NON-STATIONARY MULTIPLE-POINT GEOSTATISTICAL MODELS

SEBASTIEN STREBELLE[1] and TUANFENG ZHANG[2]
[1] *ChevronTexaco Energy Technology Company,*
6001 Bollinger Canyon Road, San Ramon, CA 94583, USA
[2] *Department of Geological and Environmental Sciences*
Stanford University, Stanford, CA 94305, USA

Abstract. During the last few years, the use of multiple-point statistics simulation to model depositional facies has become increasingly popular in the oil industry. In contrast to conventional variogram-based techniques such as sequential indicator simulation, multiple-point geostatistics enables the generation of facies models that capture key depositional elements (e.g. curvilinear channels) characterized by unique and predictable shapes. In addition, multiple-point geostatistics is more intuitive because the complex mathematical expression of the variogram is replaced with an explicit three-dimensional training image that depicts the geometrical characteristics of the expected facies.

In multiple-point geostatistics, the stationarity assumption that underlies the inference of a variogram model from sparse sample data is extended to infer facies joint-correlation statistics from the training image. A consequence of this assumption is that patterns extracted from the training image can be reproduced in any region of the reservoir model where the training image is thought to be representative of the geological heterogeneity. Yet actual reservoirs are generally non-stationary: topographic constraints, sea-level cycles, or changes of sedimentation sources lead to spatial variations of facies deposition directions and facies geobody dimensions. Three-dimensional fields of location-dependent facies azimuth/dimensions representing those spatial variations are commonly estimated from well log and seismic data, or from geological interpretations based on analogs. This paper proposes a modification of the multiple-point statistics simulation program **snesim** to account for such non-stationary information.

In the original **snesim**, prior to the simulation, the multiple-point-point statistics inferred from the training image are stored in a dynamic data structure called a search tree. In the presence of a locally-varying azimuth field, the range of possible azimuths over the study field is first discretized into a small number of classes. Then, the training image is successively rotated by the average value of each azimuth class and a search tree is built for each resulting rotated training image. During the simulation, at each unsampled node, multiple-point statistics are retrieved from the search tree built for the class in which the local azimuth falls, enabling the local reproduction of patterns similar to those of the corresponding rotated training image. A similar process is proposed to account for a field of location-dependent facies geobody dimensions. The new modified **snesim** program is applied to the simulation of a fluvial reservoir with locally-variable channel orientations and widths.

O. Leuangthong and C. V. Deutsch (eds.), Geostatistics Banff 2004, 235-244.

1 Introduction

Multiple-point geostatistics has emerged recently as a practical approach to characterize and model facies at reservoir scale (Strebelle *et al*, 2002). The first step of this approach is the construction of a three-dimensional training image describing the facies thought to be present in the study area. The training image captures the geometrical characteristics of each facies, as well as the complex spatial relationships among multiple facies. The training image is a purely-conceptual geological model; it contains no absolute location information and in particular, is not conditioned to any actual field data. In reservoir modeling applications, non-conditional object-based modeling techniques appear to be well-suited to create such three-dimensional conceptual models. The second step of this approach consists of inferring from the training image statistics on the joint-correlation of facies at multiple locations, and using these statistics to reproduce patterns similar to those of the training image while honoring hard and soft conditioning data.

The theoretical framework of multiple-point geostatistics was developed as early as 1989 by Journel and Alabert and was revisited by Guardiano and Srivastava in 1993. The first practical implementation was proposed by Strebelle (2000), who introduced a dynamic data structure called a search tree, to efficiently store and retrieve all multiple-point statistics inferred from the training image. During the last few years, multiple-point geostatistics has been shown to overcome the major limitations of traditional facies modeling technologies:

- Multiple-point statistics (MPS) simulation enables improved modeling of curvilinear and large-scale continuous facies patterns, such as sinuous channels, relative to variogram-based techniques (Strebelle *et al*, 2002). In addition, the training image is much easier to analyze/discuss than a variogram model.
- In contrast to object-based modeling techniques (Holden *et al*, 1996; Viseur, 1997; Lia *et al*, 1998), MPS simulation is a very flexible data integration tool. In particular, MPS models honor all conditioning well data, i.e. reproduce at all well data locations the facies connectivity/geometry observed in the training image, with no limitation on the number of wells (Strebelle and Journel, 2001).

