, y_p as:

$$I(u;y_p)=\begin{cases}1 & \text{if } Y(u)\le y_p\\ 0 & \text{if } Y(u)>y_p\end{cases}$$

Where p is the cdf value corresponding to the threshold y_p. When the stationarity hypothesis is assumed, the expected value of the indicator transform defines the marginal cdf of the variable *Y(u)*, regardless the location *u*:

$$E\left[I(u;y_p)\right]=\text{Prob}\left\{Y(u)\le y_p\right\}=F(y_p)$$

Using ordinary or simple Kriging, and provided the indicator direct variograms or covariances, this global cdf can be conditioned by the *n* surrounding data values at any location *u* and for several y_p cut-off's. The simple indicator Kriging SIK system of equations is expressed as:

$$\sum_{\beta=1}^{n}\lambda_\beta(u_\alpha;y_p)\cdot C_I(u_\alpha-u_\beta;y_p)=C_I(u_\alpha-u;y_p)\quad\forall\alpha=1,\ldots,\text{n}$$

The resultant SIK weights, $\lambda_\beta(u;p)$, are then used to estimate the conditional local probabilities by the expression (Deustch and Journel, 1998):

$$\begin{aligned}F^*_{SIK}(u_0;y_p\,|\,(n))&=[\text{Prob}\{Y(u)\le y_p\,|\,(n)\}]^*_{SIK}\\&=\sum_{\alpha=1}^{n}\lambda_\alpha(u_0;y_p)I(u_\alpha;y_p)+\left[1-\sum_{\alpha=1}^{n}\lambda_\alpha(u;y_p)\right]F(y_p)\end{aligned}$$

The construction of such conditional cumulative distribution functions by the indicator kriging approach often leads to order relation deviations. This means that the discrete probabilities independently estimated for each threshold can lead to a ccdf models that can be decreasing for certain intervals, or can be lower than zero or bigger than 1. These order relation deviations can be corrected by simple methods (Deustch and Journel, 1998).

Traditional Sequential Indicator Simulation commonly uses simple indicator Kriging to build a local ccdf model which is used to draw a random value by Monte Carlo Simulation. This value is added to the data set and used in the construction of the ccdf's of the remaining nodes.

As the discrete probabilities are estimated independently for each cutoff and using indicator direct variograms, the information about the interclass correlation is not included, thus in SIS results maps classes appear as superimposed patches of low and high values with neither gradation nor order in their sequence.

The implementation of the full indicator cokriging (coIK) is supposed to palliate this problem, since it includes all the information available from the indicator direct and cross variogram for every pair of threshold combinations. The coIK ccdf estimate for a given cut-off y_{p_0} is expressed as (Deustch and Journel, 1998; Goovaerts, 1994):

$$F^*_{coIK}(u_0;y_{p_0}\,|(n))-F(y_{p_0})=\sum_{p=1}^{P}\sum_{\alpha=1}^{n}\lambda_{\alpha,p}(u;y_{p_0})\cdot\left[I(u_\alpha;y_p)-F(y_p)\right]$$

And the system of equations is written:

$$\sum_{p'=1}^{P}\sum_{\beta=1}^{n}\mu_{\beta,p'}(u_\beta;y_{p_0})\cdot C_I(u_\alpha-u_\beta;y_{p,}y_{p'})=C_I(u_\alpha-u_0;y_{p,}y_{p_0})$$
$$\forall\alpha=1\text{ to }n,\quad p=1\text{ to }P$$

coIK requires the simultaneous modeling of a $P\times P$ matrix of indicator direct and cross variograms:

$$\Gamma_I(h)=\begin{bmatrix}\gamma_I(h;y_1,y_1) & \cdots & \gamma_I(h;y_1,y_k)\\ \vdots & \ddots & \vdots\\ \gamma_I(h;y_k,y_1) & \cdots & \gamma_I(h;y_k,y_k)\end{bmatrix}$$

This matrix must be positive semi-definite to ensure that the variance of any linear combination of the transform $I(u;y_p)$ is non negative (Journel and Huijbregts, 1978). The Linear Model of Corregionalization (LMC) allows the fulfilling of this condition by modelling all the direct and cross variogram as a linear combination of a limited number of variogram functions (Goovaerts; 1994, 1998):

$$\Gamma_I(h)=\sum_{l=1}^{L}B^l\cdot g_I(h)$$

Where $g_I(h)$ is a permissible variogram model with standardized sill, and B^l is a $P\times P$ matrix containing the sill contributions, $b^l_{pp'}$, of the model $g_I(h)$ for each indicator direct and cross variogram.

