*Mathematical Geology, Vol. 33, No. 8, November 2001 (© 2001)*

# Direct Sequential Simulation and Cosimulation[1]

**Amilcar Soares[2]**

*Sequential simulation of a continuous variable usually requires its transformation into a binary or a Gaussian variable, giving rise to the classical algorithms of sequential indicator simulation or sequential Gaussian simulation. Journel (1994) showed that the sequential simulation of a continuous variable, without any prior transformation, succeeded in reproducing the covariance model, provided that the simulated values are drawn from local distributions centered at the simple kriging estimates with a variance corresponding to the simple kriging estimation variance. Unfortunately, it does not reproduce the histogram of the original variable, which is one of the basic requirements of any simulation method. This has been the most serious limitation to the practical application of the direct simulation approach. In this paper, a new approach for the direct sequential simulation is proposed. The idea is to use the local sk estimates of the mean and variance, not to define the local cdf but to sample from the global cdf. Simulated values of original variable are drawn from intervals of the global cdf, which are calculated with the local estimates of the mean and variance. One of the main advantages of the direct sequential simulation method is that it allows joint simulation of* $N_v$ *variables without any transformation. A set of examples of direct simulation and cosimulation are presented.*

**KEY WORDS:** stochastic simulation, multivariable joint simulation, collocated cokriging.

## INTRODUCTION

Given its simplicity of implementation, sequential simulation has become the most common and popular stochastic algorithm to reproduce the spatial distribution and uncertainty of variables of different resources in Earth sciences. Although they use the same sequential procedure, the different versions of sequential simulation require transformations of original variables and different approaches to estimate local distribution functions.

Direct sequential simulation, which means without any transformation of the original variable, has been used widely for spatial characterization of categorical variables—e.g., lithotypes, soil types, forest species, etc. (Soares, 1998). For sequential simulation of continuous variables (e.g., grades, porosity, etc.), the other approaches—sequential indicator simulation (SIS) and sequential Gaussian

[1]Received 5 June 2000; accepted 27 November 2000.

[2]CMRP – Instituto Superior Técnico, Av. Rovisco Pais, 1096 Lisboa Codex; e-mail: ncmrp@alfa.ist.utl.pt



0882-8121/01/1100-0911$19.50/1 

simulation (SGS)—require transformation of the original variables into a set of indicator variables or a standard Gaussian variable. Local estimation of the probability distributions is performed within the indicator or the multiGaussian formalism, respectively. The pros and cons of the use of each of these algorithm for the spatial characterization of continuous variables has been reported extensively (see for example Goovaerts, 1997). It is worth mentioning that the main drawbacks of both methods are direct or indirectly related to the need to transform the original variable. Journel (1994) introduced the first step for the simulation of a continuous variable without any prior transformation of it: he showed that the direct simulation of a continuous variable succeeded in reproducing the covariance model, provided that simulated values are drawn from local distributions centered at the simple kriging estimates with a variance corresponding to the simple kriging estimation variance.

This, called Journel's theorem (Caers, 2000), guarantees that the spatial covariance of the original variable is reproduced but not the histogram, which is one of the main requirements of any simulation algorithm. This has been the most serious limitation on the practical application of the direct sequential simulation approach.

Caers (1999) proposed the use of postprocessing in order to transform the resulting "pseudo" simulated values into another set of values, which reproduces approximately the histogram of original variable and preserves the data exactitude. However this posterior transformation can, in some situations, destroy the variogram reproduction.

Caers (2000) proposed a direct indicator simulation by introducing a set of linear constraints after the kriging of the mean and variance, thereby avoiding the posterior transformation.

In this paper a new approach is proposed for the direct sequential simulation based on the principle introduced by Journel (1994). This simple algorithm succeeds in reproducing the variogram and histogram of a continuous variable. But the main advantage of the proposed algorithm is that it allows the cosimulation procedure without calling for any transformation of the original variables.

## DIRECT SEQUENTIAL SIMULATION

Let us consider the continuous variable $Z(x)$ with a global cdf $F_z(z) = \text{prob}\{Z(x) < z\}$ and stationary variogram $\gamma(h)$. We intend to reproduce both $F_z(z)$ and $\gamma(h)$ in the final simulated maps.

The sequential simulation algorithm of a continuous variable follows the classical methodological sequence:

i) Randomly select the spatial location of a node $x_u$ in a regular grid of nodes to be simulated.
ii) Estimate the local cumulative distribution function at $x_u$, conditioned to the original data $z(x_\alpha)$ and the previous simulated values $z^s(x_i)$.

iii) Draw a value $z^s(x_i)$ from the estimated local cdf.
iv) Return to step (i) until all nodes have been visited by the random path.

In step (ii) the estimation of local cdf is usually performed with the indicator formalism (SIS) or with the multiGaussian approach (SGS), both requiring transformation of the original variable.

The principle behind direct sequential simulation can be summarized as follows:

If the local cdfs are centered at the simple kriging estimate

$$z(x_u)^* = m + \sum_{\alpha} \lambda_\alpha(x_u)[z(x_\alpha) - m] \tag{1}$$

$x_\alpha$ being the conditioning data (original and previously simulated values), with a conditional variance identified with the simple kriging variance $\sigma^2_{sk}(x_u)$; no matter what probability distribution we choose, the spatial covariance model or variograms are reproduced in the final simulated maps. Bourgault (1997) and Caers (1999) have given a practical demonstration of this statement with various types of distributions. The problem is, except for a few parametric distributions (e.g., Gaussian), this simulation approach does not reproduce the histograms. The main reason for this problem is that the local cdf can not be fully characterized only by the local mean and variance.

