

# The Components of Geostatistical Simulation

Carol A. Gotway<sup>1</sup> and Brian M. Rutherford<sup>2</sup>

---

**Abstract.** There are many approaches to geostatistical simulation that can be used to generate realizations of random fields. These approaches differ fundamentally in a number of ways. First, each approach is inherently different and will produce fields with different statistical and geostatistical properties. Second, the approaches differ with respect to the choice of the features of the region that are to be modeled, and how closely the generated realizations reproduce these features. Some fluctuation in the statistical and geostatistical properties of different realizations of the same random field are natural and desirable, but the proper amount of deviation is an open question. Finally the approaches differ in how the conditioning information is incorporated. Depending on the source of randomness and the uncertainty in the given data, direct conditioning of realizations is not always desirable. In this paper, we discuss and illustrate these differences in order to emphasize the importance of these components in geostatistical simulation.

---

## INTRODUCTION

Geostatistical simulation, as illustrated in Figure 1, provides a way to quantify uncertainty in the prediction of a complex system response. Some information is usually available on a process of interest (e.g., measurements on a hydrogeologic parameter of interest in a particular region), but the transfer function (e.g., a groundwater flow model) may require a detailed spatial map of the parameter on which the measurements were obtained. (Such a map can be created through spatial prediction techniques such as kriging.). The value of the transfer function can then be computed over the mapped values to obtain a prediction of a system response (e.g., groundwater travel time), but an estimate of the uncertainty associated with the prediction, often as important as the predicted value itself, is typically much more difficult to infer. One alternative is to generate realizations of the random field that share the available information on the parameter of interest and use the transfer function to compute a system response for each. If the realizations characterize the spatial uncertainty in the parameters of interest, the resulting distribution of predicted system responses will reflect this uncertainty.

<sup>1</sup> Assistant Professor of Biometry, University of Nebraska, Lincoln, NE

<sup>2</sup> Senior Member of Technical Staff, Sandia National Laboratories, Albuquerque, NM

This work was supported by the United States Department of Energy under Contract DE-AC04-94AL85000.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

62  
**MASTER**

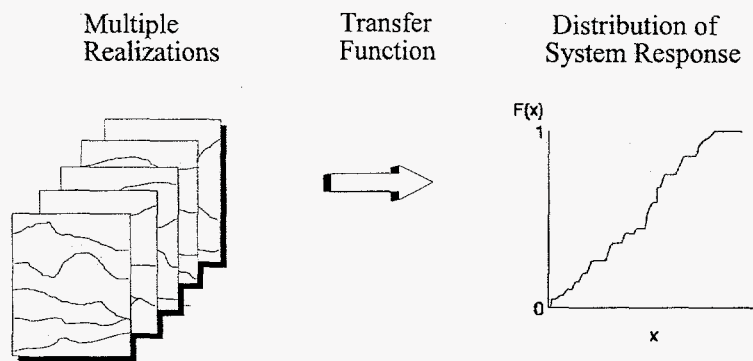


Figure 1. Illustration of Stochastic Simulation (Journel, 1989).

Although originally introduced as a geostatistical approach to uncertainty analysis (Journel, 1989), geostatistical simulation has grown to provide a variety of new approaches to stochastic modeling that can incorporate complex spatial relationships. These methods are frequently used in hydrology, petroleum engineering, and the environmental sciences, not only for uncertainty analysis, but also for risk analysis, decision analysis, and stochastic modeling.

Many different approaches to geostatistical simulation can be used to construct the realizations. For continuous variables these include the turning bands method (Journel, 1974), spectral methods (Borgman et al., 1984; Gutjahr, 1989), LU or Cholesky decomposition (Alabert, 1987; Davis, 1987), sequential Gaussian simulation (Journel, 1989), and fractal approaches (Hewett and Behrens, 1990). Approaches to simulation of discrete or categorical variables include Boolean methods (Serra, 1982; Jeulin, 1987; Chautru, 1989) and the more general random closed set approaches (Stoyen et al., 1987), and the use of truncated Gaussian random functions (Matheron et al., 1987). Sequential indicator simulation (Journel and Alabert, 1989), simulated annealing (Kirkpatrick et al., 1983; Deutsch and Journel, 1992) and genetic algorithms (Whitley, 1994) can be used with both continuous or categorical variables.

