

Choosing and Using Simulation Algorithms

Donald E. Myers¹

Abstract.--In geostatistics simulation means generating another "reality", i.e., another realization of a random function. Usually the random function is only characterized in terms of the first two moments, the mean and a spatial correlation function such as the covariance or variogram. Interpolation methods allow predicting or estimating values at non data points of the (same?) realization. Simulation algorithms generate additional realizations, often conditioned to the original data.

Originally simulation was primarily of interest as a planning tool but now some are using simulation in lieu of interpolation. Interpolation methods smooth the data, but simulation algorithms enhance the variation. In some applications the smoothing produces results that are thought to be less realistic than simulation.

There are multiple simulation algorithms available, the user must choose between them. Simulated results are often used as an input for other modeling steps. There are both practical and theoretical aspects to the problem of choosing and using a simulation algorithm. Multiple simulation algorithms will be briefly reviewed and properties discussed along with typical applications. Practical and theoretical aspects of equivalence will be discussed including the sense in which properties of the random function are reproduced

INTRODUCTION

As the term is commonly used in statistics and probability, simulation means generating a set of numbers using a given probability distribution, standard limit theorems ensure that the empirical distribution (and its characteristics) will approximate those of the original distribution. This easily extends to vector valued random variables and a joint distribution. In either case the distribution function is an essential tool. The other essential tool is a random number generator for the Uniform Distribution on an interval.

In geostatistics the underlying model is a random function defined in 1,2 or 3 dimensional space,i.e., a random variable whose values are functions rather than

¹*Professor of Mathematics, University of Arizona, Tucson, AZ 85721*

numbers. Simulation means generating an alternative "reality", i.e., another one of the functions hence preserving certain characteristics of the random function. Since simulation is a numerical technique the realization is only known at a finite number of points. In practice it is more like the problem of simulating a vector valued random variable except that the joint distribution function is seldom known. This has led to the development of a number of algorithms with differing theoretical and practical characteristics. A user must choose between algorithms and the choice of an algorithm is driven in part by the ultimate objective, i.e., the intended use of the simulated values and over time these have changed.

OBJECTIVES AND MODELS

Early applications of geostatistics were principally in mining and hydrology, objectives for simulation were mainly related to these. In mining simulation was proposed as a planning tool, Journel (1974). In hydrology simulated values of hydrological parameters such as hydraulic conductivity are used as an input for flow models. Simulation can be used to design sampling plans by generating multiple realizations, optimizing the plan for each and then optimizing between plans. Because interpolation smoothes the data, interpolated surfaces are sometimes thought to be unrealistic and simulation is used instead, Dalle-Gupta et al (1995). Simulated data has also been used to study the behavior of various variogram estimators. Consider a random function model

$$Z(x) = m(x) + Y(x) \quad [1]$$

where $Y(x)$ is a zero mean random function with some form of stationarity assumed, i.e., second order or intrinsic of order k , and $m(x)$ is the expected value of $Z(x)$, usually assumed to be a linear combination of known basis functions such as monomials in the position coordinates. $Y(x)$ is characterized by a spatial structure function such as a covariance, variogram or generalized covariance. In addition $Y(x)$ is usually assumed to have a known univariate stationary distribution.

A QUICK OVERVIEW OF COMMON ALGORITHMS

Turning Bands

Let $X(t)$ be a random function defined in 1 space, s a random unit vector in n -space and x a point in n -space then the scalar product $\langle x, s \rangle$ is the projection of the point x onto the direction s . By integrating with respect to the random direction s , one obtains a random function $Z(x)$ defined in n -space. This provides a relationship between the covariance function $C_1(r)$ for $X(t)$ and the covariance $C_n(r)$ of $Z(x)$. Given a covariance for $n=3$ one can easily solve for the corresponding covariance in 1-space. Then by writing the 1-dimensional covariance as a convolution one

obtains a function that can be used in a Box-Jenkins, i.e., moving average, simulation algorithm (in 1-space). To discretize the algorithm the user must make choices, first, one can not use an infinite number of directions. They should be equally spaced and in 3-dimensions geometry intervenes to show that 15 is maximum number. Theoretically the Moving Average simulation is an integral and must be approximated by an infinite series but only a finite number of terms are used. Instead of simulating a value at each point on a line, one simulates a value for each of a set of disjoint intervals. When the lines are rotated in different directions we obtain the "turning bands". Given a point x then one determines the interval on each direction that the point projects into and looks up the simulated value for that interval. The simulated value for the point x is then simply a sum of those values.

While the integral equation linking the 1-dimensional covariance and the n -dimensional covariance is easily solved for $n=3$, it is not so simple for $n=2$. This led to the use of spectral methods, see Mantoglou and Wilson (1982) for applications in hydrology and Booker (1985) for solutions of the integral equation when $n=2$. There is another important difference between the cases for $n=2$ and $n=3$. For any even integer it is easy to choose equally spaced directions in 2-space (equally spaced points on the unit circle). There is no upper bound on the number of equally spaced directions.