To assure the positive definiteness of the LMC, the matrix B^l must be positive semi definite and the same permissible variogram models, $g_I(h)$, must be fitted to direct and cross variograms.

However, as mentioned above, the LMC fails to reproduce the extreme continuity of indicator cross variograms for widely separated cut-off's. This excessive continuity can not be fitted by any permissible model, Moreover the changing shape of variograms, from exponential like direct variograms to Gaussian type cross variograms, difficult or hinder the fitting of a valid positive semi-definite correlation matrix under the LMC approach.

This extraordinary variogram continuity can be explained in regard to the indicator cross variogram, which can also be expressed as:

$$2\gamma_I(h;y_p,y_{p'})=E\left\{\left[I(u;y_p)-I(u+h;y_p)\right]\left[I(u,y_{p'})-I(u+h;y_{p'})\right]\right\}$$

Where and *u* and *u+h* are data locations separated by a vector *h*. The indicator cross variogram increases only if there is a class transition in the same direction for both cut-off's, y_p and y_k. This is, when the next conditions are fulfilled:

$$y(u) \le y_p,\ \ y(u+h) > y_p,\ y(u) \le y_{p'} \text{ and } y(u+h) > y_{p'} \text{ or}$$
$$y(u) > y_p,\ \ y(u+h) \le y_p,\ y(u) > y_{p'} \text{ and } y(u+h) \le y_{p'}$$

But as the difference between the thresholds y_p and y_k gets bigger, these conditions are seldom satisfied, particularly when the nugget effect is very low and the variable is very continuous or approach to the bigaussian model. Then, the experimental variogram yields a long tail of zero or close to zero values for the close range. This extreme continuity is often more pronounced than the continuity of the Gaussian variogram model, and the LMC provides a very coarse description of it when the complete matrix of direct and cross variograms is considered.

As the LMC is inadequate for fitting the full variogram matrix, and being it the only model of corregionalization available, the adjacent Cut-offs Indicator Cokriging (acoIK) is a viable and practical alternative for introducing the interclass correlation information in the SIS algorithm.

For each one of the *p* cut-offs, the correspondent ccdf value estimation by acoIK requires only the LMC of the indicator direct and cross variograms corresponding to the adjacent thresholds $y_{p_{0-1}}$, y_{p_0} and $y_{p_{0+1}}$, this is, two 2 x 2 variogram matrices for the first and last cut-offs, and *p*-2 3 x 3 matrices for the intermediate cut-offs. The modeling of each one of these matrices is performed independently, although some consistency between the resultant *p* LMC should be kept. Restricting in this way the extent of variogram matrices allows a satisfactory fitting of the LMC, without the hindrances of fitting a single full variogram matrix.

The acoIK system is similar to full coIK but restricted to $p_0 - 1$ to $p_0 + 1$ thresholds for each cut-off p_0 (Goovaerts, 1994):

$$\sum_{p'=p_0-1}^{p_0+1} \sum_{\beta=1}^{n} \nu_{\beta,p'}(u_\beta; y_{p_0}) \cdot C_I(u_\alpha - u_\beta; y_{p,} y_{p'}) = C_I(u_\alpha - u_0; y_{p,} y_{p_0})$$
$$\forall \alpha = 1 \text{ to } n, \quad p = p_0 - 1 \text{ to } p_0 + 1$$

And the estimate is calculated by:

$$F^*_{acoIK}(u_0; y_{p_0} \mid (n)) - F(y_{p_0}) = \sum_{p=p_0-1}^{p_0+1} \sum_{\alpha=1}^{n} \nu_{\alpha,p}(u_\alpha; y_{p_0}) \cdot \left[I(u_\alpha; y_p) - F(y_p) \right]$$

Order relation issues are expected after the ccdf estimation for all cut-off’s, but they can be solved in similar way as the indicator Kriging results. Once a valid ccdf is build, this can be used to draw a random number, which should carry the information of the inter-class correlation if the acoIK is implemented in the indicator simulation algorithm.

Status

The `sisim_adj` program exists and has been used for a number of numerical experiments. The motivation for this approach is the impossibility of using an LMC for full cokriging in SIS (the program for that is also available, but has limited applicability). Testing and verification of the results was ongoing at the time of this paper going to press. It is unclear that including data from adjacent cutoffs improves the unnatural variations in SIS.