SEQUENTIAL UPDATING OF LOCAL DISTRIBUTION FUNCTIONS: A NEW IMPLEMENTATION OF THE SEQUENTIAL SIMULATION PRINCIPLE

Roland FROIDEVAUX
FSS Consultants SA
Geneva, Switzerland

Abstract

This paper presents a new implementation of the sequential simulation principle, within a multiGaussian framework. In this approach, the local conditional distribution functions, from which simulated values are drawn by Monte-Carlo, are updated iteratively rather than re-estimated at each step.

This new implementation offers several significant advantages:

- The local distribution functions, from which simulated values are drawn, are conditional to all hard and previously simulated data, rather than to data within a search neighbourhood only.
- There is no need to assign existing hard data to the nearest grid nodes. The local means and variances are estimated from the available data at their exact locations.
- The updating process does not involve any longer the solving of a linear system of equations. This, in turns, relaxes the constrain of using only spatial correlation models which verify the positive definite condition. In particular it allows to use directly an experimental variogram surface.

The performances of this new approach is illustrated by four examples using well data and a seismic attribute map.

1. Introduction

Sequential simulation is a wide class of simulation algorithms, all based on a recursive implementation of the Bayes axiom whereby the modelling of the multivariate distribution function, which fully describes a random function Z at any location u, is replaced by the product of a set of univariate conditional cdfs:

\[
F(u_1, \ldots, u_n; z_1, \ldots, z_N | (N_0)) = F(u_N; z_N | (N_0 + N - 1)) \cdot F(u_{N-1}; z_{N-1} | (N_0 + N - 2)) \cdot \ldots \\
\cdot F(u_1; z_1 | (N_0))
\]

with
and \( N_0 \) denotes the number of original data values.


Sequential Gaussian simulation is an implementation of the sequential simulation paradigm under the multiGaussian random function model which is used to simulated continuous variables. In its traditional implementation, sequential Gaussian simulation proceeds as follows (see Gooverts, 1977, p.380).

1. The set of data values \( \{ z_1, \ldots, z_N \} \) is transformed into a corresponding set of normal scores \( \{ g_1, \ldots, g_N \} \) using an appropriate transform \( G(u) = \Phi(Z(u)) \) and a multiGaussian hypothesis is assumed

2. The normal scores values \( \{ g_1, \ldots, g_N \} \) are assigned to nearest node of the grid to be simulated

3. A random path visiting each node of the grid is defined

4. At each grid node the local mean and variance of the local Gaussian ccdf is estimated by simple kriging. A simulated value \( g(u) \) is drawn from this local ccdf and treated added to the data set

5. Once all the nodes have been visited the simulated normal scores are back-transformed into simulated values of the original variable using the inverse of the Gaussian transform used to calculate the original normal scores: \( Z(u) = \Phi^{-1}(G(u)) \)

The new approach proposed in this paper differs from this classical implementation in that the local ccdfs are not estimated at each grid node before drawing a simulated value. Rather, the local ccdfs are initialized before performing any conditioning or simulation and then are updated sequentially after each drawing of a simulated value.

2. The Sequential Updating Approach

Consider the N grid nodes \( u_i, i = 1, \ldots, N \) discretizing the domain \( D \) to be simulated and denote by \( k \) the iteration index for visiting all the nodes.

At the initial stage \( (k = 0) \), prior to any conditioning or simulation, the parameters (mean and variance) of the local multiGaussian cdf \( G(u) \) are all equal, whatever the location \( u \):
Once a first value \((k = 1)\) \(g(u_\alpha)\) has been drawn at location \(u_\alpha\), it can be shown that the mean and variance of the local ccdfs \(G(u|1)\), conditional to this initial value, are equal to:

\[
m_1(u) = \rho(u - u_\alpha) \cdot g(u_\alpha) \tag{1}
\]

\[
\sigma_1^2(u) = 1 - \rho^2(u - u_\alpha) \tag{2}
\]

With \(\rho(u - u_\alpha)\) being the spatial correlation coefficient between location \(u\) and \(u_\alpha\).