One important assumption underlying the inference of multiple-point statistics from the training image and their reproduction in the MPS model is the stationarity of the field under study: facies relative proportions, geometries, and associations are expected to be reasonably homogeneous over the field. Yet, most actual reservoirs are not stationary. Local topographic constraints such as the presence of a salt dome, seal level cycles, or changes of sedimentation sources, lead to significant spatial variations of facies deposition directions and facies geobody dimensions.

In this paper, we first review the implications of the stationarity assumption in multiple-point geostatistics. Then we propose modifying the MPS simulation program **snesim** (Strebelle, 2000) to reproduce pre-defined non-stationary information such as locally-varying facies azimuth and/or facies geobody dimension data.

2 Stationarity

Geostatistics relies on the concept of Random Function. The Random Function represents the statistical model of spatial variability of some property over some study field. In traditional geostatistics, the Random Function model is generally limited to some one-point and two-point statistics moments, namely a cumulative distribution function and a variogram model. In multiple-point geostatistics, the Random Function model consists of the multiple-point facies joint-correlation moments that can be inferred from the training image. The inference of statistics representing the Random Function model requires some repetitive sampling. For example, a porosity cumulative probability distribution is typically inferred from the histogram of porosity data collected from all well logs available over the study field. However, when pooling sample data together into a single histogram, the modeler makes an assumption of stationarity: all porosity sample values are assumed to originate from the same unique population, regardless of their location in the reservoir. Another stationarity decision is commonly taken whenever a variogram is computed by pooling information at similar lag distances together into a single scatter plot.

In multiple-point geostatistics, the stationarity assumption carries over to higher order statistics: multiple-point statistics moments are inferred from training patterns present in the training patterns regardless of the location of these patterns in the training image. As a consequence, non-stationary features of the training image cannot be preserved in MPS models. Figure 1 shows a clearly non-stationary training image wherein ellipses are South West-North East-oriented in the left half of the image, and North West-South East-oriented in the right half. The resulting model generated by the MPS simulation program **snesim** displays a mix of ellipses oriented in both directions over the whole field.

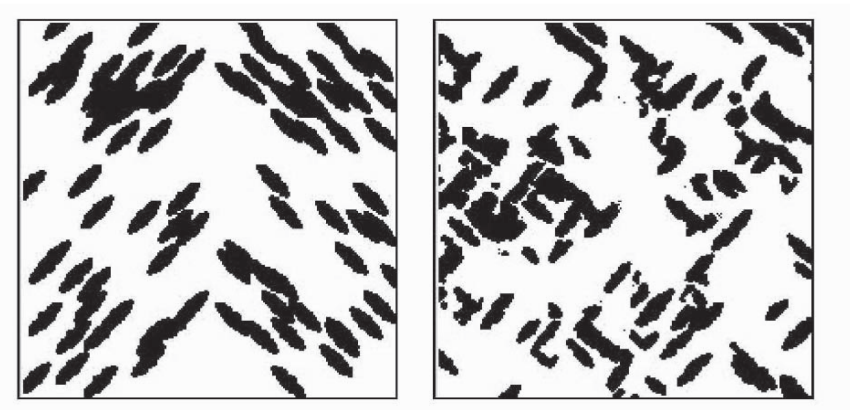

Figure 1. Non-stationary training image (left), and resulting MPS model (right). The specific locations of the South West-North East and North West-South East-oriented ellipses in the training image are not preserved in the MPS model.

The non-stationary features of the training image are not captured in MPS models. Therefore, we propose using a stationary training image and applying rotation and affinity transforms to the training image to reproduce non-stationary features to MPS models. Prior to that, the implementation of the original MPS simulation program **snesim** is briefly recalled.

3 Multiple-point statistics simulation implementation

The MPS simulation program **snesim** proposed by Strebelle (2000) is a pixel-based direct sequential simulation algorithm: all simulation grid nodes are visited only once along a random path and simulated node values become conditioning data for cells visited later in the sequence. Let S be the categorical variable (depositional facies) to be simulated, and s_k, k=1…K, the K different states (facies types) that the variable S can take. At each unsampled node $\mathbf{u}$, d_n denotes the data event consisting of the n conditioning data $S(\mathbf{u_1})=s(\mathbf{u_1})\dots S(\mathbf{u_n})=s(\mathbf{u_n})$, closest to $\mathbf{u}$. The conditional probability distribution function (cpdf) at $\mathbf{u}$ is inferred by scanning the training image to find all training replicates of d_n (same geometric configuration and same data values as d_n), and identifying the conditional facies probabilities as the facies proportions obtained from the central values of the training d_n –replicates.