For any particular type of data (continuous, discrete, or categorical), different approaches will assign different statistical and spatial features to the generated fields. These differences could be due to: (1) differences in the statistical model used to characterize the phenomenon under study. These differences are most blatantly represented by the criteria used to describe spatial

data relationships and characterize available information. We refer to such criteria as *summary functions*. Examples of summary functions include the univariate data distribution, the covariance function, indicator semivariograms, and the shape and intensity of primary elements in Boolean simulations; (2) differences in algorithms (e.g., using sequential Gaussian simulation versus using LU decomposition); (3) the acceptable deviation between the summary functions specified to the algorithm and those calculated from resulting realizations; and (4) the use of conditioning information. Since all of these differences will translate into differences in the nature of the spatial heterogeneity reflected in the resulting realizations, and hence into differences in uncertainty assessments, risk calculations, etc., and ultimately into differences in conclusions and managerial decisions, it is important to consider each of these components carefully when using geostatistical simulation. In the remainder of this paper, we elaborate on these components of geostatistical simulation, discuss some of the different choices within each component, and illustrate the effect of these differences on the nature of the simulated fields and on functions derived from them.

## COMPONENTS OF GEOSTATISTICAL SIMULATION

When using geostatistical simulation to characterize the spatial variation of parameters of interest, the modeler has to make several important choices. Each choice will affect the nature of the resulting realizations and the values of functions derived from them. Thus, careful thought with regard to the major components of geostatistical simulation is crucial in modeling the uncertainty of spatial functions.

### **Choice of statistical model and selection of summary functions**

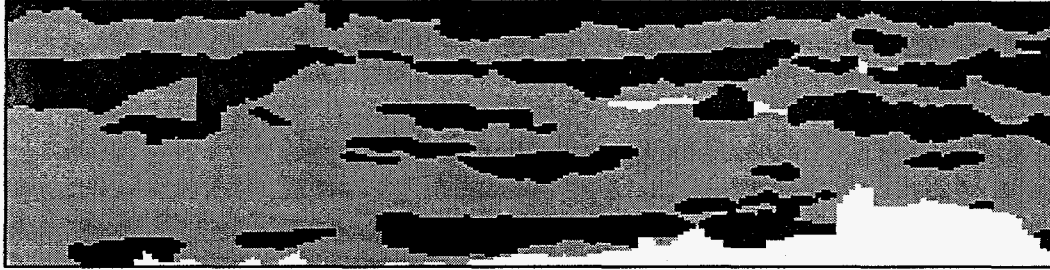
For geostatistical simulation of a spatial process, a statistical model must describe a multivariate probability distribution for the spatial data of interest. Given such a distribution, realizations can be generated by sampling randomly from this specified distribution. For example, in LU decomposition, the underlying multivariate distribution is Gaussian, and the parametric form for this distribution can be found in most introductory statistical texts. Similarly, general formula for the multivariate distributions of Boolean random sets, as well as formula for distributions of other random sets, are given in Stoyen et al., (1987). In other cases, such as the sequential indicator simulation model, or certain Markov Random Field Models, the exact form of the multivariate distribution is intractable or unknown, and is conditionally specified using a local neighborhood structure.

Once a particular model has been chosen, the multivariate distribution is specified through the use of summary functions. For example, generation of realizations from a multivariate Gaussian distribution requires specification of the

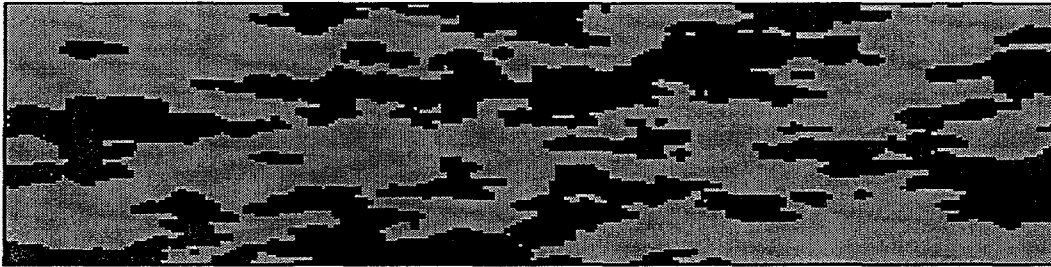
process mean and covariance structure which is typically determined through the semivariogram. In Boolean models, the multivariate distribution depends on the intensity of the Poisson process, the shape of the primary elements, and the probability distributions of their sizes. In sequential indicator simulation, the multivariate distribution is specified nonparametrically through the use of indicator semivariograms. Consequently, using a particular simulation algorithm that requires a specified set of summary functions is equivalent to specifying a multivariate probability model for the spatial process. These models necessarily emphasize different spatial features of the process, so it is important to determine what features are most important in the application of interest and how best to quantify them from the available information. In applications where a transfer function is evaluated over the generated realizations, it is important to determine the features of the process that will greatly influence the value calculated through the transfer function. For example, in reservoir modeling, it is often important to characterize those features that influence the flow of fluids through reservoirs, and in environmental risk analysis, it may be important to characterize those features that control the spread of a contaminant of special interest.

