Sequential Gaussian Simulation

J. Jaime Gómez-Hernández
Research Institute of Water and Environmental Engineering
Universitat Politècnica de València
Camino de Vera, s/n
46022 Valencia
Spain

February 4, 2021

Definition

Sequential Gaussian simulation is a computer-based technique for the generation of realizations $z(\mathbf{x})$ from a multiGaussian random function $Z(\mathbf{x})$ defined on a finite point set $\mathcal{D}$, generally discretizing into N voxels a one-, two-, or three-dimensional area of interest. There are many other techniques capable of generating realizations from a multiGaussian random function, but what makes unique the sequential Gaussian simulation algorithm is the possibility of generating realizations with a reasonable computing effort over arbitrarily large domains with N easily reaching values larger than millions. This is achieved thanks to the sequential simulation principle derived and implemented by André Journel's group at Stanford University in the late 1980s (Gómez-Hernández and Srivastava, 2021).

Random Function

A random function $Z(\mathbf{x})$ can be defined as a rule that assigns a realization to the outcome of an experiment

$$Z(\mathbf{x}) \sim \{z(\mathbf{x},\theta)\}, \forall \theta \in \Omega, \mathbf{x} \in \mathcal{D}. \tag{1}$$

In the case of numerical simulations in discretized domains, $\mathcal{D}$ is defined as the set of N points falling at the centers of the voxels discretizing the area of study; $\mathbf{x}$ is any such point, θ is the outcome of an experiment and Ω is

the sample space containing all possible outcomes. In the remainder, the dependency of any realization on θ will be dropped and lower capital letters will be used to refer to a realization or the values it could take, and upper capital letters will be used for random variables or random functions.

A random function can also be defined by all the n-variate distributions

$$F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; z_1, z_2, \ldots, z_n) = \\ \text{Prob}(Z(\mathbf{x}_1) \leq z_1, Z(\mathbf{x}_2) \leq z_2, \ldots, Z(\mathbf{x}_n) \leq z_n), \tag{2}$$

for any subset of n points out of the N points discretizing the study area. In the previous expression $Z(\mathbf{x}_i)$ is a random variable defined at $\mathbf{x}_i$ through the ensemble of realizations.

When the random function is multiGaussian each random variable can take any value in $\Re$ and the distributions in Eq. (2) have the following expression (Anderson, 1984)

$$F(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n; z_1, z_2, \ldots, z_n) = \\ (2\pi)^{-\frac{n}{2}} |\Sigma|^{\frac{1}{2}} \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \cdots \int_{-\infty}^{z_n} e^{-\frac{1}{2}(\mathbf{z}-\mu)^T \Sigma^{-1} (\mathbf{z}-\mu)} d\mathbf{z}, \tag{3}$$

where μ is a column vector with the expected values of the random variables

$$\mu = \left\{ \begin{array}{l} E\{Z(\mathbf{x}_1)\} \\ E\{Z(\mathbf{x}_2)\} \\ \cdots \\ E\{Z(\mathbf{x}_n)\} \end{array} \right\}, \tag{4}$$

$\mathbf{z}$ is an n-component column vector with the integration variables, and Σ is a matrix of covariances

$$\Sigma = \left(\begin{array}{cccc} C(Z(\mathbf{x}_1), Z(\mathbf{x}_1)) & C(Z(\mathbf{x}_1), Z(\mathbf{x}_2)) & \cdots & C(Z(\mathbf{x}_1), Z(\mathbf{x}_n)) \\ C(Z(\mathbf{x}_2), Z(\mathbf{x}_1)) & C(Z(\mathbf{x}_2), Z(\mathbf{x}_2)) & \cdots & C(Z(\mathbf{x}_2), Z(\mathbf{x}_n)) \\ \vdots & \vdots & \ddots & \vdots \\ C(Z(\mathbf{x}_n), Z(\mathbf{x}_1)) & C(Z(\mathbf{x}_n), Z(\mathbf{x}_2)) & \cdots & C(Z(\mathbf{x}_n), Z(\mathbf{x}_n)) \end{array} \right). \tag{5}$$

The attractiveness of the multiGaussian distribution is that it is fully characterized by its expected value and its covariance between any two points in its domain. At the same time, this is its major drawback since such a characterization prevents any control on the higher-order moments, which are fully determined as functions of the mean and covariance. Once a multiGaussian random function is chosen, the user cannot expect higher-order

moments different from the ones to be derived from the distributions in Eq. (3). For instance, multiGaussianity will always prevent high continuity of extreme values, a property much desirable in the geosciences when modeling high-permeability fracture zones or impermeable shale barriers.

Sequential Simulation

The sequential simulation algorithm was proposed by André Journel in the late 1980s to address the problem of how to generate a realization from a non-Gaussian random function defined on the basis of indicator functions (functions that take a value of 0 or 1 depending if the argument is above or below a given threshold). The first publicly available code was published by Gómez-Hernández and Srivastava (1990) and it describes the principles of the algorithm and many implementation issues. Soon after, it was realized that sequential simulation could also be applied for the generation of realizations from other random functions, particularly, from multiGaussian random functions, and sequential Gaussian simulation was born (Deutsch and Journel, 1992; Gómez-Hernández and Journel, 1993).

