A Two Step Approach for Block Simulation

S. Elogne[1], O. Leuangthong[1], and J. Ortiz[2]
[1]Centre for Computational Geostatistics – Edmonton, Alberta
[2]Department of Mining Engineering – Santiago, Chile

An important problem in mining is the estimation of global recoverable reserves. A consistent block simulated value requires (i) a fine discretization of the block into a number of nodes, (ii) simulated values at those nodes and (iii) a regularization at the block scale. In presence of large domains with relatively wide spacing data locations, this approach is computationally heavy. A two-step block simulation approach is proposed to handle such problem. The method first partitions the domain into two sets of blocks, uses the point support scales to get simulated values at blocks belonging the first set, that are then used to simulate at second set of blocks. An illustration of the method is given with real data example.

Introduction

Assume that n data $\mathbf{Z} = \left(z(\mathbf{u}_1), \ldots, z(\mathbf{u}_n)\right)$ collected at known spatial locations $\mathbf{u}_1, \cdots, \mathbf{u}_n$ come from a second order stationary process Z defined over a finite domain of interest D with positive volume A central problem in mining exploration, petroleum, reserves estimation, environmental sciences and others, is to simulation/prediction grades at block support scale (Journel and Huijbrets, 1978). Geostatistical simulation is widely used as it provides an assessment of uncertainty and reproduces the spatial variability and distribution of a regionalized variable. A recommended and accurate approach will be to provide L simulated realizations at point support scale before regularizing at the block scale. This approach can prove computationally inefficient, if the number of discretized nodes within the domain is extremely large due to the computer limitations. In such case, the impracticability of approach can lead to a coarser discretization point scale process, where one has to cope with the uncertainty at block scale.

Another approach for global recoverable reserves is the change of support models. From the support v to the support V, (assuming the volume of v smaller that the one of V) this approach (Journel and Huijbregts, 1978) defines the dispersion variance D(v/V) and expresses it in term of the second order stationary semivariogram of the point support as

$$D^2(v/V) = \frac{1}{|V|^2}\int_{V^2} ds\, ds'\, \gamma(\mathbf{s}-\mathbf{s}') - \frac{1}{|v|^2}\int_{v^2} ds\, ds'\, \gamma(\mathbf{s}-\mathbf{s}') \qquad (1)$$

with the double summation being approximated by its discretized form as

$$\frac{1}{|V|^2}\int_{V^2} ds\, ds'\, \gamma(\mathbf{s}-\mathbf{s}') \simeq \frac{1}{n^2}\sum_i \sum_j \gamma(\mathbf{s}_i - \mathbf{s}_j) = \bar{\gamma}_V.$$

If v is a point support volume, then the variance reduction factor defined as the ratio between the variance of the block support and the variance of the point support is given by (Journel and kyriakidis, 2003)

$$f \simeq 1 - \frac{\bar{\gamma}_V}{\sigma^2}$$

and the variance of the block at support V is approximated by

$$\sigma_V^2 \simeq f\, \sigma^2.$$

This paper proposes *a two-step block simulation* (TSBS) method based on simulating the nodes within those blocks that contain sample data based by performing the LU conditional simulation, and then using the previous realizations for simulating blocks realizations on the entire domain. The new approach is proved to be computationally efficient compared to the conventional approach described above which provides L simulated realizations on the entire domain D.

The remainder of the paper is structured as follows. Section 2 reviews the conventional approach considered in this paper. In Section 3, the two-steps block simulation is fully described and Section 4 studies and compares the numerical complexities of the conventional approach and of the two-steps block simulation approach. A real data example is carried out in Section 5 and finally Section 6 summarizes the potential of the method and discusses futures directions.

Conventional Approach Considered (CAC)

One of the consistent and recommended (Journel and Kyriakidis, 2004) geostatistical approach for global reserve recoverable estimation first transformed the original data into a transformed data set

$$\mathbf{Y} = \Big(y(\mathbf{u}_1), \ldots, y(\mathbf{u}_n)\Big) \tag{2}$$

which is normally distributed with n correlated components through a non-decreasing function as

$$y(\mathbf{u}) = G^{-1}[F(z(\mathbf{u}))] \tag{3}$$

where F is an estimator of the cumulative distribution of the collected data and G is the known cumulative Gaussian distribution. The transformed data is then used to simulate L equiprobable realizations at unknown locations of the domain D under the Gaussian space by lower-upper decomposition approach (Alabert, 1987) or by sequential approach (Deutsch and Journel, 1998). Then a back transformation operation of the Gaussian simulated data are conducted through the following

$$z^{(l)}(\mathbf{u}_0) = F^{-1}[G(y^{(l)}(\mathbf{u}_0))] \tag{4}$$

Then a simulated value at the block V at the *l-th* realization is obtained by regularizing the above point scale simulated values over the block V as

$$\begin{aligned} z_V^{(l)}(\mathbf{u}_V) &= \frac{1}{|V|}\int_V d\mathbf{u}\ z^{(l)}(\mathbf{u}) \\ &\simeq \frac{1}{\sum_{k=1}^{n_V}\omega(\mathbf{u}_k)}\sum_{j=1}^{n_V}\omega(\mathbf{u}_j) z^{(l)}(\mathbf{u}_j \in V) \end{aligned} \tag{5}$$

for l=1,..,L where $\omega(\mathbf{u})$ is a kernel function allowing weights to the location $\mathbf{u}$. This methodology provides simulated blocks realization where statistics, like histogram or semivariogram can be computed.