Emphasizing different features through the use of different summary functions can give rise to realizations that appear quite different. As an example, consider the top image in Figure 2. It shows a geologic section (obtained from digitization) of interbedded debris flow (fine grained) and sheet flood (coarse grained) units produced as part of site-characterization studies at the Greater Confinement Disposal Facility in Southern Nevada (Rutherford and Gotway, 1994; McCord et al., 1996). Recharge through sections of this type is an important component of such studies and the position, shape, and amount of each type of unit are important for modeling vadose zone flow. The last three images in Figure 2 are unconditional realizations generated from statistics inferred from the training image using sequential indicator simulation, Boolean simulation, and a more general random set approach. Sequential indicator simulation uses two summary functions, the proportion of debris flow and the indicator semivariogram. The indicator semivariogram attempts to incorporate the connectivity of the permeable and impermeable zones. Although the shape of the debris flow units can be controlled to some degree by the anisotropy reflected in the indicator semivariograms, the shapes are somewhat limited. The Boolean model uses summary functions that specify the intensity of the process, and the shape and size of the debris flow units. Here the emphasis is on shape, rather than connectivity. The last realization is based on the theory of random closed sets (Kendall, 1974). In this application the summary functions are described by a system of structuring elements, and any random set can be characterized through its intersection with these elements. The last realization was generated using 19 structuring elements consisting of lines, triangles and rectangles. Here there is an attempt to account for both shape and connectivity.

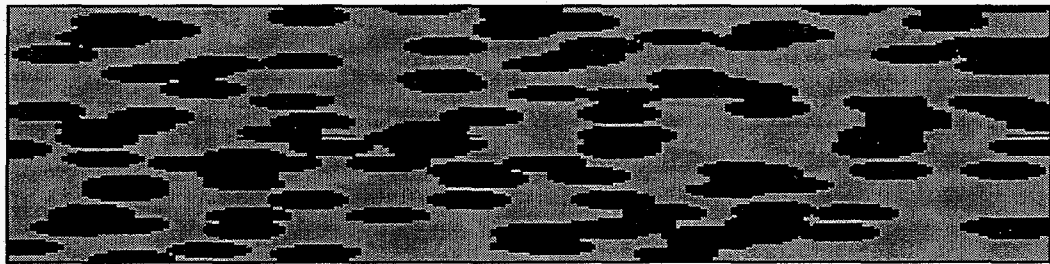
**GCD Training Image**



**Sequential Indicator Realization**



**Boolean Realization**



**Hitting Functions Realization**



Figure 2. The Use of Different Summary Functions in Lithologic Simulation.

The use of multiple-point statistics (Deutsch and Journel, 1992; Guardiano and Srivastava, 1993;), might also be used to incorporate more features of importance. Note that in such applications, the goal is not to reproduce the training image, but to capture essential features of the image (not always visible by eye) that are most important to the transfer function of interest.

### **Choice of Simulation Algorithm**

In addition to the selection of an appropriate set of summary functions that determine the modeling approach, different algorithms for generating the realizations often impact the results. In implementing most geostatistical simulation algorithms, there are a number of parameters that must be made specific, often based on a tradeoff between speed and accuracy. Examples include the number of turning bands lines, the number of conditioning values retained in sequential simulation, the choice of annealing schedule and convergence criteria, and the number of structural elements in random set generation. In many cases, blatantly improper choices can be clearly reflected in the realizations, e.g., the appearance of banding effects in turning bands method when the number of lines is small. In other cases, apparently reasonable selections can also have an effect on the realizations and on the resulting uncertainty distributions derived from them. In a comparative study of several geostatistical simulation algorithms using a variety of exhaustive data sets and transfer functions, Gotway and Rutherford (1994) illustrate these algorithmic differences. Figure 3 shows the uncertainty distribution of the average of local geometric means computed from conditional realizations of a Gaussian field. Although the shape of the uncertainty distributions is similar, they differ in the upper tail which could be crucial for risk assessments. Comparing the distributions obtained using LU Decomposition, Sequential Gaussian, and Turning Bands algorithms shows the effect of algorithmic differences on the response uncertainty distributions. Comparing the distribution obtained using sequential indicator simulation to the others illustrates the effect of choice of summary functions. As shown in Gotway and Rutherford (1994), these differences are much more pronounced for unconditional realizations.