The idea proposed in this paper is to use the estimated local mean and variance, not to define the local cdf but to sample from the global cdf. It is a similar procedure to the sequential indicator simulation: in this algorithm the global histogram remains with the same number of classes in any sequential step; locally the conditioning data determine which classes will be sampled to generate a new simulated value. For example, at a given step of the SIS procedure, suppose the local neighborhood values of $x_u$ belong just to the first two classes of a ten-class histogram. Consequently, the simulated value $z^s(x_u)$ is normally drawn from those two classes.

In the proposed algorithm the cumulative distribution function $F_z(z)$ is the same throughout the sequential procedure. Intervals of $z$ are chosen from $F_z(z)$, defining a new $F'_z(z)$ and then simulated values $z^s(x_u)$ are sampled from the chosen distribution $F'_z(z)$. The intervals are "centered" at the simple kriging estimate $z(x_u)^*$, the interval range being dependent on the SK estimation variance $\sigma^2_{sk}(x_u)$.

One way to define these intervals is to select a subset of $z(x_i)$ values of the experimental histogram (the one we wish to reproduce in final simulated maps) in such way that the mean and variance of the selected $n$ values $z(x_i)$ is equal to the local SK estimate $z(x_u)^*$ and SK variance $\sigma^2_{sk}(x_u)$ respectively:

$$\frac{1}{n}\sum_{i=1}^{n} z(x_i) = z(x_u)^* \qquad \frac{1}{n}\sum_{i=1}^{n} [z(x_i) - z(x_u)^*]^2 = \sigma^2_{sk}(x_u)$$

Then, a simulated value $z^s(x_u)$ is drawn from the $F'_z(z)$ of the $n$ selected values.

*Note*. In extreme situations of no spatial structure (spatial covariance as a pure nugget effect) those intervals are of equal length, or the equivalent distributions $F'_z(z)$ of equal variance. It means that once the intervals, or $F'_z(z)$, are chosen randomly from $F_z(z)$, this corresponds to the composition method (Ripley, 1987) of drawing values from a global cdf (see Fig. 3).

Another way, which is much simpler to implement, to define the range of such sampling intervals of the global cdf, according to the value $\sigma^2_{sk}(x_u)$, is by using a Gaussian distribution.

Suppose $\varphi$ is the normal score transform of original $z(x)$ values

$$y(x) = \varphi(z(x)) \quad \text{with} \quad G(y(x)) = F_z(z(x)) \tag{2}$$

The local SK estimate $z(x_u)^*$ has the equivalent Gaussian value $y(x_u)^*$, $y(x_u)^* = \varphi(z(x_u)^*)$ which, together with the SK standardized estimation variance $\sigma^2_{sk}(x_u)$, can define a Gaussian cdf $G(y(x_u)^*, \sigma^2_{sk}(x_u))$.

This Gaussian cdf can be used to define the interval of the cdf of $z(x)$ which must be sampled:

Generate a value $p$ from a uniform distribution $U(0,1)$
Generate a value $y^s$ from $G(y(x_u)^*, \sigma^2_{sk}(x_u))$

$$y^s = G^{-1}\left(y(x_u)^*, \sigma^2_{sk}(x_u), p\right) \tag{3}$$

Finally, a simulated value $z^s(x_u)$ is obtained by the inverse transform $\varphi^{-1}$:

$$z^s(x_u) = \varphi^{-1}(y^s) \tag{4}$$

This means that $z^s(x_u)$ is sampled from intervals of $F_z(z)$ defined by the local estimates $z(x_u)^*$ and $\sigma^2_{sk}(x_u)$.

It is important to note that

1) the Gaussian transformation is used solely for sampling intervals of the distribution $F_z(z)$. It does not have any role in the estimation of local cdf; hence, no Gaussian hypothesis of the transformed values is assumed. The entire sequential procedure is performed with the original variable $Z(x)$.
2) Because we use a nonlinear function $\varphi$ to obtain the intervals of global cdf, we can not guarantee that the expectation of $z^s(x_u)$ is equal to $z(x_u)^*$:

$$\mathrm{E}\{y^s\} = y(x_u)^* = \varphi(z(x_u)^*) \quad \text{but not} \quad \mathrm{E}\{z^s(x_u)\} = z(x_u)^*.$$

This theoretical limitation of the sampling method did not show any practical impact in most of the cases, and it has to be balanced against its simplicity of implementation. However, in those situations when the histogram of $z(x)$, used for normal score transform (2), has few data, mostly in low frequency classes, the following bias can occur: $\mathrm{E}\{z^s(x_u)\} \neq z(x_u)^*$. This means that we are trying to do

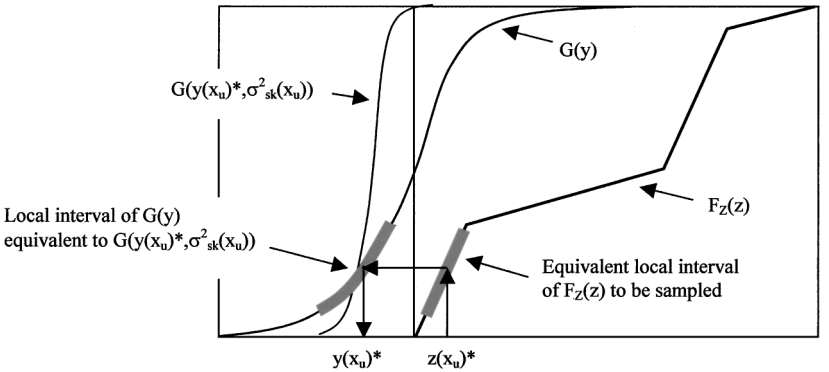


**Figure 1.** Sampling of global distribution $F_z(z)$ by intervals defined by the local mean and variance of $z(x_u)$: The value $y(x_u)^*$ corresponds to the local estimate $z(x_u)^*$. The simulated value $z^s(x_u)$ drawn from the interval of $F_z(z)$ is defined by $G(y(x_u)^*, \sigma^2_{sk}(x_u))$.

a correspondence (4) between a Gaussian cdf centered at $y^*(x_u)$ and a $Z(x_u)$ cdf which is not centered at $z^*(x_u)$ given the scarcity of data. In Appendix a correction for the local bias, of these particular cases, is proposed.