Instead of repeatedly scanning the whole training image at each unsampled node to search for training replicates of the local conditioning data event, Strebelle (2000) proposed storing ahead of time all conditional facies probabilities that can be inferred from the training image in a dynamic data structure called a search tree. More precisely, given a conditioning data search window W, which may be a search ellipsoid defined using GSLIB conventions (Deutsch and Journel, 1998), τ_N denotes the data template (geometric configuration) constituted by the N vectors $\{\mathbf{h}_\alpha, \alpha=1\dots N\}$ corresponding to the N relative grid node locations included within W. Prior to the simulation, the training image is scanned with τ_N, and the numbers of occurrences of all training data events associated with τ_N are stored in the search tree. During the simulation, at each unsampled node $\mathbf{u}$, τ_N is used to identify the conditioning data located in the search neighborhood W centered on $\mathbf{u}$. d_n denoting the data event consisting of the n conditioning data found in W (original sample data or previously simulated values, $n \le N$), the local probability distribution conditioned to d_n is retrieved directly from the above search tree; the training image need not be scanned anew.

Theoretically, a large data template τ_N should be used to capture the large-scale features of the training image. However, such large template would increase dramatically the memory used to build the search tree and the cpu-time needed to retrieve conditional probabilities from it. One practical solution to capture large-scale structures while keeping the size of the data template τ_N reasonably small ($N \le 100$) is to use a multiple grid simulation approach (Strebelle, 2000). In **snesim**, this approach consists of simulating a series of G increasingly-finer grids, the g-th ($1 \le g \le G$) grid comprising each 2^{G-g}-th node of the final (finest) simulation grid. After the data template $\tau_N=\{\mathbf{h}_\alpha, \alpha=1\dots N\}$ has been defined on the finest grid, its components $\mathbf{h}_\alpha$ are rescaled proportionally to the node spacing within the grid being simulated. Thus the rescaled data template $\tau_N^g=\{\mathbf{h}_\alpha^g=2^{G-g}.\mathbf{h}_\alpha, \alpha=1\dots N\}$ is used to build the search tree and search for conditioning data when simulating the g-th grid.

In the next two sections, we show how rotation and affinity transformations can be applied to the data template τ_N prior to building the search tree, to integrate location-dependent azimuth and geobody size information into MPS models.

4 Integration of azimuth data

In this section, we first study the simple case in which the main direction of continuity of the facies geobodies is assumed to be constant over the field under study, but possibly different from the main direction of continuity of the training facies. Then we extend this technique to handle 2D or 3D fields of location-dependent azimuths. Only azimuths defined in the *xy*-plane are considered in this section because, in practice, dip is typically taken into account by the layering of the stratigraphic grid in which the facies model is built.

4.1 CONSTANT AZIMUTH

Consider the case in which the facies geobodies in the MPS model should have a constant principal direction of continuity, yet possibly different from that of the training image. Let θ be the difference in degrees counter-clockwise between those two directions.

Given a training image and a data template $\tau_N=\{\mathbf{h}_\alpha, \alpha=1\ldots N\}$, Zhang (2002) proposed modifying the **snesim** algorithm as follows. First the search tree is built from the training image using τ_N. Then, a new data template $\tau_N(\theta)$ is created from τ_N by the following method:

1. Rotate by θ each single component $\mathbf{h}_\alpha$ of τ_N.
   In 2D, the coordinates $(x_\alpha(\theta),\ y_\alpha(\theta))$ of the rotated component $\mathbf{h}_\alpha(\theta)$ are computed from the coordinates (x_α, y_α) of the original component $\mathbf{h}_\alpha$ as:
   $x_\alpha(\theta)= x_\alpha \cos\theta + y_\alpha \sin\theta$
   $y_\alpha(\theta)= - x_\alpha \sin\theta + y_\alpha \cos\theta$
2. Relocate the rotated components $\mathbf{h}_\alpha(\theta)$ to the nearest nodes of the simulation grid currently simulated.

