Mathematical Geology, Vol. 35, No. 8, November 2003 (© 2003)

Spatial Connectivity: From Variograms to Multiple-Point Measures[1]

Sunderrajan Krishnan[2] and A. G. Journel[2]

Anisotropy and curvilinearity are common characteristics of geological structures. Traditional measures of connectivity such as the variogram are rectilinear in that they do not take into account the curvilinearity of these structures. Recent developments in geostatistics have demonstrated and simulated the effect of curvilinearity and multiple-point (mp) connectivity on the output of transfer functions such as flow simulators. A set of curvilinear channels and set of elliptical lenses may share the same variogram and rectilinear connectivity but would yield different flow responses because of their different curvilinearity. A measure of curvilinearity generalizing the variogram measure is therefore proposed. The proposed measure is directional with a tolerance cone and depends on distance with a tolerance, as with an experimental variogram.

KEY WORDS: multiple-point connectivity, curvilinearity.

INTRODUCTION

Geological patterns typically involve simultaneously many locations over possible long distances, they are typically anisotropic and are not rectilinear. Examples are folded surfaces, river channels, and even fault planes that are actually never exactly planar. Modeling of such curvilinear patterns requires measuring the connectivity in the space of the indicators of such structures; the traditional tool offered by geostatistics is the 2-point statistics covariance/variogram which relates any two points in space, for example establishing the probability that any two locations $\mathbf{u}$, $\mathbf{u} + \mathbf{h}$, distant of vector $\mathbf{h}$ be in the same facies. Although already difficult to infer, such variogram statistics is largely insufficient to characterize the shape and spatial continuity of the structure under study, and it should come at no surprise that a model based on only variogram(s) cannot reproduce accurately the structure. Yet continuity of that structure in space may be critical for the application at hand,

[1]Received 23 October 2002; accepted 19 July 2003.

[2]Department of Geology and Environmental Science, Stanford, California 94305; e-mail: sunderk@stanford.edu or journel@pangea.stanford.edu

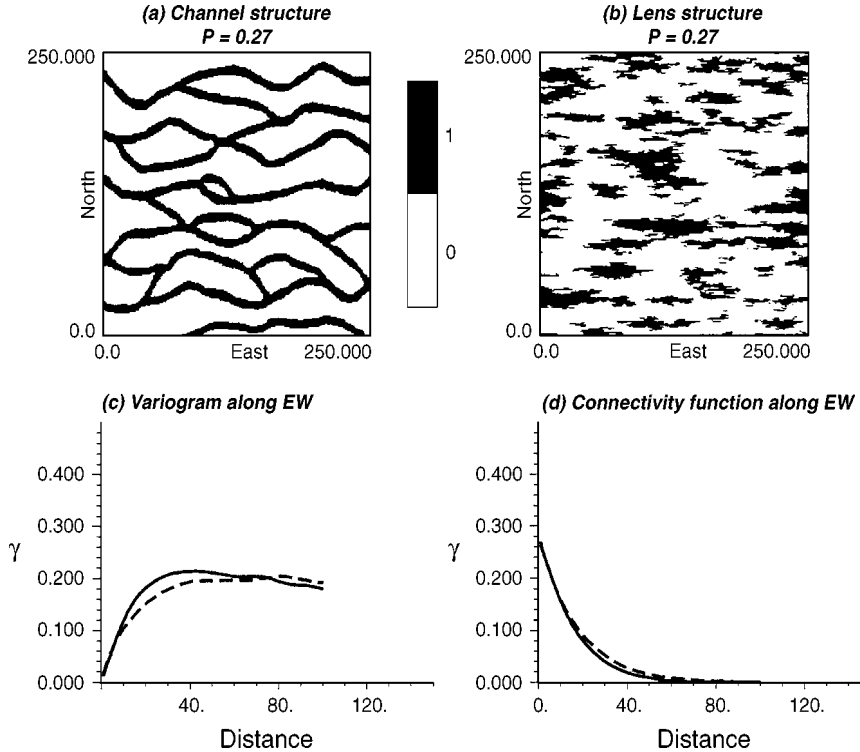

Figure 1. Comparison of EW rectilinear connectivity measures. (a) Channel structure, (b) Lens structure, (c) indicator semivariograms in EW direction, (d) rectilinear multiple-point connectivity functions (EW) (continuous line: channel, broken line: lens structure).

e.g., flow and transport in hydrogeology and gallery design in subsurface mining. We need richer measures of continuity/spatial connectivity which would involve multiple-point (mp) statistics, with much more than 2 points.