Figure 1 illustrates the sampling of global distribution $F_z(z)$ by intervals defined by the local mean and variance of $z(x_u)$. The value $y(x_u)^*$ corresponds to the local estimate $z(x_u)^*$. The simulated value $z^s(x_u)$ is drawn from the interval of $F_z(z)$ defined by $G(y(x_u)^*, \sigma^2_{sk}(x_u))$.

After calculating the normal score transform of $z(x)$ (for example with a graphical method, Goovaerts, 1997) the direct sequential simulation (DSS) can be described in the following steps:

1. Define a random path over the entire grid of nodes $x_u$, $u = 1$, $N_s$, to be simulated.
2. Estimate the local mean and variance of $z(x_u)$, identified, respectively, with the simple kriging estimate $z(x_u)^*$ and estimation variance $\sigma^2_{sk}(x_u)$ conditioned to the experimental data $z(x_i)$ and previous simulated values $z^s(x_i)$.
3. Define the interval of $F_z(z)$ to be sampled, by using the Gaussian cdf:

$$G\big(y(x_u)^*, \sigma^2_{sk}(x_u)\big), \quad \text{where } y(x_u)^* = \varphi(z(x_u)^*).$$

4. Draw a value $z^s(x_u)$ from the cdf $F_z(z)$.
   Generate a value $p$ from a uniform distribution $U(0,1)$
   Generate a value $y^s$ from $G(y(x_u)^*, \sigma^2_{sk}(x_u))$: $y = G^{-1}(y(x_u)^*, \sigma^2_{sk}(x_u), p)$
   Return the simulated value $z^s(x_u) = \varphi^{-1}(y^s)$.
5. Loop until all $N_s$ nodes have been visited and simulated.

## Examples

Figure 2 shows the spatial location of a set of 50 samples with the histogram (bi-modal) which is the target of the simulation procedure.

The results of the presented DSS method are shown in three different situations:

a) Simulation of a pure nugget effect variogram without taking into account the data (nonconditional simulation). The histogram of simulated values perfectly matches the target (Fig. 3).