More generally, at any iteration \(k\) and after having drawn a simulated value \(g(u_\alpha)\), the mean and variance of the local ccdfs \(G(u|k)\), conditional to this new value, are estimated by:

\[
m_k(u) = m_{k-1}(u) + \rho(u - u_\alpha) \cdot \frac{\sigma_{k-1}(u)}{\sigma_{k-1}(u_\alpha)} \cdot (g(u_\alpha) - m_{k-1}(u_\alpha)) \tag{3}
\]

and

\[
\sigma_k^2(u) = \sigma_{k-1}^2(u) \cdot (1 - \rho^2(u - u_\alpha)) \tag{4}
\]

Note that if \(k = 1\) (first iteration) we have:

\[
m_{k-1}(u) = m_{k-1}(u_\alpha) = 0 \quad \text{and} \quad \sigma_{k-1}^2(u) = \sigma_{k-1}^2(u_\alpha) = 1
\]

and equations (3) and (4) identify equations (1) and (2).

Thus, the key idea of sequential updating simulation is to visit randomly all grid nodes, to draw by Monte-Carlo a simulated value at each location, and to condition all local means and variances to this newly simulated value before moving to the next location.

Because of the iterative way in which the local ccdfs are updated, the equations (3) and (4) allow to generate a correlated gaussian field only if the spatial correlation function (variogram) is defined by a single structure with no nugget effect.
In order to reproduce a multi-structure variogram model, the local cdf \( G(\mathbf{u}) \) needs to be interpreted as a linear combination of \( Ns+1 \) (structure 0 is the nugget effect) independent random functions \( G^l(\mathbf{u}) \) with parameters:

\[
m_o^l(\mathbf{u}) = 0; \quad \forall l; \quad \forall \mathbf{u}
\]

\[
\sigma_o^{l2}(\mathbf{u}) = C'(0) = \text{increment of the } l_{th} \text{ structure of the variogram.}
\]

Thus, at each location \( \mathbf{u}_\alpha \), a set of simulated values \( \{g^l(\mathbf{u}_\alpha), l = 0, \ldots, Ns\} \) is drawn from the corresponding set of ccdf and recombined into a single simulated value:

\[
g(\mathbf{u}_\alpha) = \sum_{l=0}^{Ns} g^l(\mathbf{u}_\alpha)
\]

Then the local means and variances over the entire grid are updated as follows:

\[
m_k^l(\mathbf{u}) = m_{k-1}^l(\mathbf{u}) + \rho(\mathbf{u} - \mathbf{u}_\alpha) \cdot \frac{\sigma_{k-1}^l(\mathbf{u})}{\sigma_{k-1}^l(\mathbf{u}_\alpha)} \cdot (g^l(\mathbf{u}_\alpha) - m_{k-1}^l(\mathbf{u}_\alpha))
\]

(5)

\[
\sigma_k^{l2}(\mathbf{u}) = \sigma_{k-1}^{l2}(\mathbf{u}) \cdot (1 - \rho^2(\mathbf{u} - \mathbf{u}_\alpha))
\]

(6)

In practice, the updating of the local cccdfs is performed only within correlation distance from location \( \mathbf{u}_\alpha \).

Remarks

1- One of the fundamental difference between the Sequential Updating approach and the classical Sequential Gaussian simulation is the fact that there no kriging system to solve. This eliminates the numerical problems sometimes found when solving a linear system of equations and relaxes the constraints imposed on the choice of variogram models. Indeed since there are no kriging system to solve, the need to respect the positive definite condition is not any longer mandatory.

2- In the Sequential Updating approach, the locals cccdfs, from which simulated values are drawn, are conditioned to all hard data and previously simulated values: there are no neighbourhood search nor a maximum number of data to be considered.

3- Equations (3), (4), (5) and (6) are only approximations. Indeed at each iteration only the local means and variances are updated. Not the correlogram. This means that instead of using a non-stationary posterior correlogram we use the prior, stationary one. However, considering the results
which are presented further down, it seems that this departure from exactitude does not have any significant negative impact.