Methodology of the two step block approach (TSBS)

In this section, we describe the two step block simulation method.

Step one of the TSBS

1. Infer relevant statistics of the data, including representative histogram and semivariogram. Declustering or debiasing technique may be required. Simulation is typically based on multivariate Gaussiannity framework. A normal score transformation is performed on the original data. The normal score transformation is a commonly used approach in nonparametric statistics which is defined as values of the order statistics in a sample from a Gaussian distribution. This approach has proved to be a powerful tool in geostatistics when dealing with non Gaussian distribution.

2. Consider a block grid discretization od the entire domain of interest into n_b non-overlapping blocks after a practical block size has been chosen

3. Consider a partition of the domain D into two sets of blocks D_1 and D_2, that is

$$D = D_1 \cup D_2 \tag{6}$$

4. The choice of subsets D_1 and D_2 can be made subjective as many different partitions of the domain can be proposed. We believe that a relevant approach for block's partition is as follows: the subset D_1 called the occupied block set is the set of blocks containing at least one sample data locations. The second part D_2 called the empty block set denotes the set of blocks where no collected data falls in. The cardinality of D_1 and D_2 are denoted by n_o and n_e respectively.
5. A local discretization is then applied inside the blocks D_1 into m nodes and L simulated realizations are accried out on those local dicretized nodes using the CAC approach described above. This gives L simulated realizations at only blocks belonging to D_1, that is we end the first step with L equiprobables simulated values at n_o blocks.

Step two of the TSBS

The second and final step of the TSBS approach is to use the previously simulated blocks of D_1 to simulate at blocks of D_2.

1. Analyze of the spatial variability of the occupied block support over the L realizations is performed as well as the spatial continuity.
2. For each realization of the previous L simulated values at block support, one realization from the SGS/LU algorithm is obtained after common technique. A back transformation of this realization gives a full block realization. The final result produces L simulated values at block support.

Implementation

In this section we study the numerical complexity of the TSBS approach and compare it to the one of the conventional approach, using the LU approach. The numerical complexity of both steps using the SGS algorithm is not discussed here. Assume the following conditions.

- (C1) The sample size n is sufficiently large (acceptable) to obtain relevant and consistent statistics at the point scale.
- (C2) The quantity n_b is sufficiently large (acceptable) to obtain relevant and consistent at the block support scale.
- (C3) The number of occupied blocks n_o is sufficiently large (acceptable) to obtain relevant and consistent at the block support scale.
- (C4) The ratio between the n_o and n_b is small, that is there exists a positive number α<1 such that

$$n_o = \alpha\, n_b. \tag{7}$$

Or equivalently,

$$n_e = (1-\alpha)\, n_b. \tag{8}$$

While conditions (C1) and (C2) can be viewed as minimum requirements, condition (C3) is not straightforward derived from the previous conditions. This assumption (C3) is not too restrictive in practice since economical reasons lead to sparsely collected data. Thus the probability that all the locations fall into a few number of blocks is significantly reduced.

Recall that the lower-upper decomposition simulation requires one inversion of a matrix and two LU decompositions (Alabert, 1987). Recall that a LU decomposition of a positive definite matrix A of size K has a numerical complexity that scales proportionally to third power of K. Similarly any inverse matrix operation of matrix A of size K has a numerical complexity that scales proportionally to third power of K.

Numerical Complexity of the Conventional Approach

Since the number of discretized nodes into each block is m, the numerical complexity of the CAC using the LU decomposition is as

$$T_1 = O\left(n_b^3 m^3\right) + O(n^3)$$

which is clearly reduced to

$$t_{cac} = O\left(n_b^3 m^3\right) \tag{9}$$

which bears an heavy computational burden if the product of the number of blocks n_b by the number of nodes m is large. The computer storing process is not discussed here.

Numerical Complexity of the TSBS Approach

As the TSBS method uses two steps, the computational time for the first step is

$$T_2' = O\left(n_o^3 m^3\right) + O(n^3) + O(n^3)$$

while the one of the second step is given as

$$T_2'' = O\left(n_e^3\right) + O(n_o^3) + O(n_o^3).$$

Hence the computation time for the TSBS approach is deduced as

$$\begin{aligned} t_{tsbs} &= O\left(n_o^3 m^3\right) + O\left(n_e^3\right) \\ &= \alpha^3 O\left(n_b^3 m^3\right) + (1-\alpha)^3 O\left(n_b^3\right) \qquad \left[\text{from the equations (7) and (8)}\right] \\ &= \alpha^3 O\left(n_b^3 m^3\right) \\ &= \alpha^3 O\left(t_{cac}\right) \end{aligned} \tag{10}$$

A comparison between the times consuming for the CAC and TSBS approaches is obtained through the ratio α. If α=1, that is n_e=0, then the computational times are the same. If α<1/2 or equivalently if n_b>2n_o, then a clearly computational advantage is attributed to the TSBS approach.