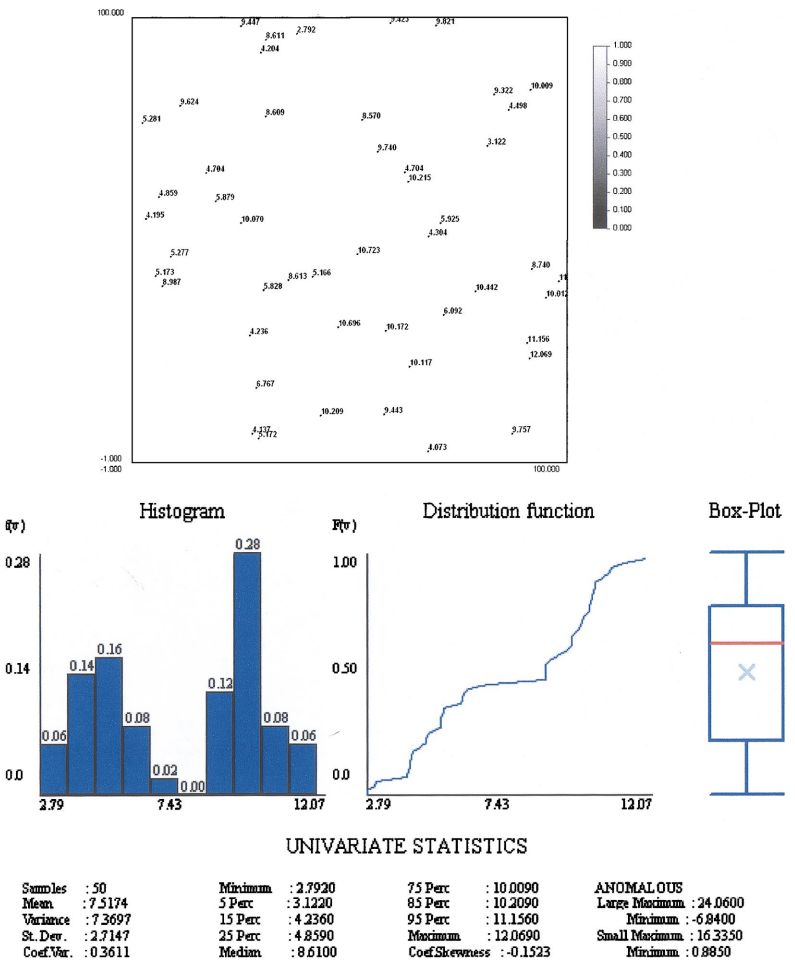


**Figure 2.** Spatial location of a set of 50 samples and the histogram.

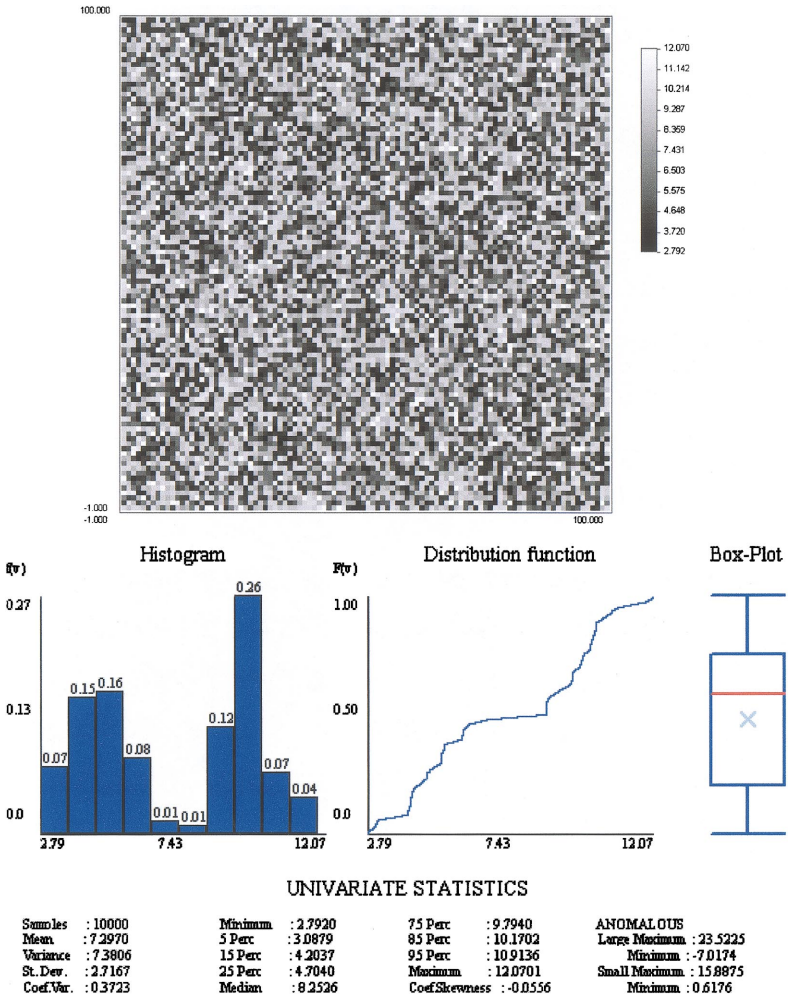


**Figure 3.** Simulation of a pure nugget effect variogram without taking into account the data (nonconditional simulation). Histogram of simulated values.

b) Simulation of a 20 m range (1/5 of the side of area A) isotropic variogram without taking into account the data. Figure 4 shows one simulated realization and the corresponding histogram and variogram.
c) Simulation of the same variogram as (b) but conditioned to the data. Figure 5 shows one simulated realization and the corresponding histogram and variogram.

In all situations there is a satisfactory match of the targets.

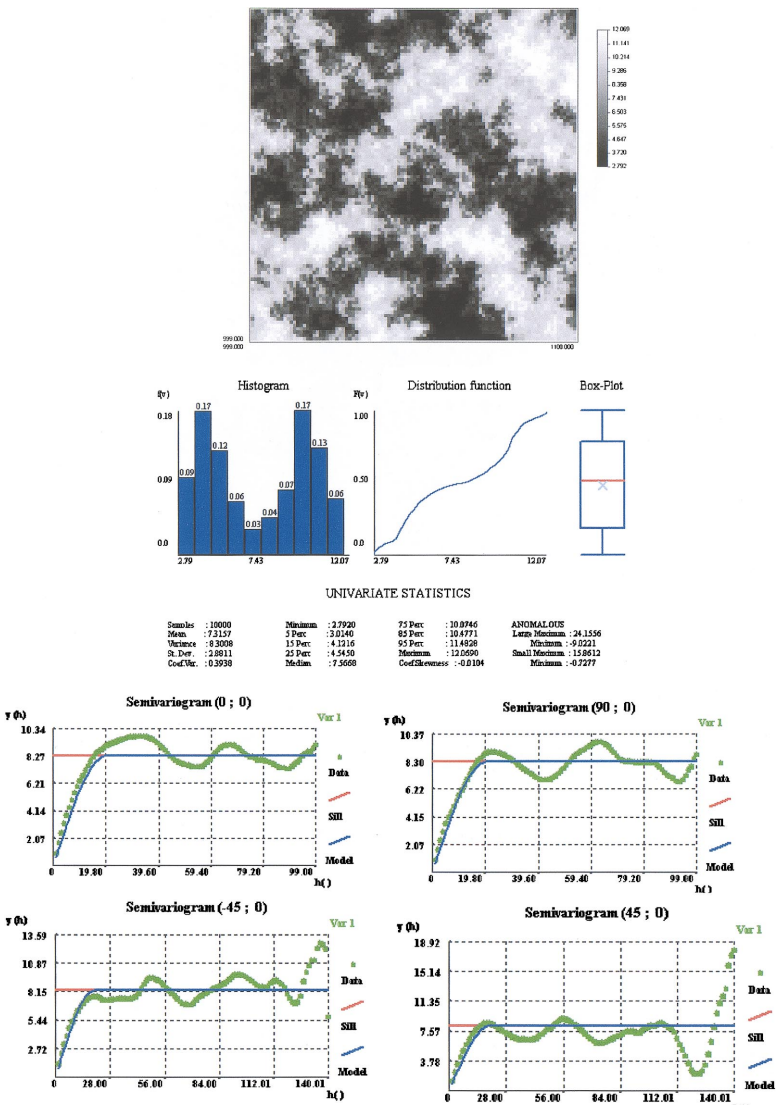


**Figure 4.** Simulation of a 20 m range (1/5 of the side of area A) isotropic variogram without taking into account the data (nonconditional simulation). Histogram and variograms of simulated values.