The basic idea behind sequential simulation is quite simple (Gómez-Hernández and Cassiraga, 1994). Given a random function defined over a finite point set of size N, the probability distribution in Eq. (2) can be decomposed, by recursively applying the definition of conditional probability, as follows:

$$\begin{aligned} F(Z(\mathbf{x}_1), Z(\mathbf{x}_2), \ldots, Z(\mathbf{x}_N)) \;=\; & F(Z(\mathbf{u}_1)) \cdot \\ & F(Z(\mathbf{x}_2)|Z(\mathbf{x}_1)) \cdot \\ & F(Z(\mathbf{x}_3)|Z(\mathbf{x}_1), Z(\mathbf{x}_2)) \cdots \\ & F(Z(\mathbf{x}_N)|Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_{N-1})). \end{aligned} \quad (6)$$

Using this decomposition, drawing a realization from a full N-variate distribution can be replaced by drawing, sequentially, N values from N conditional univariate distributions. The only requirement would be the ability to compute the conditional distribution of any random variable given any number of random variables. What, for the case of a multiGaussian random function, is trivial.

The conditional distribution of a random variable given any number of random variables when these random variables are members of a multiGaussian random function is a Gaussian distribution with mean and covariance given by the solution of a set of normal equations (Anderson, 1984), also

known as the simple kriging equations. Its expression is

$$\text{Prob}(Z(\mathbf{x}_i) \leq z_i | Z(\mathbf{x}_1) = z_1, Z(\mathbf{x}_2) = z_2, \ldots, Z(\mathbf{x}_n) = z_n) = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{z_i} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz \qquad (7)$$

with μ being the simple kriging estimate of $z(\mathbf{x}_i)$ using as data $\{z_1, z_2, \ldots, z_n\}$ at locations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, respectively; and σ^2 the corresponding simple kriging variance.

The conditional distribution in Eq (7) is generally approximated, when the number n is large, by restricting the data used to solve the kriging equations to a predefined maximum number within a given search neighborhood around $\mathbf{x}_i$. Only the subset of $z_1, z_2, \ldots, z_n$ that falls within the neighborhood, and only the closest values up to certain maximum number are used to build the simple kriging system. This approximation is necessary since, for a realization with N larger than a few hundreds, the decomposition in Eq. 6 results in conditional distributions depending in too many conditioning data; determining the parameters of those conditional distributions would amount to solving kriging systems with hundreds of equations; in order to keep the computation of the conditional realizations —that has to be repeated for each point— at a reasonable computing cost, there is a need to reduce the size of the kriging system, what can be achieved using the concept of a search neighborhood and limiting the maximum number of points.

Sequential Gaussian Simulation Algorithm

Drawing a realization from a multiGaussian random function can be achieved as follows:

1. Define a random permutation of the natural numbers 1 to N that will serve to define how the expression (6) is built and also the order in which the points will be visited,
2. Following the order in the random permutation visit each point and
   (a) Search for the values that have already been simulated falling within a search neighborhood centered at the point to simulated,
   (b) Solve the simple kriging equations to determine the mean and the variance of the Gaussian conditional distribution,
   (c) Draw a random number from the conditional distribution and assign it to the point.

3. Return to the beginning for the generation of a new realization.

Remarks

Equation (6) is valid for any permutation of numbers 1 to N; a random permutation is suggested because it was found that the approximation of the conditional distribution retaining only the closest simulated points introduces some artifacts in the realizations when a structured sequence is used (for instance, simulating the points following the rows or columns of a two-dimensional realization).

Conditional realizations, that is, realizations that honor the observed data at data locations can be generated simply by including their locations in the domain $\mathcal{D}$ and by including the observed values as known (already simulated) values at their locations prior to the start of any simulation.

There are other implementation issues, many of which can be consulted in the review paper by Gómez-Hernández and Srivastava (2021).

Cross-references

Journel, Andre
Kriging
Monte-Carlo Method
Multivariate Analysis
Random Variable
Realizations
Simulation
Spatial Statistics

References

Anderson TW (1984) Multivariate statistical analysis. Wiley, New York

Deutsch CV, Journel AG (1992) GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press, New York

Gómez-Hernández JJ, Cassiraga EF (1994) Theory and practice of sequential simulation. In: Armstrong M, Dowd P (eds) Geostatistical Simulations, Kluwer Academic Publishers, pp 111–124

Gómez-Hernández JJ, Journel AG (1993) Joint sequential simulation of Multi-Gaussian fields. In: Soares A (ed) Geostatistics Tróia '92, Kluwer Academic Publishers, Dordrecht, vol 1, pp 85–94

Gómez-Hernández JJ, Srivastava RM (1990) ISIM3D: An ANSI-C three dimensional multiple indicator conditional simulation program. Computer and Geosciences 16(4):395–440

Gómez-Hernández JJ, Srivastava RM (2021) One step at a time: The origins of sequential simulation and beyond. Mathematical Geosciences 53(2)