Real Data Application: Walker Lake Data set

All the programs are running under the Gslib software (Deustch and Journel, 1998). We present a 'well-known' Walker Lake data example to demonstrate the TSBS approach described in the paper. Any computational comparison has not been conducted here since proofs for time consuming of both approaches have been provided. In this section, we compare the performances of both methods described in the paper. The data consists of a training set of n=470 samples collected over a domain D=[0,260m]x[0,300m]. Figure (1) shows relevant information about the data, like the locations map of the data and the histogram. For simplicity and for a comparison purpose, the omni directional semivariogram modeling is considered.

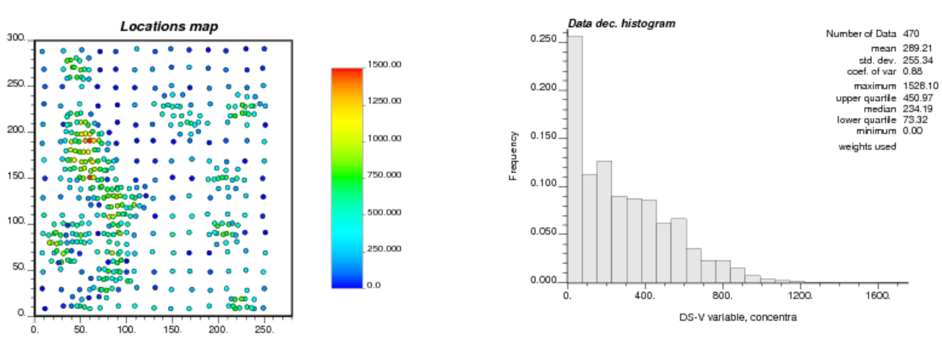

Figure 1: Data configuration and empirical histogram estimates of the collected data

Now a practical block grid discretization of the domain D by 26x30 is considered with a constant nodes discretization by 5x5 within each block, whilst the number of simulated realizations is L=100. For simplicity, an isotropic assumption will be formulated as the goal of this paper is to assess the proposed methodology and to compare it to the conventional approach described above. The normally transformed data set with n correlated components through has a fitted semivariogram model given by

$$\gamma_{NS}(\mathbf{h}) = 0.135 + 0.865\,\mathrm{S}phe_{36.9}(\|\mathbf{h}\|) \tag{12}$$

as described in Figure (2).

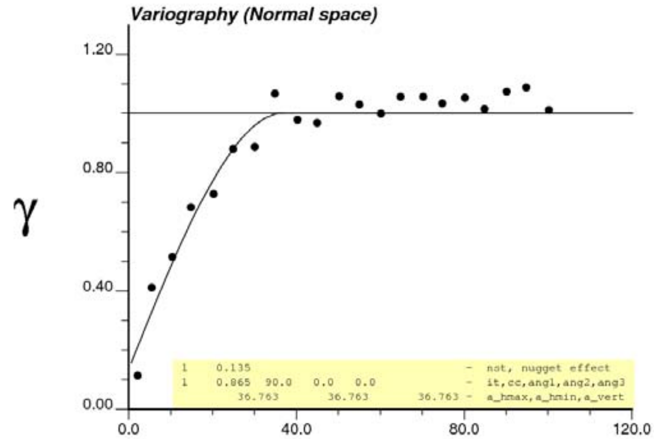

Figure 2: Empirical semivariogram estimates and the superimposed fitted model of the normal score data.

The CAC employing the sequential Gaussian algorithm (SGS) to simulate at the point support before regularizing at the desired block scale as is fully computed and described in Figure (3). The TSBS approach is now performed. The number of blocks obtained after selecting our practical block grid size discretization of 26x30 gives obviously n_b=780 leading to

$$n_o = 298, \quad n_e = 482, \quad \frac{n_o}{n_b} = \alpha \simeq 0.38 \quad \text{and} \quad \alpha^3 \simeq \frac{1}{20}. \tag{13}$$

This later results indicates that the TSBS approach can be approximately 20 times much faster than the CAC, which can be significant if the TSBS times consuming exceeds the hour. Figure (4) describes the first step of the TSBS approach while Figure (5) gives relevant information on the spatial dispersion of the blocks realizations of D_1. Comparisons between histograms estimates and cumulative distribution function estimates using the CAC method and the first step of TSBS are plotted in Figure (6). Finally Figure (7) provided some final results of the TSBS after the second step has been performed. Quantile quantile plots between arbitrarily chosen realizations are also provided.

Both reference distributions at block support are compared in Figure (8) showing a good match of the approach. The new approach gives a higher probability of the first class than does the CAC. We may suspect the uncertainty brought by the variogram model used for each partial block support realization and the impact of the seed. Indeed, one simulation from one simulated value using the sequential Gaussian algorithm may be investigated in more detail to analyze the sensitivity of the variogram model employed. At the end, Figure (9) studies the spatial variablity of both approaches. That plot indicates some good matching between on the average semivariogram estimates at block support.

Discussion and Future Work

A new approach for fast recoverable estimation at block support scale is proposed. This method uses a conventional approach to simulate realizations at blocks containing only observed data locations and uses those realizations to obtain blocks on the entire domain. Although it is proved to be computationally efficient, the following points need to be addressed.