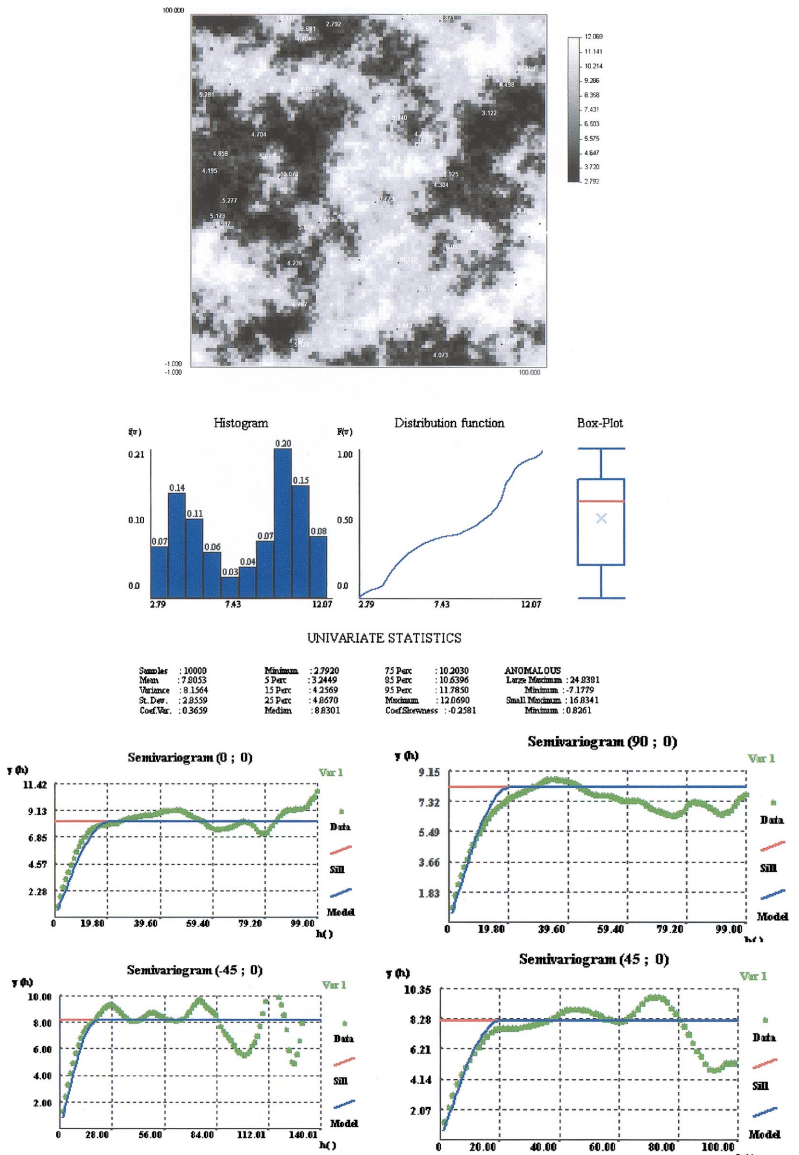


**Figure 5.** Simulation of a 20 m range (1/5 of the side of area A) isotropic variogram, taking into account the 50 samples. Histogram and variograms of simulated values.

## DIRECT SEQUENTIAL COSIMULATION

One of the main advantages of the proposed algorithm over traditional SIS and SGS is that it allows a joint simulation dealing directly with the original variables. Instead of simulating $N_v$ variables simultaneously, each variable is simulated in turn conditioned to the previously simulated variable (Gomez-Hernandez and Journel, 1993; Goovaerts, 1997).

Suppose just two variables, $Z_1(x)$ and $Z_2(x)$. Choosing the primary variable, say $Z_1(x)$, as the most important or with a more evident spatial continuity (Almeida and Journel, 1994), the joint simulation algorithm is described in detail as follows:

1. Define a random path visiting each node of a regular grid of nodes.
2. At each node $x_u$
   Simulate the value $z_1^s(x_u)$ using the DSS algorithm described in step 2 above:
   - Identify the local mean and variance of $z_1(x)$ as the SK estimate and estimation variance $z_1(x_u)^*$ and $\sigma^2_{sk}(x_u)$; calculate $y(x_u)^* = \varphi_1(z(x_u)^*)$, $\varphi_1$ being the normal score transform of the primary variable $z_1(x)$;
   - Generate a value $p$ from a uniform distribution $U(0, 1)$;
   - Generate a value $y^s$ from $G(y(x_u)^*, \sigma^2_{sk}(x_u))$: $y^s = G^{-1}(y(x_u)^*, \sigma^2_{sk}(x_u), p)$;
   - Return the simulated value $z_1^s(x_u) = \varphi_1^{-1}(y^s)$ of the primary variable.

   The same DSS algorithm is applied to simulate $Z_2(x)$ assuming the previously simulated $Z_1(x)$ as the secondary variable. Colocated simple cokriging is used to calculate $z_2(x_u)^*$ and $\sigma^2_{sk}(x_u)$ conditioned to neighborhood data $z_2(x_\alpha)$ and the colocated datum $z_1(x_u)$ (Goovaerts, 1997):

$$z_2(x_u)^*_{sck} = \sum_{\alpha=1}^{N} \lambda_\alpha(x_u)[z_2(x_\alpha) - m_2] + \lambda_\beta(x_u)[z_1(x_u) - m_1] + m_2 \quad (5)$$

   - Transform $y(x_u)^* = \varphi_2(z_2(x_u)^*)$. $\varphi_2$ is the normal score transform of the $Z_2(x)$ variable.
   - Generate a value $p$ from a uniform distribution $U(0, 1)$;
   - Generate a value $y^s$ from $G(y_2(x_u)^*, \sigma^2{}_{sk}(x_u))$: $y^s = G^{-1}(y_2(x_u)^*, \sigma^2_{sk}(x_u), p)$;
   - Return the simulated value $z_2^s(x_u) = \varphi_2^{-1}(y^s)$ of the secondary variable.
3. Loop until all nodes are simulated.