Figure 1 provides an introductory example. Figure 1(a) gives a 2D view of what could be seen as a conceptual depiction of a braided river channel meandering over an area sufficiently small not to reveal any trend, say 250×250 m: the channel proportion is $p = 0.27$. Figure 1(b) gives another depiction, of a lenticular system this time, with the same sand proportion $p = 0.27$ and the same general EW direction of continuity. Clearly, the channel structure displays large scale continuity: the channels cross the whole EW extent of the figure, although in a curvilinear fashion.

Denote by $I(\mathbf{u}) = 1$ or 0 presence or absence of sand at any particular location of coordinates vector $\mathbf{u}$. The following statistics were calculated from the $250 \times 250 = 62{,}500$ pixel values constituting the two images of Figure 1(a) and (b):

- mean: $\mathrm{E}\{I(\mathbf{u})\} = p = 0.27$. Hence variance: $\mathrm{E}\{[I(\mathbf{u}) - p]^2\} = p(1-p) = 0.197$.
- semivariogram in the EW direction, with no angle tolerance: $\gamma(\mathbf{h}) = \frac{1}{2}E\{[I(\mathbf{u}+\mathbf{h}) - I(\mathbf{u})]^2\}$, see Figure 1(c).

Recall that an indicator semivariogram is related to the noncentered covariance K(**h**), or 2-point probability to be in sand, by the relation:

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) = p - K(\mathbf{h}) \tag{1}$$

where $K(\mathbf{h}) = E\{I(\mathbf{u}) \cdot I(\mathbf{u}+\mathbf{h})\} = \mathrm{Prob}\{I(\mathbf{u}) = 1, I(\mathbf{u}+\mathbf{h}) = 1\}$

$C(\mathbf{h}) = K(\mathbf{h})\text{-}p^2$:centred indicator covariance

$C(0) = \mathrm{Var}\{I(\mathbf{u})\} = p(1-p)$:indicator variance

The lesser the $\gamma(\mathbf{h})$, the greater the probability $K(\mathbf{h})$ of two points $\mathbf{u}$ and $\mathbf{u}+\mathbf{h}$ to be simultaneously in sand. Figure 1 indicates that the channel and lens structures have similar EW variograms with a slight continuity advantage to the lens structure (lesser γ-values). The 2-point statistics $K(\mathbf{h})$ or $\gamma(\mathbf{h})$ cannot differentiate Figure 1(a) from (b), they are an inappropriate or insufficient measure of spatial connectivity.

A mp rectilinear connectivity function had been introduced (Strebelle, 2002) which generalizes the previous 2-point probability and statistics $K(\mathbf{h})$. Define the function:

$$\begin{aligned} K(\mathbf{h}; n) &= \mathrm{E}\{I(\mathbf{u})I(\mathbf{u}+\mathbf{h})\ldots I(\mathbf{u}+n\mathbf{h})\} \\ &= \mathrm{Prob}\{I(\mathbf{u}) = 1, I(\mathbf{u}+\mathbf{h}) = 1, \ldots, I(\mathbf{u}+n\mathbf{h}) = 1\} \end{aligned} \tag{2}$$

where $\mathbf{h}$ is a unit vector in any given direction, $K(\mathbf{h}; n)$ gives the probability of observing a continuous string of n points in sand.

Remarks

- $K(\mathbf{h}; 0) = p = \mathrm{E}\{I(\mathbf{u})\}$:channel proportion, a 1-point statistic.
- $K(\mathbf{h}; 1) = K(\mathbf{h})$: the traditional noncentered indicator covariance.
- $K(\mathbf{h}; n)$ measures only straight rectilinear connectivity; any undulation would break that connectivity and reduce the K-value.

  Figure (d) gives the two mp K-connectivity functions for the two channel and lens structures, plotted as function of the mp string length $n\mathbf{h}$ in the EW direction. Again the two curves are almost identical, with again a slight advantage (greater connectivity) given to the lens structure!

A measure of spatial connectivity in 2D and 3D space must accept, provide a tolerance, for curvilinearity, just like an experimental variogram allows for an angle tolerance around the target direction, see next section.