- On the first hand, discretizing only occupied blocks and taking those blocks as realizations for inferring the whole domain, may bear a huge uncertainty if the number of occupied blocks n_o is not 'enough' for an 'acceptable modeling.' This case may happen in presence of small sample data. In this case, some empties blocks should be considered as occupied blocks to get a reasonable blocks inference process at the second step of the TSBS method.
- On the other hand, discretizing occupied blocks and simulating using the LU decomposition conditionally to the data is not always possible, depending of the number of blocks considered and the number of nodes in each block (large number of occupied blocks for instance). In such case, some occupied blocks should be considered as empties blocks before the simulation in the first step of the TSBS method takes place.
- In the general case, a flexible and rigorous algorithm (up to the practitioner) for the selection blocks in the first step of the TSBS should be proposed in the future.

References

Alabert. F., (1987). The practice of fast conditional simulations through the LU decomposition of the covariance matrix. Mathematical coregionalization: Mathematical Geology, Geology, Vol. 19, No 5, 1987.

Davis, M., W. (1983). production of conditional simulations via the LU decomposition of the covariance matrix. Mathematical Geology, vol. 19, n.2, page 91-98.

Cressie, N., 1993, Statistics for spatial data: Wiley, New York, 900 p.

Deutsch, C., V. (2002) Geostatistical Reservoir Modeling. Oxford University Press, 2002.

Deutsch C.V. and Journel A.G., 1998, GSLIB: Geostatistical Software Library and User's Guide, 2nd edition, Oxford University Press, New YorkDeutsch, C.V. (2002) *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 376 p.

Deutsch C.V., 2006. A sequential indicator simulation program for categorical variables with point and block data: BlockSIS. Computers \& Geosciences, 32, p. 1669-1681.

Elogne, S. N., Perrin, O. and Thomas-Agnan, C., (2008). Non parametric estimation of smooth stationary covariance functions by interpolation methods. Statistical Inference for Stochastic Processes, vol. 11, p. 177-205.

Elogne, S. N., Hristopulos, D. T. and Varouchakis, M. (2008). An application of Spartan spatial random fields in environmental mapping: focus on automatic mapping capabilities. Stochastic Environnemtal Research and Risk Assessment, vol. 22, p. 633-646.

Emery X., (2006). Two ordinary kriging approaches to predicting block grade distributions. Mathematical Geology, vol. 38, No 7, p. 801-819.

Journel, A. and Huijbregts, J. (1978) Mining Geostatiscis, London: Academic Press, 600 p.

Journel, A. and Kyriakidis, P. C. (2004). Evaluation of Mineral Reserves: A Simulation Approach, Oxford University Press.

Ortiz J., Leuangthong O. and Deutsch C. V A MultiGaussian Approach to Assess Block Grade Uncertainty, CIM Conference and Exhibition, Edmonton 2004, May 9 to 12, 2004.

Verly, G., 1984. The block distribution given a point multivariate normal distribution. In G. Verly, M. David, A. G. Journel, and A. Marechal, editors.