3. Conditioning to data

In the Sequential Updating approach, conditioning to existing hard data is achieved as an initial updating of the local cdfs $G(u)$:

- The data set $\{z(u_j), j = 1, \ldots, n\}$ is first transformed into a corresponding set of normal scores $\{g(u_j), j = 1, \ldots, n\}$

- Then, each normal score value $g(u_j)$ is split into its structural components:
  $$g'(u_j) = g(u_j) \cdot C'(0)$$

- Finally, each set of values $\{g'(u_i), i = 0, \ldots, N3\}$ is used, successively, to update the a priori parameters $m_0(u)$ and $\sigma_0^2(u)$ at each location $u$ using equations (5) and (6).

Remarks

1. Unlike the traditional implementation of sequential gaussian simulation, the conditioning data are not re-assigned to nearest grid node: the exact location of each datum is used for performing the updating.

2. The order in which the $n$ conditioning data are used for updating varies randomly from one realization to the next.

4. Accounting for collocated information

Collocated information classically denotes a secondary attribute, statistically correlated with the attribute of main interest, and available at each grid node. This information is typically used to estimate the a priori mean, and variance, of the main attribute. The collocated cokriging approach (Almeida, 1993, Xu et als, 1992) is one of the classical way to perform an estimation based on both the hard data on the main attribute and the secondary information.

In the Sequential Updating approach, this process is split in two distinct phases: first, the a priori local cdfs $G(u)$ are initialized using the collocated secondary information as follows:
• The co-located attribute \( z_2(u_i), i = 1, \ldots, N \) are transformed into their normal scores \( g_2(u_i), i = 1, \ldots, N \).

• Then, \( m_0'(u) \) and \( \sigma_0'^2(u) \), the parameters of the local a priori cdf \( G'(u) \), are initialized to:

\[
m_0'(u_i) = g_2(u_i) \cdot C'(0) \cdot \rho
\]

\[
\sigma_0'^2(u_i) = C'(0) \cdot \rho
\]

with

\[
\rho = \text{linear correlation coefficient between } G(u) \text{ and } G_2(u)
\]

Once this initialization is completed, the conditioning to hard data and the simulation itself proceed as described in the previous paragraphs.

5- Examples

Four examples will be used to illustrate how the Sequential Updating approach performs. In all cases, the attribute considered is the effective thickness of a gas reservoir (i.e. the thickness of the reservoir above the oil-gas contact) tested by three wells. A seismic attribute map (two-ways travel time) is also available and provides indirect information on the reservoir thickness. Figure 1 presents the location map of the three wells, with the measured thickness, as well as the two-ways travel time map. The outline of the shaded area on the location map represents the limit of the effective portion of the reservoir. Outside this area, the structure is completely under the gas-water contact. The a priori distribution model assumed for thickness is a Gaussian distribution with a mean of 15m and a standard deviation of 5m. This model implies that the thickness of the reservoir can become null and that the three wells are located in the thickest part of the reservoir.

In all four cases, 50 realizations of the reservoir effective thickness will be generated using various variogram models and accounting for the various types of available information.

In the first case we consider a non-conditional simulation of the thickness using an hypothetical, strongly anisotropic, variogram model (large range 1000m, short range 100m, direction 70°) and a 30 percent nugget effect.

In the second case we perform a conditional simulation of the thickness, taking into account the three well data only, and using a more realistic variogram model describing smooth spatial continuity (gaussian model with a large range of 1100m, a short range of 700m and a direction of 135°) without any nugget effect.

In the third case a conditional simulation is performed again, but this time accounting for both the hard
and soft information. The variogram model used is the same than in the second case. The statistical correlation coefficient between the effective thickness and the two-ways travel time is assumed to be -0.9, i.e indicative of a very strong negative relationship between the two attributes.

In the last case, we perform again a non-conditional simulation but this time the measure of spatial correlation is not deduced from a specified variogram model but is read directly from a variogram map calculated from an other seismic attribute.

Figures 2 to 5 present three realizations for each case, together with relevant statistics: maps of the local means and standard deviations (average of the 50 realizations), histogram and distribution function of the simulated values (for the 50 realizations) and average variogram calculated, again, on the 50 realizations.

For case 1 (Figure 2) we see that both the a priori cdf model and the a priori variogram model are almost perfectly reproduced and that the three realizations presented are realistic with respect to the prior model.