During the simulation, at each unsampled node, the rotated data template $\tau_N(\theta)$ is used to search for nearby conditioning data, and the corresponding conditional probability distribution function (cpdf) is retrieved from the search tree, which was built using the original data template τ_N.

However, as described in the previous section, **snesim** uses a multiple-grid simulation approach that consists of simulating a series of increasingly-finer grids. Thus, at the early stage of the simulation, the rotated components $\mathbf{h}_\alpha(\theta)$ are relocated to the closest nodes of some coarse grids, entailing drastic approximations regarding the actual locations of the conditioning data. Such approximations lead to the inaccurate estimation of facies probability distributions and the poor reproduction of training patterns. However, because the simulation grids used in **snesim** are regular Cartesian grids and the distance between nodes is the same along both x and y-directions, the components $\mathbf{h}_\alpha(\theta)$ of the rotated data template match exactly existing grid nodes for θ=0, 90, 180, or 270 degrees. This is a property that we will use in the next sub-section.

An alternative approach consists of keeping the original data template τ_N to search for conditioning data, but rotating that data template to build the search tree prior to the simulation. In this case, the rotated data template is built in a slightly different way than in Zhang's method:

1. Rotate by $-\theta$ each single component $\mathbf{h}_\alpha$ of the original data template τ_N.
   In 2D, the coordinates $(x_\alpha(-\theta),\ y_\alpha(-\theta))$ of the rotated component $\mathbf{h}_\alpha(-\theta)$ are computed from the coordinates (x_α, y_α) of the original component $\mathbf{h}_\alpha$ as:
   $x_\alpha(-\theta) = x_\alpha \cos\theta - y_\alpha \sin\theta$ and : $y_\alpha(-\theta) = x_\alpha \sin\theta + y_\alpha \cos\theta$
2. Relocate the rotated components $\mathbf{h}_\alpha(-\theta)$ to the nearest nodes of the training image grid. When using **snesim**, the training image is assumed to have the same node spacing as the (finest) simulation grid.

The resulting rotated data template $\tau_N(-\theta)$ is used to build the search tree from the training image. The exact same result can be obtained by rotating the training image by θ, then building the search tree from that rotated training image using the original data template τ_N. During the simulation, at each unsampled node, τ_N is used to search for nearby conditioning data, and the corresponding cpdf is retrieved from the above search tree.

The most critical advantage of that new technique over Zhang's original method is that the relocation of the rotated components $\mathbf{h}_\alpha(-\theta)$ to the training image grid entails only minor approximations of the actual locations of the conditioning data, thus resulting in a reasonably good reproduction of the training patterns. As an application, this modified **snesim** program was used to model a horizontal 2D section of a fluvial reservoir. The training image depicts the prior conceptual geometry of the sinuous sand channels expected to be present in the subsurface (Figure 2). The size of that image is 250*250=62,500 pixels, and the channel proportion is 27.7%. A non-conditional simulated realization was generated using the same direction of continuity as that of the training image (Figure 2), then two additional models were created using different arbitrary main directions of continuity: 20 and 50 degrees (Figure 3). All models reproduce equally well the patterns displayed in the training image.

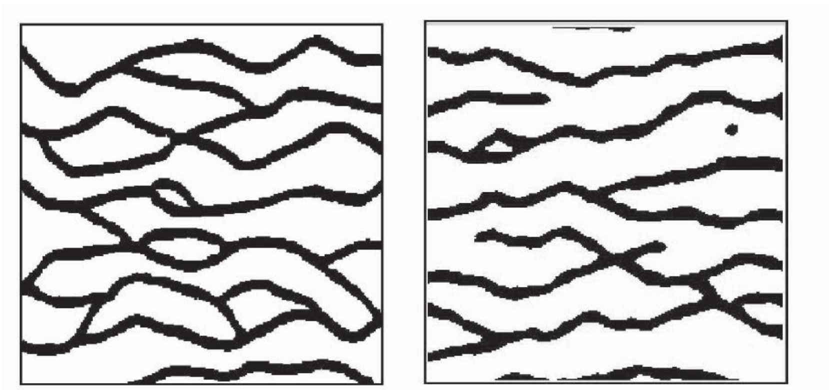

Figure 2. Training image used for the simulation of a 2D horizontal section of a fluvial reservoir (left), and reference MPS model (right).