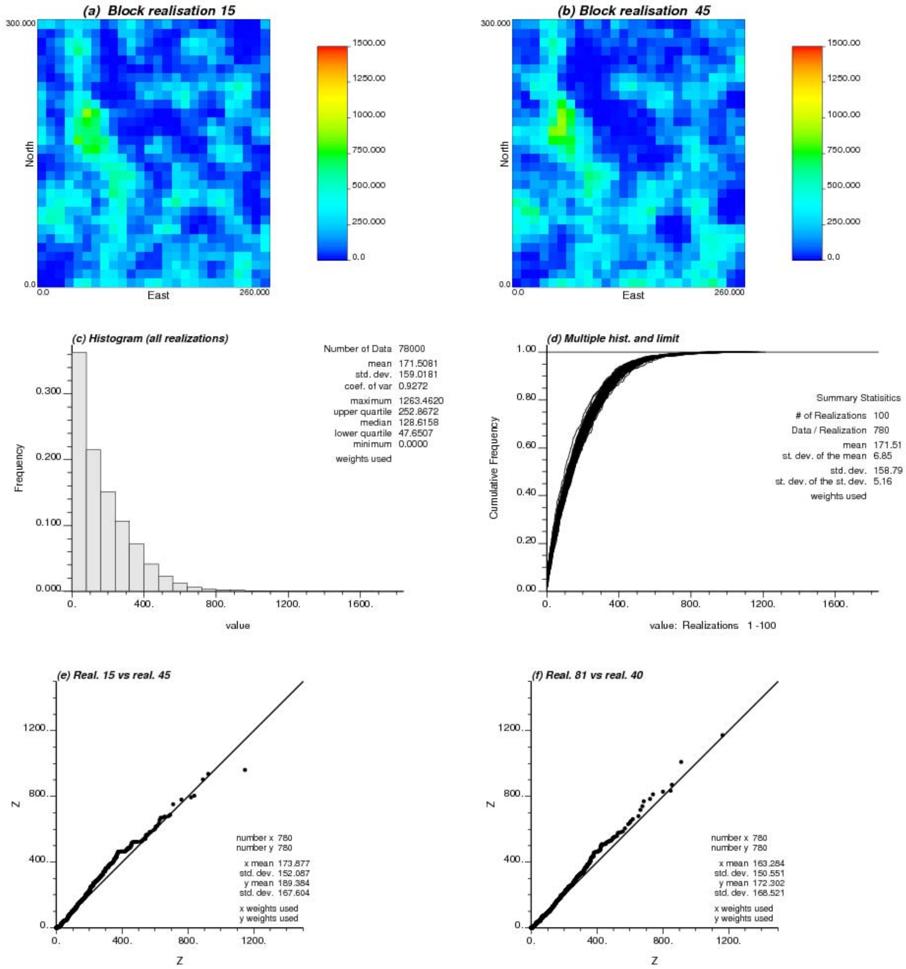

Figure 3: The CAC approach shown in six plots after the point support simulation and regularization processes have been carried out; (a) realization 15; (b) realization 45; (c) empirical histogram reproduction at block scale using the L=100 simulated realizations; (d) empirical cumulative distribution function at block scale using the L realizations; (e) quantile quantile plot of block realization 15 against 45 and (f) quantile quantile plot of realization 81 against realization 40.

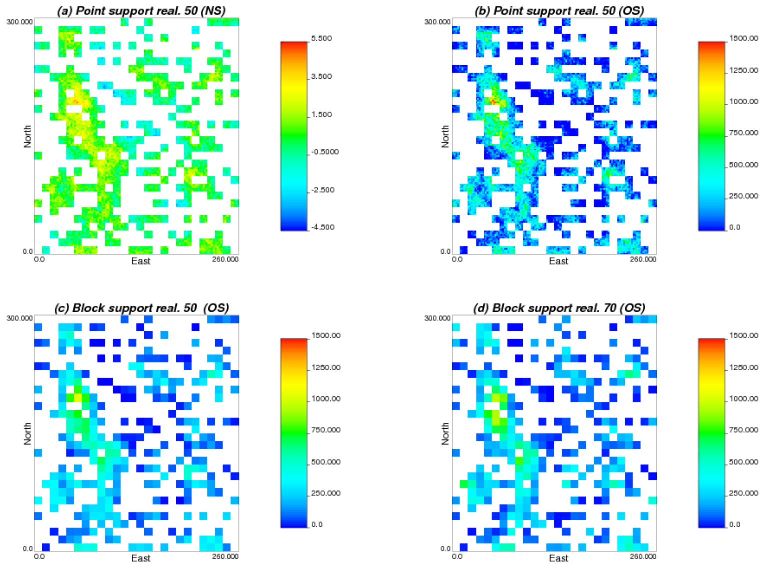

Figure 4: Description of the first step of the TSBS approach given in four plots. The part of the domain D with no data is clearly the empty part block set D_2 while the part with plotted data is the D_1 block set. (a) Point support realization 50 under the normal space (NS) at discretized nodes of blocks of D_1; (b) Back transformation of the normally simulated realization given in (a) to the original space (OS); (c) regularization of point support realization given in (b) to the block scale yielding to the realization 50 of blocks of D_1; realization 70 at blocks of D_2.

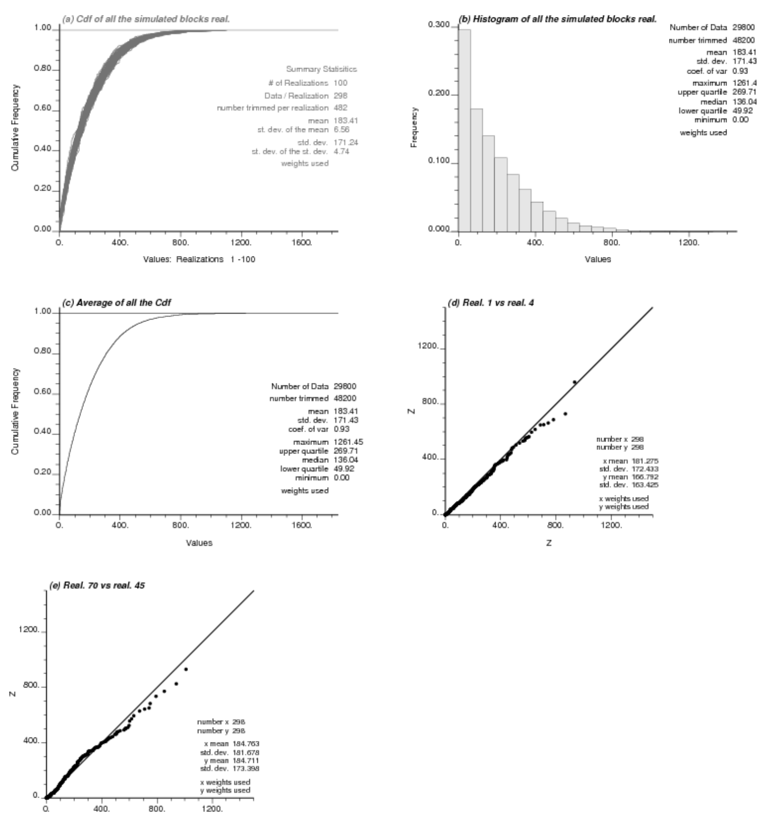

Figure 5: Study of the spatial dispersion of the blocks simulated realizations of D_1 obtained from the first step of the TSBS method. (a) Plot of all the L=100 estimates at blocks of D_1 ; (b) histogram estimates using all the L simulated realizations at blocks of D_1 ; (c) average of all the c.d.f. estimates given in (a); (d) quantile quantile plot of blocks (of D_1) realizations 1 and 4 and (e) same as in (d) by using the realizations 70 and 45.