- For multiple categories indicators $I(\mathbf{u}; k)$ or indicators $I(\mathbf{u}; z_k)$ based on multiple thresholding of a continuous variable $Z(\mathbf{u})$, expression (2) generalizes into the multivariate distribution function or spatial law (Goovaerts, 1997, p. 65). Indeed, for

$$\begin{aligned} I(\mathbf{u}; z_k) = 1 \quad & \text{if } Z(\mathbf{u}) \leq z_k,\ k = 1, \ldots, K > 2 \\ = 0 \quad & \text{if not.} \end{aligned}$$

The n-point indicator covariance (2) is generalized into the multivariate distribution of the continuous random function $Z(\mathbf{u})$:

$$\begin{aligned} K(\mathbf{h}_1, \ldots, \mathbf{h}_n; z_1, \ldots, z_n) &= \mathrm{E}\{I(\mathbf{u}_1; z_1), \ldots, I(\mathbf{u}_n; z_n)\} \\ &= \mathrm{Prob}\{Z(\mathbf{u}_1) \leq z(\mathbf{u}_1), \ldots, Z(\mathbf{u}_n) \leq z(\mathbf{u}_n)\} \end{aligned} \tag{3}$$

The problem with this multivariate distribution is that it depends on too many parameters: $2n$ for expression (refeq:multin). One would reduce it to expression (2) by considering all equal threshold values and equidistant locations: $\mathbf{u}_j = \mathbf{u} + j\mathbf{h}$.

A CURVILINEAR CONNECTIVITY MEASURE

Similar to the tolerance allowed for calculation of an experimental variogram (see program GAMV of the public domain software GSLIB, Deutsch and Journel, 1998, p. 53), consider the truncated tolerance cone shown in Figure 2(a). In a 2D space, that cone is defined by

- the target direction *azimuth*,
- an angle tolerance $\pm$ atol parameter,
- a band with bandwidth parameter *bandw*, controlling deviation from the central target direction at larger distances, and
- a maximum distance *maxlag* parameter.

Consider a seed location $\mathbf{u}$ in channel: $I(\mathbf{u}) = 1$. Within the previously defined tolerance cone, we would like to connect $\mathbf{u}$ through a channel path to another maximally distant location $\mathbf{u}'$, that distance being measured along the axis of the cone. More precisely, the algorithm proposed in program *curvcon* proceeds as

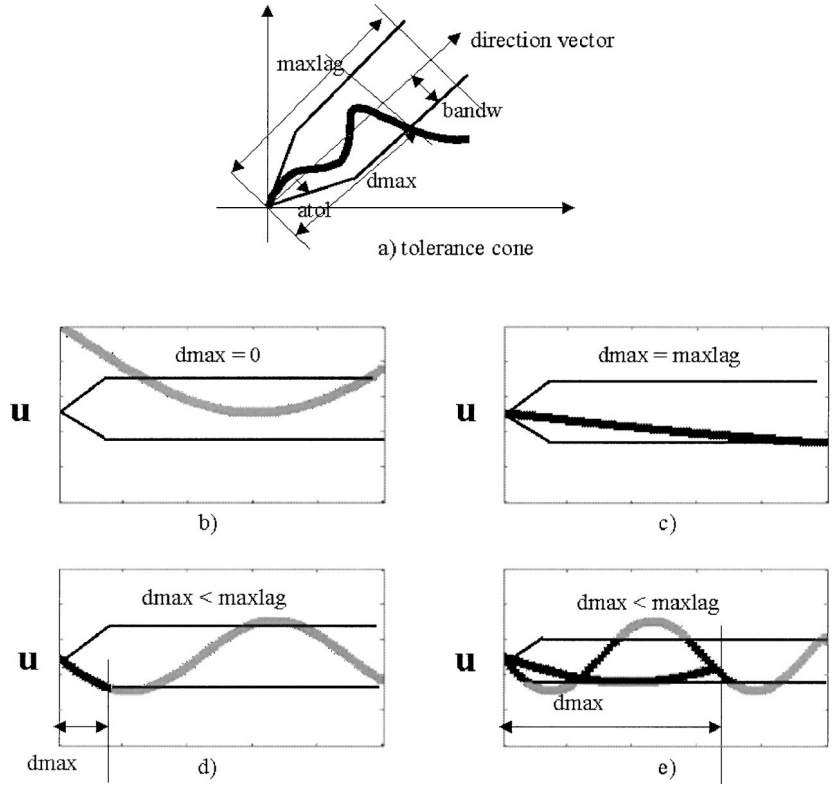

Figure 2. Tolerance cone and dmax connectivity distance. (a) Definition of the tolerance cone in 2D, (b)–(e): the channel must pass by the cone apex **u**; and only its intersection with the cone contributes to dmax.