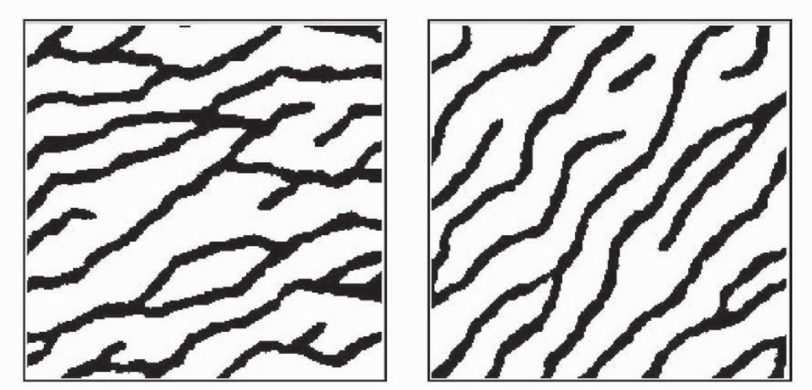

Figure 3. Fluvial reservoir MPS models obtained with a 20 degree counter-clockwise azimuth difference with the training image (left), and a 50 degree difference (right).

4.2 LOCATION-DEPENDENT AZIMUTHS

In reservoir facies modeling applications, it commonly is observed that the principal direction of continuity of the facies varies from one region of the reservoir to another. For example, topographic constraints, such as changes in the slope gradient, may lead to the formation of several sand fairways with different depositional directions. Data regarding such variations can be derived from different sources. In particular, local depositional directions can be obtained from geological interpretation (Harding *et al*, this meeting), or can be computed from seismic data (Strebelle *et al*, 2002).

Suppose that local azimuths can be estimated at each location **u** of the reservoir, and that $\theta(\mathbf{u})$ denotes the difference in degrees counter-clockwise between the azimuth value estimated at node **u** and the azimuth of the (stationary) training image. Given a data template τ_N, the method previously presented for a constant azimuth field can be extended to the location-dependent azimuth field $\theta(\mathbf{u})$ as follows:

1. Consider the range $[\theta_{min}, \theta_{max}]$ of all azimuth values estimated over the entire study field. Discretize that range into a small number L of classes, using regularly-spaced threshold values: $\theta_i = \theta_{min} + i^*(\theta_{max} - \theta_{min})/L$, i=0… L.
2. Using the method described in the previous sub-section, compute for each class $[\theta_i; \theta_{i+1}]$ the search tree corresponding to the rotated data template $\tau_N(-\theta)$ where θ is the central value of the class: $\theta = (\theta_i + \theta_{i+1})/2$
3. During the simulation, at each node **u** to be simulated, use the original data template τ_N to search for nearby conditioning data, and retrieve the local cpdf from the search tree corresponding to the class of azimuth angles to which $\theta(\mathbf{u})$ belongs.

If the range $[\theta_{min}, \theta_{max}]$ of azimuth angles is greater than 90 degrees, Zhang's original technique can be used to decrease the range of the individual discretized classes $[\theta_i, \theta_{i+1}]$. For example, consider the simulation of node **u** where $\theta(\mathbf{u}) = \theta_{min} + 100°$. The rotated data template $\tau_N(90°)$ can be used to search for conditioning data (recall that 0, 90, 180, and 270 degrees are the only rotation angles for which Zhang's method requires no data relocation). Then the resulting cpdf can be retrieved from the search tree corresponding to the class of azimuth angles to which (θ_{min}+100°)-90°= θ_{min}+10° belongs. Therefore, in any case, the maximum range of azimuth values to discretize is 90 degrees. The number L of azimuth classes should depend then on the uncertainty about the local azimuth values. With L=5 classes, the range of each class is 18 degrees. This is equivalent to estimating local azimuth values with an error of ± 9 degrees.

One limitation of the above technique may be the memory demand because one search tree per azimuth class needs to be built. However, one can consider one azimuth class after the other, i.e. build the search tree corresponding to a given azimuth class, simulate all grid nodes corresponding to that class, then delete that search tree prior to considering the next azimuth class. Building, then deleting search trees is a relatively fast process compared to the actual grid simulation process.