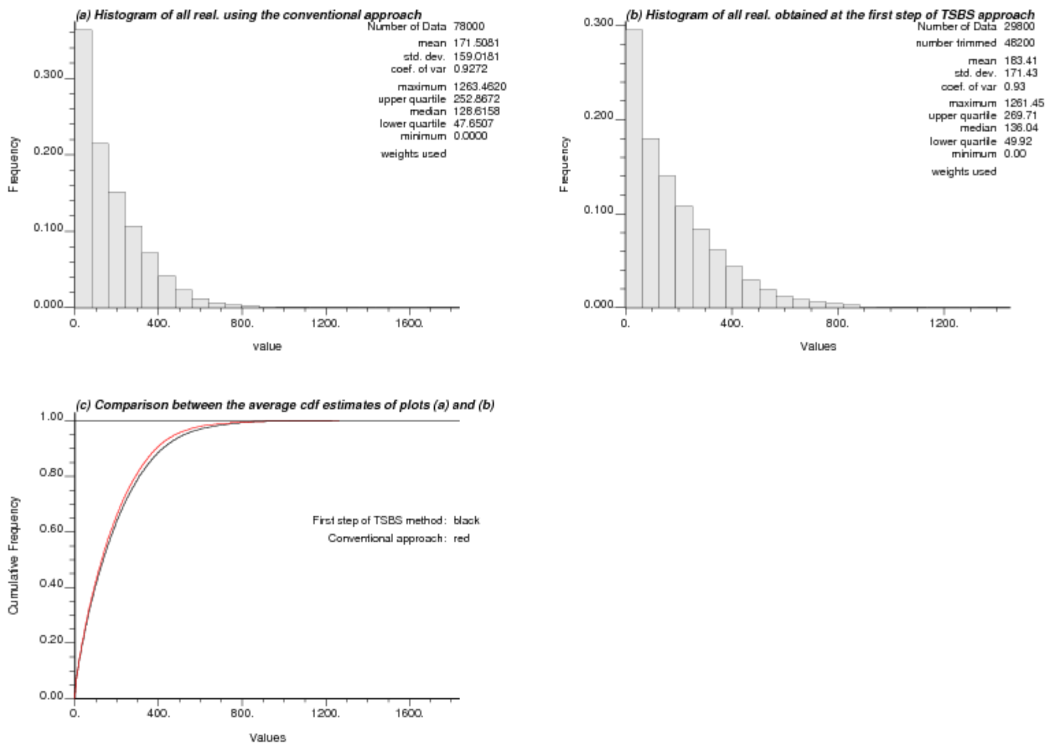

Figure 6: Comparison study between the block distribution obtained from the CAC and the distribution obtained at the first step of the TSBS approach. (a) Histogram estimates of all realizations at block scale using the CAC; (b) Histogram estimates of all realizations at block scale using the first step of the TSBS method; (c) comparison between the averages of all the c.d.f. estimates using the blocks realizations obtained for the CAC and those obtained from the first step of the TSBS method.