- Loop through all nodes **u** of the study area.
- If $I(\mathbf{u}) = 0$, set $dmax = 0$ and go to next node **u**, see Figure 3(a).
- If $I(\mathbf{u}) = 1$, i.e., location **u** belongs to the target facies (sand),
  - call MATLAB function *bwselect* to determine within the tolerance cone with apex at **u**, the set (body) of all locations $\mathbf{u}'$ belonging to the same facies, i.e., $I(\mathbf{u}') = 1$, see Figure 2(b)–(e).
  - loop through all these locations $\mathbf{u}'$ and determine the maximum distance *dmax* to the apex **u**. Increment the n proportions of continuous (and curvilinear) channel paths of length $n\mathbf{h} \leq dmax$, recall definition (2).

Figure 2(b) gives an example of a noncontributing channel which does not pass through apex location **u**. Figure 2(c) shows a channel contributing up to distance $dmax = maxlag$.

```
                  Parameters for CURVCON
                  **********************

START OF PARAMETERS:

channels.dat                        .....    data file
1                                   .....   column number
-1e21  1e21                         ..... trimming limits
curv_con.out                        ..... output file
250  0.5  1                        ..... nx, xmn, xsize
250  0.5  1                        ..... ny, ymn, ysize
1  0.5  1                          ..... nx, zmn, zsize
2                                   ..... no directions
0.0  45.0   10.0  50.0             ..... azm, atol, bandh,maxlag
90.0  45.0   10.0 50.0             ..... azm, atol, bandh,maxlag
0                                   .....0/1 standardize
0                                   ..... 0/1 erosion
erode_template.dat                  ..... erosion template file
0                                   ..... 0/1 extension
ext_template.dat                    ..... extension template file
```

Figure 3. Parameter file for program *curvcon*.

Figure 2(d) and (e) shows examples of channels undulating out and in the tolerance cone and contributing up to distance $dmax < maxlag$.

Figure 3 shows the parameter file of program *curvcon* as applied to calculation of curvilinear directional connectivity on the two sand structures of Figure 1(a) and (b). The two directions are NS and EW, the tolerance angle is $atol = \pm 22.5^\circ$, the bandwidth $bondw = 10$, the maximum cone distance is $maxlag = 100$, the distance unit is the lag spacing of the underlying data grid.

The resulting two sets (NS and EW) of two connectivity functions (channel vs. lens structure) are given in Figure 4. If the two structures present about the same NS connectivity, the mp EW connectivity of the channel structure is clearly greater than for the lens structure as compared to the poor differentiation given by the 2-point measure of Figure 1(c) and (d). Note at $n\mathbf{h} = 0$, the mp measure gives back the channel proportion, here $p = 0.27$.

Program *curvcon*

Figure 3 shows the parameter file for program *curvcon*. This program has been written following the style and notations of the GSLIB variogram program *gamv* (Deutsch and Journel, 1998, p. 53–55). The description of input parameters is given below:

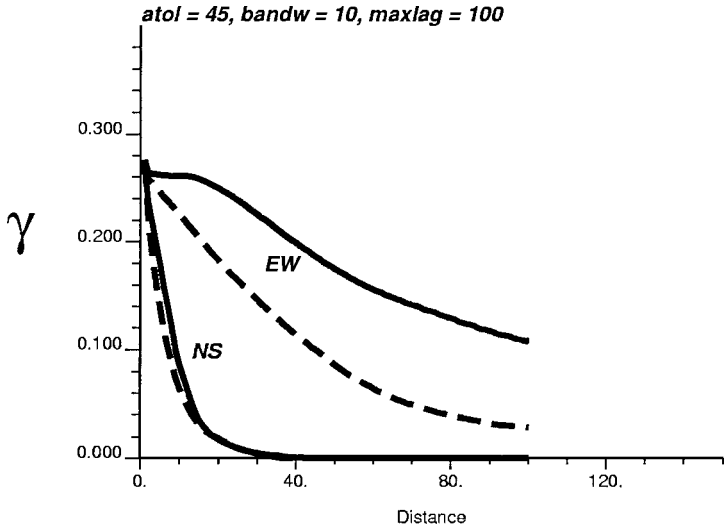

Figure 4. Curvilinear directional connectivity functions. For the structures shown in Figure 1, connectivity function is calculated along NS and EW directions with a tolerance cone (continuous line: channel, dashed line: lens structure).