Figure 4 shows a 2D azimuth field and a resulting simulated realization using the fluvial reservoir training image of Figure 2. The reproduction of the training patterns is similar to that in the reference simulated realization of Figure 2. Note also that, although only five azimuth classes were used, the discretization of the range of possible azimuths did not create any artifact in the simulated realization.

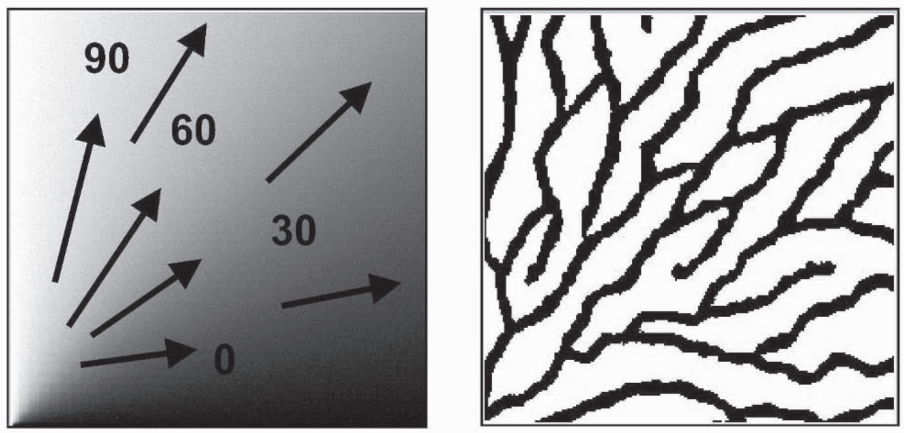

Figure 4. 2D location-dependent azimuth field (left), and resulting MPS model obtained using the fluvial reservoir training image of Figure 2 (right).

5 Integration of geobody dimensions data

Facies geobody dimensions that may depend, for example, on the distance to the sedimentation source, represent another traditional non-stationary feature of hydrocarbon reservoirs. A technique similar to that presented in the previous section to impose locally-varying azimuths is proposed to integrate geobody dimensions data, using some affinity transform of the data template used to build the search tree. For the sake of simplicity, we assume in this section that an isotropic rescaling factor (same affinity ratio in x, y, and z-directions) is sufficient to describe the variations of geobody dimensions in the volume under study.

Consider first the case in which the facies geobodies should have constant dimensions over the study field, yet possibly different from the dimensions of the training geobodies. Let λ be the ratio between target and training facies dimensions. Given a training image, and a data template $\tau_N=\{\mathbf{h}_\alpha, \alpha=1\ldots N\}$, a new data template $\tau_N(1/\lambda)$ is obtained from τ_N by the following method:

1. Rescale by $1/\lambda$ each component $\mathbf{h}_\alpha$ of τ_N. In 2D, the coordinates ($x_\alpha(1/\lambda)$, $y_\alpha(1/\lambda)$) of the rescaled component $\mathbf{h}_\alpha(1/\lambda)$ are computed from the coordinates (x_α, y_α) of the original component $\mathbf{h}_\alpha$ as: $x_\alpha(1/\lambda)=x_\alpha/\lambda$ and: $y_\alpha(1/\lambda)=y_\alpha/\lambda$
2. Relocate these rescaled components to the nearest nodes of the training image grid.

The resulting rescaled data template $\tau_N(1/\lambda)$ is used to build the search tree from the training image. The exact same result can be obtained by rescaling the training image by λ, then building the search tree from that rescaled training image using the original data template τ_N. During the simulation, at each unsampled node, the original data template τ_N is used to search for nearby conditioning data, and the corresponding cpdf is retrieved from the above search tree.

The extension of that technique to integrate location-dependent geobody dimensions data is straightforward and similar to the integration of location-dependent azimuth data. If $\lambda(\mathbf{u})$ denotes the ratio between target and training facies dimensions at the grid node location $\mathbf{u}$, then MPS simulation using locally-varying geobody rescaling factors consists of dicretizing the range of $\lambda(\mathbf{u})$ values into a smaller number of classes, and building a search tree for the average rescaling factor value of each class.

Figure 5 shows a 2D rescaling factor field and a resulting simulated realization using the fluvial reservoir training image of Figure 2. The reproduction of the training patterns is similar to that in the reference simulated realization of Figure 2.