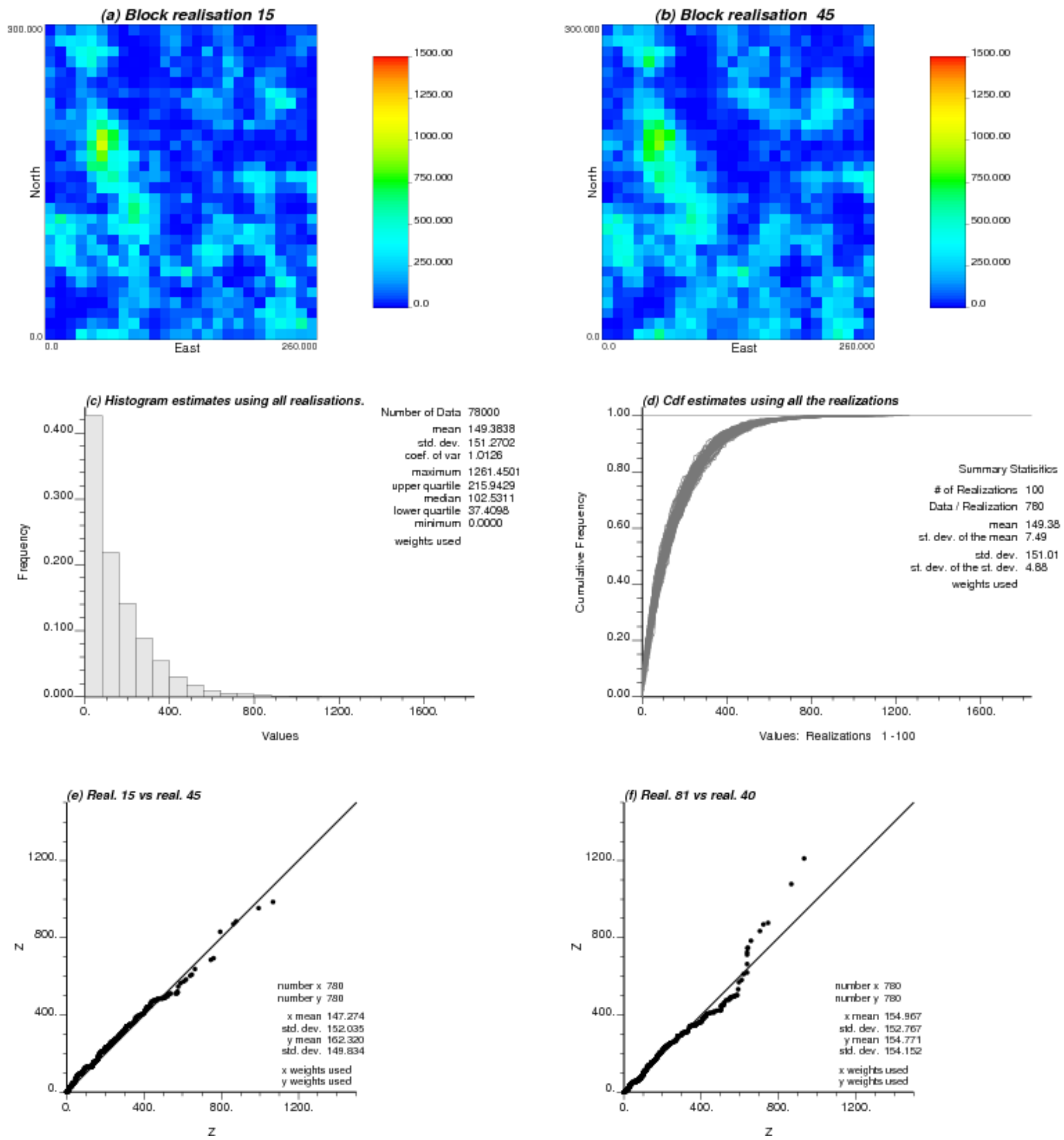

Figure 7: The second step of TSBS method. (a) Realization number 15; (b) realization number 45; (c) empirical histogram reproduction at block scale using all the L=100 simulated realizations; (d) empirical cumulative distribution function at block scale using all the L=100 simulated realizations; (e) quantile-quantile plot of block realizations 15 and 45 and (f) same as in (e) by using realizations 81 and 40.