- *datafl*: The data file with binary indicator data (0, 1) in GEO-EAS format.
- *icolvr*: the column number of the variable in the data file.
- $t_{\min}$, $t_{\max}$: the minimum and maximum trimming limits for data.
- *outfl*: The output file for the connectivity values.
- *nx, xmn, xsiz*: the number of nodes in x direction, origin, and the grid node separation.
- *ny, ymn, ysiz*: the number of nodes in y direction, origin, and the grid node separation.
- *nz, zmn, zsiz*: the number of nodes in z direction, origin, and the grid node separation.
- *ndir*: the number of directions.
- *azi(i), atol(i), bandw(i), maxlag(i)*: parameters azimuth, angle tolerance, bandwidth, and maximum lag distance for each direction.
- *standardize*: if set to 1, the connectivity value is divided by the global proportion of 1s.
- *erode*: if set to 1, will erode the structures in the image.
- *erodfl*: input file for erosion template.
- *extend*: if set to 1, will extend the structures in the image.
- *extendfl*: input file for extension template.

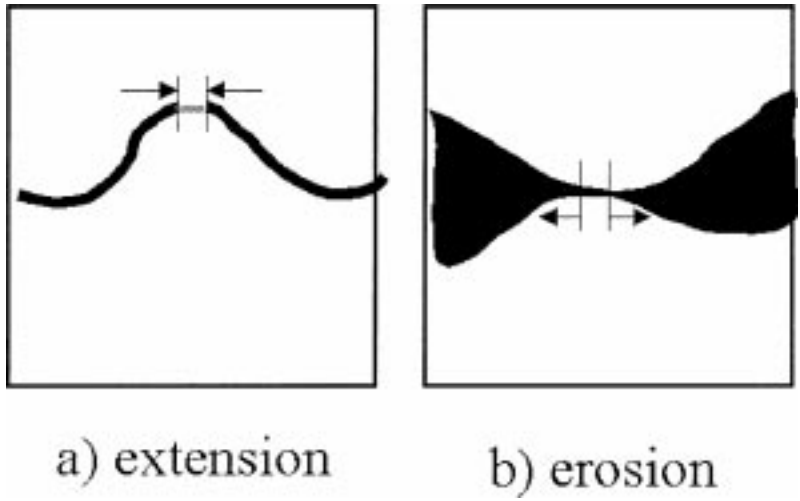

Figure 5. Tolerance for (dis)connecting bodies. (a) Extending the original two bodies found in the cone allows reconnecting the channel, (b) Eroding the original single body allows disconnecting the two distinct masses.

Tolerance for (dis)Connection

The program *curvcon* allows for small extension/erosion of the bodies found in the tolerance cones in order to

- connect bodies that are close together, see Figure 5(a).
- disconnect parts if a body that were too thinly linked together, see Figure 5(b).

The algorithm consist of extending/eroding the original bodies found by program *bwselect* before processing them. Extension amounts to add a specified number of one-indicator values to the original bodies borders in each of the x and y directions of the 2D space. Erosion amounts to substract that specified number.

If a (rather large) tolerance of 5 EW pixels is added to the lens bodies of Figure 1(b), the resulting EW mp connectivity is considerably increased, starting with the total proportion of sand increasing from $p = 0.27$ to $p = 0.42$. For a fair comparison with the original result of Figure 3, the connectivity probabilities have been standardized by that sand proportion resulting in a unit value at the origin $n\mathbf{h} = 0$, see Figure 6.

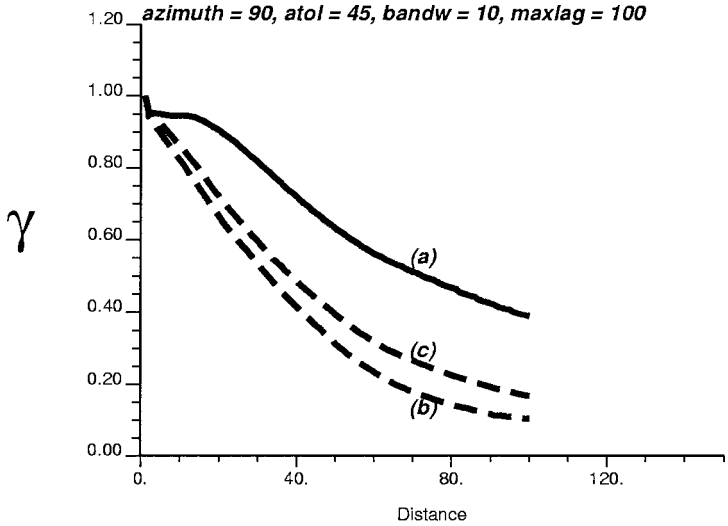

Figure 6. Standardized directional connectivity functions. Channel (a) and original lens (b) EW connectivity, and Lens connectivity (c) allowing for extension of lens bodies by 5 pixels in EW direction.