This methodology can be extended to the joint simulation of $N_v$ variables.

## Example

For illustration purposes, direct cosimulation was performed with just two variables. Figure 6 shows the histogram and the variograms of the (nonconditional) simulation of primary variable.

Three direct cosimulation tests were run to obtain a secondary variable with different correlation coefficients: $\rho = 0.9$, $\rho = 0.7$, and $\rho = 0.5$. For sake of simplicity the collocated cokriging was applied with the Markov-type approximation (Goovaerts, 1997), i.e., only the primary variable variogram and the correlation coefficient between primary and secondary variable is needed. Figure 7 shows the map of simulated primary variable and the simulated secondary variable maps for the 3 correlation coefficients.

All results show a good match of the targets of the joint simulation. Figure 8 presents the histogram and variograms of simulated secondary variable for $\rho = 0.7$.

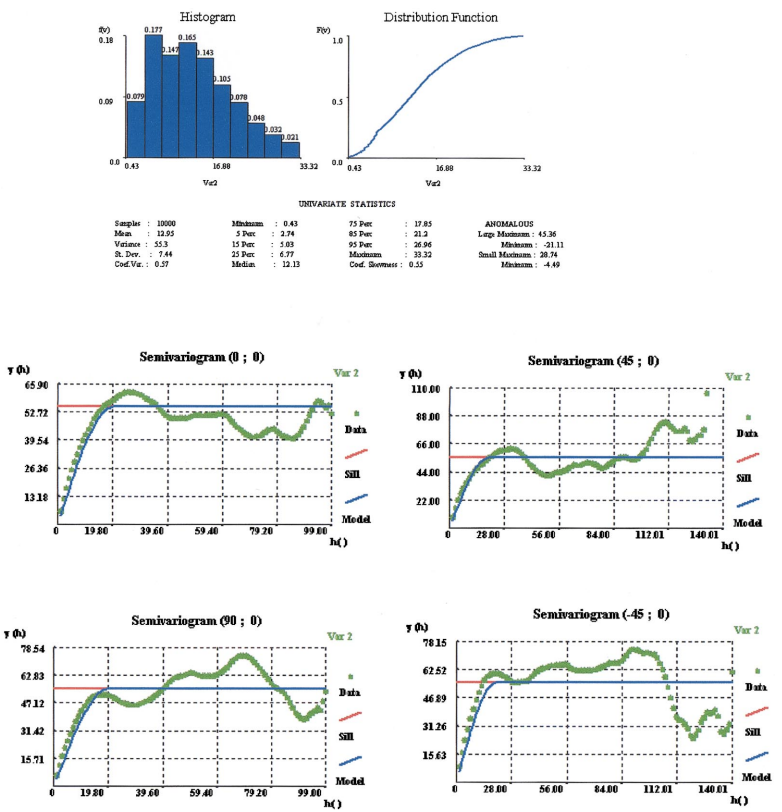


**Figure 6.** Histogram and variograms of simulated values of the primary variable.

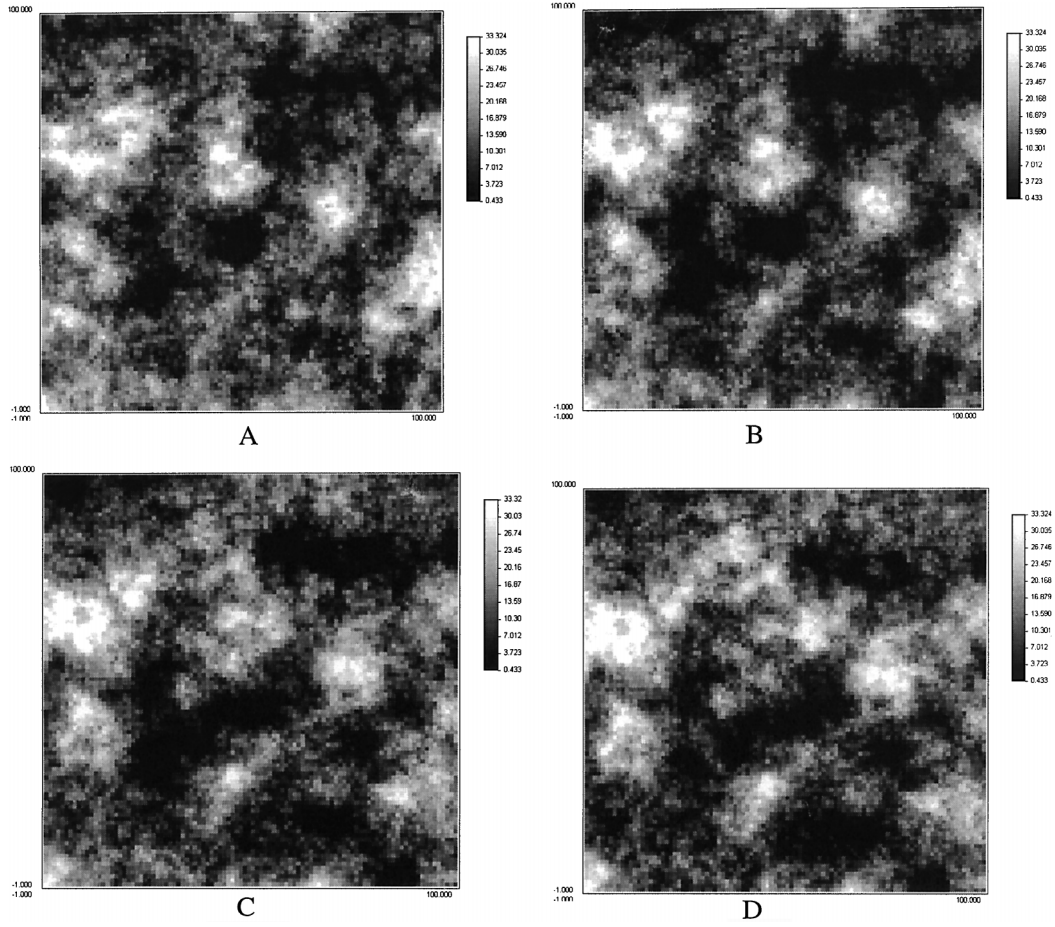


**Figure 7.** Results of the Direct Cosimulation: (A) primary variable; (B) secondary variable with $\rho = 0.9$; (C) secondary variable with $\rho = 0.7$; and (D) secondary variable with $\rho = 0.5$.

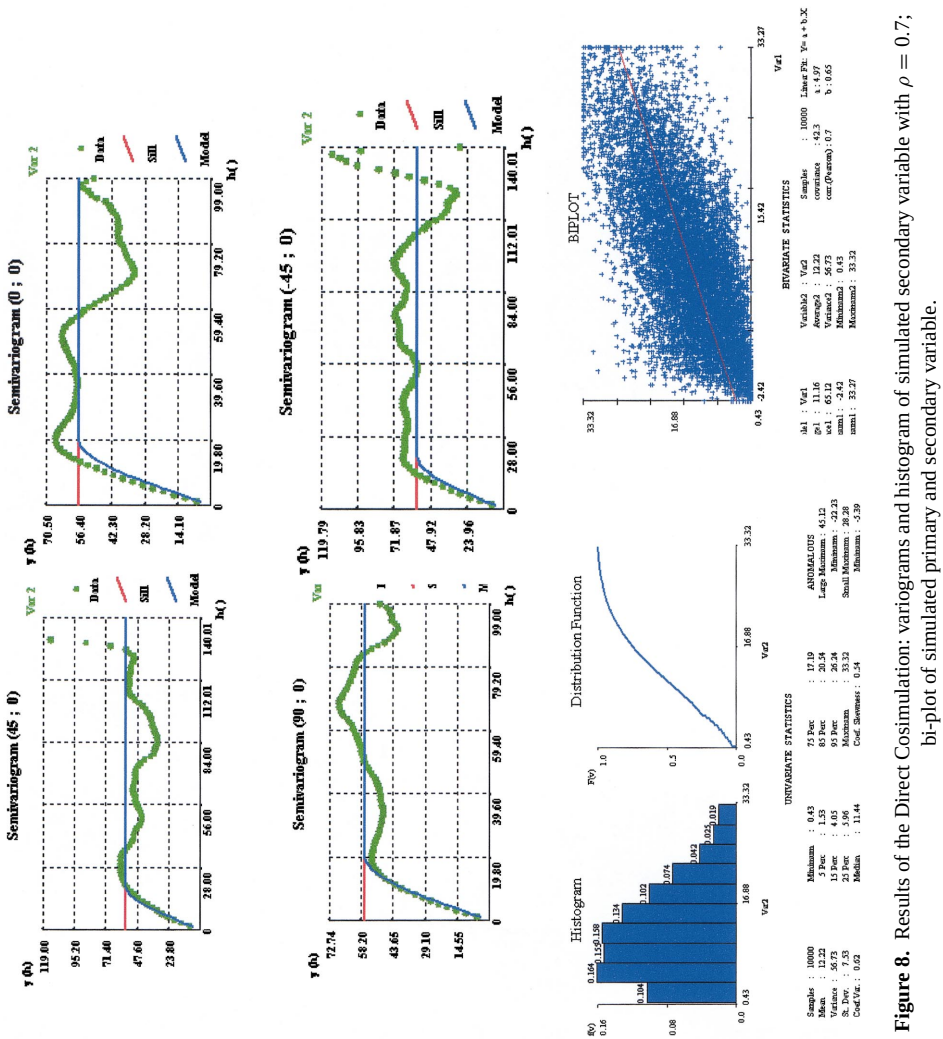