### **The Acceptable Deviation**

The acceptable deviation refers to the difference between the summary functions calculated from realizations and that specified by the model. For example, should the variogram calculated from each realization exactly match the variogram specified? Certainly, the answer to this question is no, since, in simulation, some deviation of summary functions is desirable in order to reflect uncertainty. However, the amount of deviation acceptable is an open question. Unless there is "post-generation" alteration to the realizations, these deviations depend on the specified summary functions and on the simulation algorithm. Often, simulation algorithms are compared and evaluated based on their ability to

reproduce the desired functions over an average of realizations. However, the variability in the summary functions can be quite different depending on the particular algorithm used and still, on the average, reproduce the specified summary functions. More importantly, such a reproduction assumes that the summary functions specified are in fact known, when in reality they are estimated from available data.

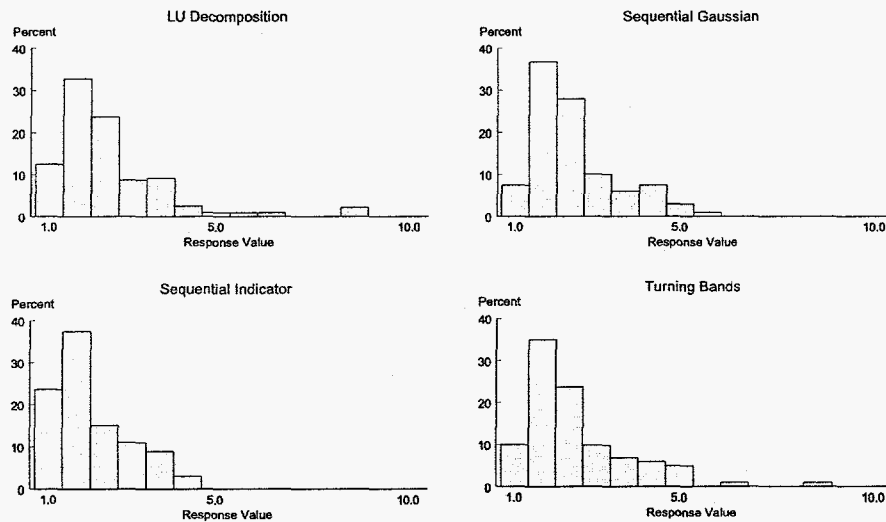


Figure 3. Response Uncertainty Distributions.

The fundamental concern with the acceptable deviation is that differences between specified summary functions and those calculated from generated realizations should reflect: (1) the uncertainty in input summary functions; (2) the natural variability that comes with random spatial sampling over a finite region to generate realizations, and not (3) the variability (or lack of variability) introduced by the simulation algorithm. None of the available algorithms account for (1), and with the exception of algorithms in which a multivariate distribution is directly sampled (such as LU decomposition) it is also not clear whether (2) is accomplished correctly. For many applications, (3) may be the primary determinate of deviations between summary functions.

### The Use of Conditioning Information

Conditioning a simulation on known values or soft information (incomplete information concerning the region of interest) must be done in a way that reflects the true conditions imposed on the simulated region by that data or information. At present, the common approach for algorithms using a sequential method of generation is to treat previously generated values as known data values in the simulation. The study in Hansen (1992) indicates that this approach might "over-condition" or "under-condition" the realizations. The result is a biased uncertainty distribution of response that may be unrealistically wide or too narrow. Moreover, if data are measured with error, conditioning on these data may not be warranted; rather some filter must first be applied to reduce the noise introduced by measurement error.

In many applications, the summary functions are derived from a training image, and simulation of complex geologic structures is often based on a training image that is derived from outcrop mapping (e.g., Figure 2). However, the goal of simulation is not to reproduce a training image. Rather, the objective is to generate realizations that account for information believed relevant to the general process under study. Such information might include the general shape and relative distributions of important units, connectivity patterns, cyclic trends, etc., that could be expected to be general properties of the process. It most probably does not include exact values obtained from the training image. In fact, Guardiano and Srivastava (1993) note that "the training image is itself is not the underlying phenomenon. It is assumed, however, that both share some structural properties such as specific multiple-point covariances." Consequently, it seems very difficult to justify using the exact values (or even some randomly selected subset of them) as conditioning information.