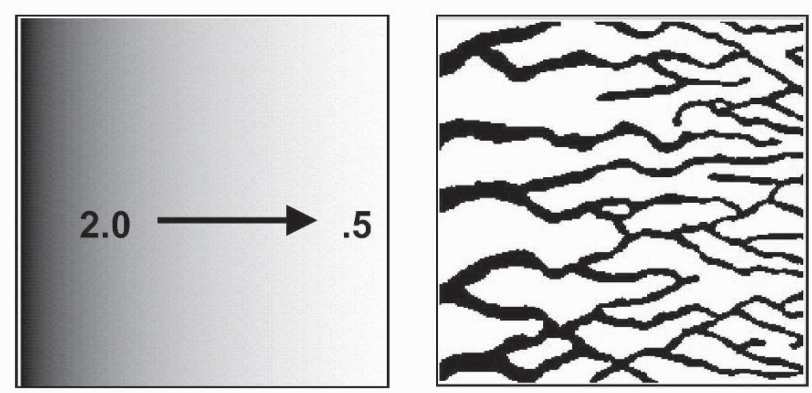

Figure 5. 2D field of location-dependent geobody dimension rescaling factors (left), and resulting MPS model obtained using the training image of Figure 2 (right).

6 Conclusion

In multiple-point geostatistics, statistics on facies joint-correlation at multiple locations are inferred from patterns displayed by a training image regardless of the location of these patterns in the training image. As a consequence, non-stationary features, such as spatial variations of facies azimuths or geobody dimensions that the training image may contain are not preserved in the multiple-point statistics simulated realizations.

To integrate variable azimuth/dimensions data, we propose applying a series of rotation/affinity transforms to a stationary training image, and building a search tree to store the multiple-point statistics inferred from each rotated/rescaled training image. During the simulation, multiple-point statistics are retrieved from the search tree corresponding to the class where the local azimuth/rescaling factor occurs. The application of that process to a 2D horizontal section of a fluvial reservoir indicates that the reproduction of the training patterns in non-stationary MPS models is similar to that observed in stationary models.

This technique can be easily generalized to create non-stationary models using several different training images thought to be representative of the geological heterogeneity in different areas of the reservoir provided that there is a smooth transition between the different training images.

References

Deutsch, C.V. and Journel, A.G., *GSLIB: Geostatistical Software Library and User's Guide*, 2nd edition, Oxford University Press, 1998.

Guardiano, F. and Srivastava, R.M., Multivariate Geostatistics: Beyond Bivariate Moments, in Soares, A., editor, *Geostatistics-Troia*, vol. 1, p. 133-144. Kluwer Academic Publications, 1993.

Holden, L., Hauge, R., Skare, Ø. and Skorstad, A., Modeling of Fluvial Reservoirs with Object Models, *Mathematical Geology*, vol. 30, no. 5, 1998, p. 473-496.

Journel, A. and Alabert, F., Non-Gaussian Data Expansion in the Earth Sciences, *Terra Nova* , no. 1, 1989, p. 123-134.

Lia, O., Tjelmeland, H. and Kjellesvik, L.E., Modeling of Facies Architecture by Marked Point Models, in *Geostatistics-Wollongong*, p. 386-398, Kluwer Academic Publications, 1997.

Strebelle, S., *Sequential Simulation Drawing Structures from Training Images*, Ph.D. Thesis, Department of Geological and Environmental Sciences, Stanford University, 2000.

Strebelle, S., and Journel, A., Reservoir Modeling Using Multiple-point Statistics, paper SPE 71324 presented at the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, Sept. 30-Oct. 3, 2001.

Strebelle, S., K. Payrazyan, and J. Caers, Modeling of a Deepwater Turbidite Reservoir Conditional to Seismic Data Using Multiple-Point Geostatistics: paper SPE 77425 presented at the 2002 SPE Annual Technical Conference and Exhibition, San Antonio, Sept.29-Oct. 2, 2002.

Viseur, S., Stochastic Boolean Simulation of Fluvial Deposits: a New Approach Combining Accuracy and Efficiency, paper SPE 56688 presented at the 1999 SPE Annual Technical Conference and Exhibition, Houston, Oct. 3-6, 1999.

Zhang, T. Program 2001 snesim version 5.0. In *Report 15, Stanford Center for Reservoir Forecasting*, Stanford, CA, 2002.