**Figure 8.** Results of the Direct Cosimulation: variograms and histogram of simulated secondary variable with $\rho = 0.7$; bi-plot of simulated primary and secondary variable.

## FINAL REMARKS

- The presented approach of direct sequential simulation uses the original variables, without any prior or posterior transformation. There is no more advantage in using the multiGaussian approach for the sequential simulation (SGS): the simplicity of implementation is identical to the proposed DSS without the disadvantages of the multiGaussian hypothesis. In those situations where spatial patterns of extreme classes need to be reproduced traditional SIS can still be used.
- Because the proposed algorithm deals with original variables, corrections for local means can be implemented easily in order to reproduce non-stationary patterns of the primary variable or to better reproduce the extreme classes of highly skewed histograms (see Appendix).
- The joint simulation of a set of $N_v$ variables is certainly the main advantage of the proposed method.

## ACKNOWLEDGMENTS

This research was supported by grants from Companhia de Geologia e Geofísica de Portugal (CGGP). I thank Jeff Caers for the discussion and review of a preliminary note on this subject. I acknowledge the constructive reviews from Sebastien Strebelle and one anonymous reviewer, which helped to improve the manuscript.

## REFERENCES

Almeida, A., and Journel, A., 1994, Joint simulation of multiple variables with a Markov-type coregionalization model: Math. Geology, v. 26, no. 5, p. 565–588.