In case 2 (Figure 3) we see that the three conditioning data are correctly accounted for (the local standard deviation map tends toward zero at data locations) and that all the realizations exhibit the smoothness typical of this type of structural attribute. In terms of reproduction of the model statistics neither the a priori cdf nor the variogram variogram model are exactly reproduced. These discrepancies represent ergodic fluctuations (see Gooverts, p. 426) characteristic of all stochastic simulation methods. These fluctuations result from the number of realizations generated, the amount and position of conditioning data and from the ratio between the variogram ranges to the limits of the area to be simulated. In our case the discrepancy between the prior and posterior cdf can be explained by the fact that all measured thicknesses are well above the model average value and, moreover, are located near the centre of the reservoir. Given the strong and smooth spatial correlation model, all realizations will contain a large high valued area. Similarly, the joint impact of the position of high valued data near the centre, the parameters of the variogram model, and the geometry of the simulated area implies an 'hole-effect' behaviour which is not explicit in the a priori model but is quite apparent in the average posterior variogram.

The results of case 3 (Figure 4) show that both the hard data and the secondary information are correctly accounted for. Comparing these results to case # 2 we see that the realizations still have their high-valued area near the centre of the reservoir (impact of hard data). However, away from the centre, the realizations are much constrained in the current case than in the previous one (impact of the collocated information). Note also that the local standard deviations of the simulated thicknesses are strongly reduced with regard to the second case. This is again the result of the strong constraining effect of the secondary information.

The only purpose of case 4 (Figure 5) is to illustrate how the Sequential Updating approach allows the direct use of a variogram surface instead of the classical parametric model. In this case the correlation map, calculated on another seismic attribute, is used as a non-parametric model. As can be seen, the posterior correlation map matches reasonably well the prior one. This flexibility on how to specify
spatial correlation opens up interesting new avenues for simulation. In particular we do not have to worry any longer on the positive definite property of the model used. However, that does not mean that any variogram map is suited to task: one has to be careful that it will not induce unwanted artefacts. A potential problem is the presence of high (positive or negative) correlation values near the edge of the correlation map. This tends to generate apparent discontinuities, or strong gradients, on the realizations as can be seen in some areas of the realizations presented here. One simple way to reduce, if not eliminate, this problem is to ensure that the variogram map is not calculated for distances greater than 50 percent of the dimensions of the simulation area and that it does not contain correlation values calculated on too few data pairs.

6- Discussion

The Sequential Updating approach, proposed in this paper, offers an alternative to the classical implementation of the sequential gaussian simulation with the following attractive features:

1- By construction, all local ccdfs are fully conditioned to all simulated values. This is not the case in the classical implementation since the local ccdfs are estimated on the basis of a limited number of data within a search ellipse.

2- It does not require the solving of any system of linear. As a result Sequential Updating is generally faster than the classical implementation and is not prone to the sometimes annoying numerical problems found in kriging.

3- Since no kriging system needs to be solved, spatial continuity for the attribute to be simulated can be specified with less restrictions than before. In particular, the positive definite requirement for a variogram model is not any longer required. An other interesting possibility offered by this approach is that the spatial continuity can be specified on the form of a variogram map rather than a parametric model.

4- The conditional hard data are not assigned to the nearest grid node, but are used at their exact locations. Although this data re-allocation has never been a serious concern in 2D, it has represented a problem in 3D if the vertical grid node spacing is larger than the vertical data spacing. In this situation several data may share the same grid node and a decision must be taken on which one takes precedence.

An additional attractive feature, although not discussed in the present paper, is the possibility to perform factorial simulation in a way similar to factorial kriging.
References


Figure 1: Data location map and travel time map

Histogram and cumulative distribution function (50 realizations)

Average variograms (50 realizations)
2: Non conditional simulation

Results

Histogram and cumulative distribution function (50 realizations)
Figure 3: Conditional simulation results (hard data only)

Average variograms (50 realizations)

\(\theta = 45^\circ\)

\(\theta = 135^\circ\)
Real # 1

Real # 2

Real # 3

Histogram and cumulative distribution function (50 realizations)

Correlation Map (Model)

Figure 4: Conditional simulation results (hard data and collocated information)

Average variograms (50 realizations)

Conditional results (hard collocated)
Figure 5: Non conditional variogram model simulation with non-parametric

Average Correlation Map (50 realizations)