### REFERENCES

- Alabert, F.G. 1987. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, 19: 369-386.
- Borgman, L.E., M. Taheri, and R. Hagen. 1984. Three-dimensional, frequency-domain simulations of geological variables. In G. Verly et al. (eds.), *Geostatistics for Natural Resources Characterization*. Reidel, Dordrecht, pp. 517-541.
- Chautru, J.M. 1989. The use of Boolean random functions in geostatistics. In M. Armstrong (ed.), *Geostatistics*, Vol. 1, Kluwer, Dordrecht, pp. 201-212.
- Davis, M.W. 1987. Generating large stochastic simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, 19: 91-98.



- Deutsch, C.V. and A.G. Journel. 1992. Annealing techniques applied to the integration of geological and engineering data. Stanford Center for Reservoir Forecasting, Stanford, CA.
- Gotway, C.A., and B.M. Rutherford. 1994. Stochastic simulation for imaging spatial uncertainty: Comparison and evaluation of available algorithms. In M. Armstrong and P.A. Dowd (eds.), *Geostatistical Simulations*, Kluwer, Dordrecht, pp. 1-21.
- Guardiano, F. And M. Srivastava. 1993. Multivariate geostatistics: Beyond bivariate moments. In Soares, A. (ed.), *Geostatistics Tró ia '92*. Kluwer, Dordrecht, pp. 133-144.
- Gutjahr, A. 1989. *Fast Fourier Transforms for Random Fields*. Technical Report No. 4-R58-2690R, Los Alamos, NM.
- Hansen, K.M. 1992. The use of sequential indicator simulation to characterize geostatistical uncertainty. Technical Report No. SAND91-0758, Sandia National Laboratories, Albuquerque, NM.
- Hewett, T. And R. Behrens. 1990. Conditional simulation of reservoir heterogeneity with fractals. *Formation Evaluation*, 330-310.
- Jeulin, D. 1987. Anisotropic rough surface modeling by random morphological function. *Acta Stereologica*, 6:183-189.
- Journel, A.G. 1974. Geostatistics for conditional simulation of ore bodies. *Economic Geology*, 69:673-687.
- Journel, A.G. 1989. *Fundamentals of Geostatistics in Five Lessons*. American Geophysical Union, Washington, D.C.
- Journel, A.G., and Alabert, F. (1989). Non-Gaussian data expansion in the earth sciences. *Terra Nova*, 1:123-124.
- Kendall, D.G. 1974. Foundations of a theory of random sets. In E. Harding and D. Kendall (eds.), *Stochastic Geometry*, Wiley, Chichester, pp. 322-376.
- Kirkpatrick, S., C.D., Gelatt, and M.P. Vecchi. 1983. Optimization by simulated annealing. *Science*, 220:671-680.

- Matheron, G., H. Beucher, C. De Fouquet, A. Galli, D. Guerillot, and C. Ravenne. 1987. Conditional simulation of the geometry of fluvio-deltaic reservoirs. SPE Paper No. 16753, 62nd Annual Technical Conference and Exhibition of the Society of Petroleum Engineers, Dallas, TX., September 27-30.
- McCord, J.T., C.A. Gotway, and S.H. Conrad. 1996. Inferring recharge rates using environmental tracers in heterogeneous vadose zones. *Water Resources Research*, to appear.
- Rutherford, B.M., and C.A. Gotway. 1994. Selecting features from spatial data for use in stochastic simulation. *Proceedings of the 35th Annual Meeting of the Institute of Nuclear Materials Management*, July 17-20, Naples, FL.
- Serra, J. 1982. *Image Analysis and Mathematical Morphology*. Academic Press, London.
- Stoyen, D., W.S. Kendall, and J. Mecke. 1987. *Stochastic Geometry and Its Applications*. Wiley, New York.
- Whitley, D. 1994. A genetic algorithm tutorial. *Statistics and Computing*, 4: 65-85.

### **BIOGRAPHICAL SKETCH**

Carol A. Gotway is Assistant Professor of Biometry at the University of Nebraska. Since graduating from Iowa State University with a Ph.D in statistics, Carol has continued her interest in spatial statistics and applied statistics in the physical and environmental sciences. She currently provides statistical consultation in many areas including agriculture, ecology, forestry, hydrology, geology, and the environmental sciences.

Brian M. Rutherford is a statistical consultant with Sandia National Laboratories in Albuquerque, NM. He graduated from the University of Arizona with a Ph.D. in systems engineering. Brian is currently involved in many Department of Energy and United States Environmental Protection agency sponsored projects concerned with nuclear waste disposal and environmental remediation.

### **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.