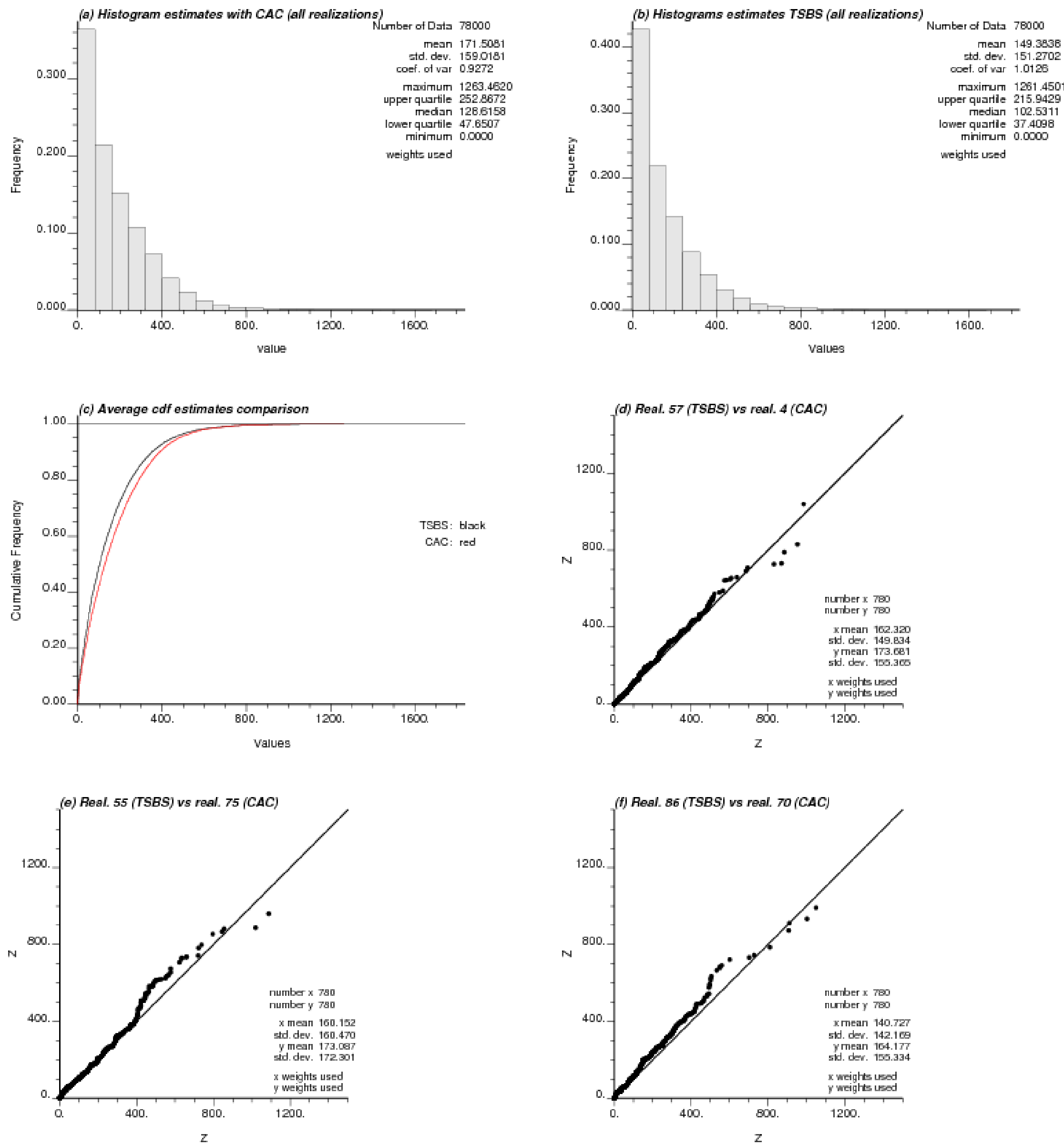

Figure 8: Comparison study of the distribution between CAC and the TSBS method plotted in (a) histogram average from CAC; (b) histogram average from TSBS; (c) average cumulative distributions estimates and (d), (e) and (f) compare some realization obtained from both approaches.

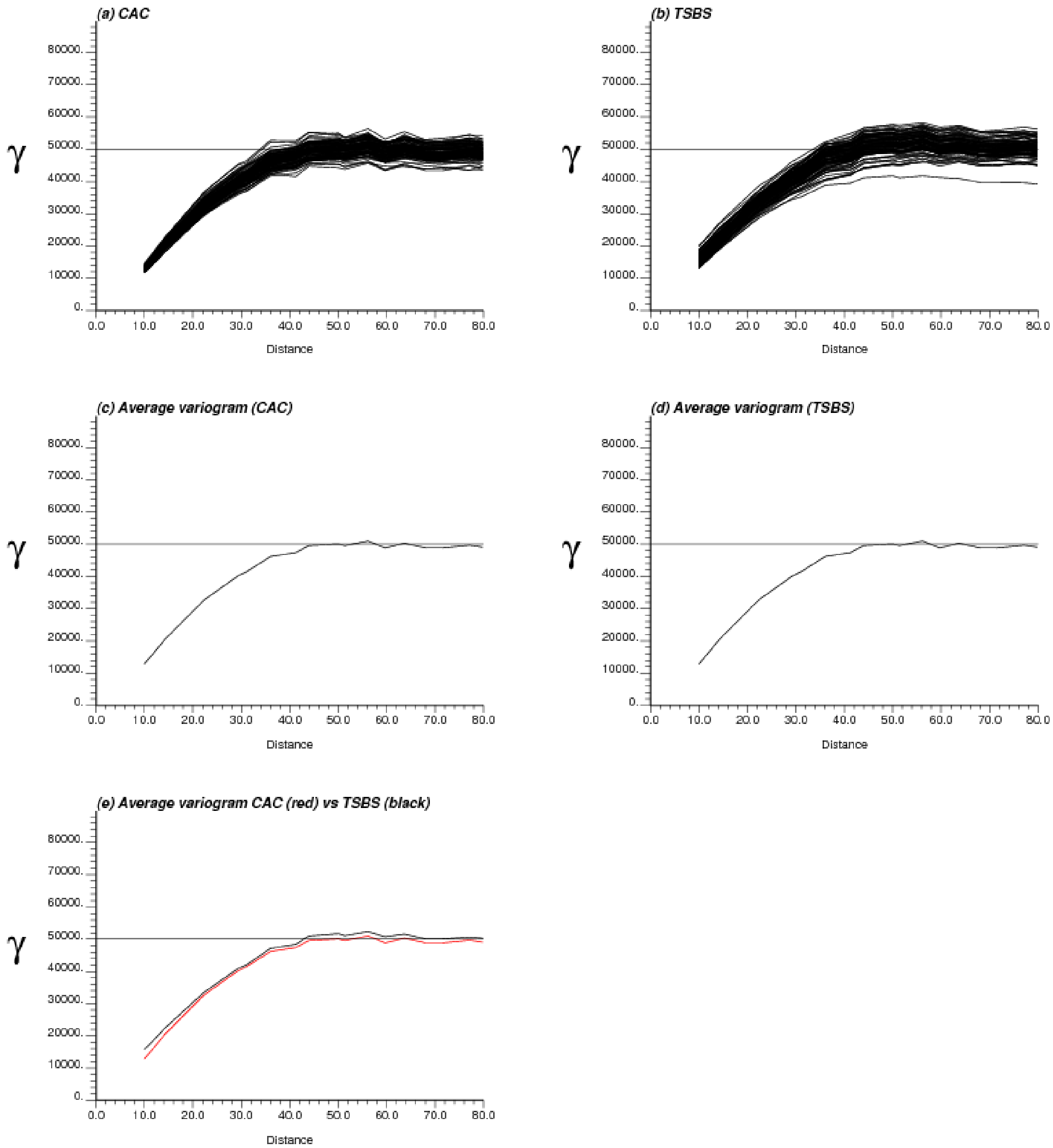

Figure 9: Spatial variability comparison between the CAC and the TSBS. (a) All the L semivariogram estimates using the CAC; (b) all the L semivariogram estimates using the TSBS method; (c) average over the L semivariogram estimates from (a); average over the L semivariogram estimates from (b); (e) comparison between (c) and (d).