Other Connectivity Measures

The mp directional connectivity measure (2) proposed is a generalization of the 2-point variogram/covariance statistics (1). For any particular application, another more specific measure can be built, but it would lack the generality of a statistics but be more physically significant. An example is given by the directional effective permeability (or conductivity) of a binary porous media obtained by solving the pressure field over the study area, assuming no flow boundaries along the faces parallel to the flow direction (Deutsch, 2002, p. 333).

For both the channel and lens structure displayed in Figures 1(a) and (b), a constant permeability of 300 mD (milliDarcies) was given to all sand nodes and 1 mD to all non-sand/mud nodes. Program *flowsim* (Deutsch, 1992) was used to give the two effective permeabilities:

56.7 mD for the channel structure (Fig. 1(a)).
9.3 mD for the lens structure (Fig. 1(b)).

Prospectives

A clearly defined mp statistics such as the n-point rectilinear covariance $K(\mathbf{h}; n)$, itself a generalization of the classical 2-point covariance $K(\mathbf{h})$, is useful

as the foundation for the next frontier geostatistics looks at spatial patterns rather than mere 2-point correlations. The mp statistics brings superior pattern resolution and the potential for much better pattern reproduction/simulation.

The problems faced by mp geostatistics are, however, multiple and formidable:

1. First, inference of mp statistics requires a vast amount of data on a regular grid, typically not available in the subsurface. One must rely on densely informed training images, such as Figure l(a) and (b), deemed representative of the structures under study. As a consequence, the mp structural function equivalent to the traditional variogram is not meant to estimate an unknown structure, instead it characterizes a known (conceptual) structure to be mapped into the study area after conditioning to the actual data. The choice is clear, either one sticks to 2-point statistics that may be inferred (typically very poorly!) from actual data and ends up ignoring fundamental prior information such as the formation is of channel-type not of discontinuities lens-type; or one decides to capitalize on that prior information, generates a conceptual training image of the type of Figure 1(a), and proceeds with mp geostatistics.

   Experience gained through the practice of variogram inference can be exported to inference of mp statistics from the training images, as was done above for the tolerance cone of directional connectivity functions.
2. Second and *no less* difficult, reproduction of those mp statistics conditional to actual local data. One could think of using iterative algorithms a la simulated annealing (Deutsch, 2002, p. 275), perturbing progressively a random field until the target mp statistics are approximately reproduced. Iterative algorithms are typically too slow to be practical for most applications involving a substantial grid (more than 10^3 nodes). Direct sequential algorithms such as pioneered by Srivastava (Guardino and Srivastava, 1992) and Strebelle (2002) offer greater immediate potential.

REFERENCES

Bour, O., and Davy, P., 1998, On the connectivity of three-dimensional fault networks: Water Resour. Res., v. 34, no. 10, p. 2611–2622.

Deutsch, C., 2002, Reservoir modeling: Oxford University Press, New York, 376 p.

Deutsch, C., and Journel, A. G., 1998, GSLIB: Geostatistical software library and user's guide, 2nd edn.: Oxford University Press, New York, 369 p.

Deutsch, C., 1992, Ammending techniques applied to research modeling and the integration of geological and engineering data: PhD Thesis, Stanford University, 306 p.

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

Guardino, F., and Srivastava, R. M., 1992, Borrowing complex geometries from training images: The extended normal equations algorithm: Annual Report of Stanford Center for Reservoir Forecasting, Stanford University.

Renshaw, C., 1999, Connectivity of joint networks with power law length distributions: Water Resour. Res., v. 35, no. 9, p. 2661–2670.

Srivastava, R. M., and Dunn, M., 1995, Modeling sub-seismic faults for the Hibernia reservoir: FSS Canada Internal report 95-07, Vancouver, BC.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics, Math. Geol., v. 34, no. 1, p. 1–21.

Tran, T., 1995, Stochastic simulation of permeability fields and their scale-up for flow modeling, PhD thesis: Stanford University, California.