Bourgault, G., 1997, Using non-Gaussian distributions in geostatistical simulations: Math. Geology, v. 29, no. 3, p. 315–334.

Caers, J., 1999, Adding local accuracy to direct sequential simulation: Stanford Center for Reservoir Forecasting, Annual Meeting 12, v. 2.

Caers, J., 2000, Direct sequential indicator simulation: Proceedings of 6th International Geostatistics Congress, Cape Town, South Africa.

Gomez-Hernandez, J., and Journel, A. G., 1993, Joint sequential simulation of multigaussian fields, *in* Soares, A., ed., Geostatistics TROIA'92: Kluwer Academic Pub., Dordrecht, The Netherlands, p. 85–94.

Goovaerts, P., 1997, Geostatistics for natural resources characterization: Oxford University Press, New York, 483 p.

Journel, A. G., 1994, Modeling uncertainty: some conceptual thoughts, *in* Dimitrakopoulos, R., ed., Geostatistics for the Next Century: Kluwer Academic Pub., Dordrecht, The Netherlands, p. 30–43.

Luís, J., and Almeida, J., 2000, Stochastic simulation of a multi-phase set of fluvial sand channels by using a geostatistical pixel-based algorithm: Proceedings of 6th International Geostatistics Congress, Cape Town, South Africa.

Ripley, B., 1987, Stochastic simulation: Wiley, New York, 237 p.

Soares, A., 1998, Sequential indicator simulation with correction for local probabilities: Math. Geology, v. 30, no. 6, p. 761–765.

## APPENDIX

### Correction for Local Bias

In those situations when the histogram of $z(x)$, used for normal score transform (2), has little data, mostly in low frequency classes, the following bias can occur: $\mathrm{E}\{z^s(x_u)\} \neq z(x_u)^*$.

In other words, if in each sequential step we calculate the following estimator of $\mathrm{E}\{Z^s(x_u)\}$: $z^{s^*}(x_u) = (1/N)\sum_{i=1}^{N} z_i^s(x_u)$ by drawing, for example, $N$ values of $z_i^s(x_u)$, using a Monte-Carlo simulation, this local bias can occur: $z^{s^*}(x_u) \neq z(x_u)^*$.

Hence, in these cases the following correction for the bias is proposed:

$$z^{s'}(x_u) = z^s(x_u) + [z^*(x_u) - z^{s^*}(x_u)]$$

where the simulated value $z^s(x_u)$, for node $x_u$, is corrected by the deviation $z^*(x_u) - z^{s^*}(x_u)$.

This correction for local bias allows a much better reproduction of those $z(x)$ histograms with few data classes.

*Important note*. The number $N$ of simulated values $z_i^s(x_u)$ (drawn by Monte-Carlo simulation) that we need to calculate $z^{s^*}(x_u)$ depends on the range of the histogram to be sampled, i.e., depends on the magnitude of the estimation variance $\sigma_{sk}^2(x_u)$. The following interative procedure to estimate $z^{s^*}(x_u)$ is proposed: we start drawing an increasing number of $z_i^s(x_u)$ values ($i = 1, 2, \ldots$), until the estimator $z^{s^*}(x_u)$ remains practically unchanged. In the examples of Figures 4 and 5, the maximum number of simulations used to estimate $z^{s^*}(x_u)$ in each sequential step was $N = 50$. The fact, that in each sequential step we have drawn up to 50 values of $z_i^s(x_u)$ did not have any significant impact on the computer time consumed.

### Correction for Local Means

The proposed DSS algorithm deals with original variables, which allows two important effects through the correction of local estimates (1) and (5):

i) Better reproduction of extreme classes of highly skewed histograms. With highly skewed distributions, direct sequential simulation sometimes does not succeed in reproducing the extreme classes proportion. This effect, common to all sequential simulation algorithms (SIS or SGS), is a consequence of the way the random path visits the nodes, close to extreme values, at the beginning of the sequential procedure (see Soares, 1998).

This can be an important issue if the proportions of extreme classes are a major concern of the case study. Soares (1998) proposed an efficient algorithm for correction of local probabilities, which entails the reproduction of the proportions of a categorical variable in the final simulated maps. Following the same idea, a correction of a local parameter—mean, proportion of extreme classes of the histogram—can be performed for better reproduction of the extreme classes. In highly skewed distributions the mean is a sensitive indicator of how the extreme classes are reproduced in simulated maps.

ii) Reproduction of nonstationary behavior of the mean, for example local areas with different means of $Z(x)$. Luís J. and Almeida J. (2000) applied the same correction for local probabilities to reproduce the local trends of a categorical variable (proportions of different types of sand channels in an oil reservoir). Within the same idea nonstationary patterns, characterized by local means of $Z(x)$, can be reproduced in final simulated maps by correcting the local estimates (1) and (5) to the targets that are the local means.

Suppose $m_z$ is the mean of original variable $Z(x)$ that we wish to be reproduced in the final simulated maps.

A deviation $e^s$ can be calculated between $m_z$ and the mean of simulated values $z^s(x)$ at a given step $s$ of the sequential procedure:

$$e^s = m_z - m_z^s$$

where $m_z^s$ is the mean of all simulated values at step $s$:

$$m_z^s = \frac{1}{N_s} \sum_{i=1}^{N_s} z^s(x_i)$$

$N_s$ is the number of simulated nodes up to step $s$.

Hence the estimated local means (1) or (5), in the cosimulation, can be corrected with the deviation $e^s$:

$$z(x_u)_s^* = z(x_u)^* + e^s$$

The remaining steps of direct sequential simulation or cosimulation are unchanged.