The Turning Bands algorithm will reproduce the covariance function in an "average" sense, i.e., if one generates multiple realizations then averages the sample covariance functions (or sample variograms) the average will approximate the original covariance (variogram).

L-U Decomposition

Given locations x_1, \dots, x_m the covariance matrix of the random variables $Z(x_1), \dots, Z(x_m)$ is positive definite and may be "factored" in at least two ways. The Cholesky decomposition produces two matrices L , U (which are respectively lower triangular and upper triangular) such that $L^T = U$. Now suppose that Y_1, \dots, Y_m are uncorrelated, mean zero, variance one random variables. The product UY is a collection of random variables whose covariance matrix is the original. Although the technique was well-known in statistics, the application and extension to spatial data is found in Davis (1987) where it is also shown that the conditioning is easily incorporated. There are at least two difficulties; most algorithms for finding the L-U decomposition require that the entire matrix be in memory at one time and this limits the size of the grid to about 1000 points. Alabert (1987) has given a modified form of the algorithm allowing the use of a moving neighborhood for the conditioning and hence the use of multiple covariance matrices instead of one large matrix. Since the matrix "slides" over the grid one must still consider the possible discrepancies that might be generated by having a grid location occur in multiple

matrices.

Although it does not completely solve the problem, Davis (1987) also proposed the use of the square root of the covariance matrix which is then approximated by a polynomial expansion which only requires computing powers of the matrix also see Dietrich and Newsam (1995). Myers (1989) extended both of these algorithms to the vector case including the undersampled case for conditioning.

One advantage of the L-U or square root algorithms is that they are dimension independent but both suffer the same difficulty with respect to the univariate distribution as does the Turning Bands method, namely that the simulated values are obtained as linear combinations of uncorrelated random variables. Hence a transformation to normality is needed along with a reverse transformation. There is a parallel processing version of the Cholesky algorithm which would allow the use of larger covariance matrices, i.e., larger grids, but it seems not to have been used in geostatistical applications as yet.

Sequential Gaussian

It is well-known that in the case of multivariate normality the Simple Kriging estimator is the conditional mean, i.e., the mean of the conditional distribution, and the Kriging variance is the conditional variance. The Sequential Gaussian simulation algorithm is based on this property. Given a set of data locations and a grid point to be simulated, one computes the Simple Kriging estimated value for the grid point and the kriging variance. These are used in the conditional normal distribution which is then "sampled". Having generated a value at the grid point, this location and value are considered to be a part of the data set and process is repeated at a new grid point. Since it is straightforward to use a moving search neighborhood for the simple kriging it is not necessary to use the entire "data" set for each grid point but rather only those that are close. The user must then select the parameters for the search neighborhood and also select the order in which grid points are simulated since points selected earlier may be used in the subsequent simulations.

Since normality is explicitly used in the algorithm it is then necessary to transform the data and to re-transform the simulated values. As with the Turning Bands and L-U decomposition methods, the covariance used pertains to and is modeled from the transformed data. In an average sense it can be shown that any randomly selected sequence of grid points produces the same set of realizations.

Simulated Annealing

The title is somewhat mis-leading since Simulated Annealing is really an optimization method. A realization is generated by minimizing the "difference"

between the sample variogram (or sample covariance) and the theoretical model. First the data is used to model a probability distribution, then using this distribution and a random number generator, a value is simulated at each grid location (independently). Although this will ensure that the univariate distribution is preserved it would in general result in a pure nugget effect variogram. Next select at random two locations, interchange the values at these two points and compute the sample variogram. If the new sample variogram is closer the interchange is retained, otherwise not. New locations are selected and the process repeated. A "temperature" function is used to ensure that the iteration stops. For a brief introduction see Deutsch(1995), Deutsch and Cockerham (1994).

VARIATIONS AND EXTENSIONS

The four methods described above essentially assume that the random function is continuous valued. At least two variations occur on this theme, coded data obtained by for example an indicator transform applied to continuous data, and categorical data. One might simulate indicator transform data for reasons that are essentially the same as those related to the use of indicator kriging as opposed to kriging of the original variable. Journel and Isaaks (1984) use an indicator transform to separate the data corresponding to a high grade mineralization from that corresponding to a low grade mineralization. Bierkens and Weerts (1994) apply indicator coding to categorical data, e.g., soil texture classes. In this case it is necessary to assign an (arbitrary) ordering to the texture classes. The use of indicators essentially corresponds to assigning a probability distribution to each point, i.e., one is simulating a probability distribution at each point. This distribution is conditioned on the "data". As in indicator or probability kriging, if the values are kriged or simulated for each cut-off separately then the order relations may not be satisfied and adjustments are needed.

SOME QUESTIONS

Usually one desires that the univariate marginal distribution be preserved and this also creates a problem. First of all one must distinguish between a theoretical distribution for the random function model and the spatial frequency distribution of a particular realization although the latter is what is usually computed. Since the simulated value at any point on the grid is obtained as a sum of independent, i.e., uncorrelated, random numbers one must expect a form of the Central Limit Theorem effect. Actually one need not resort to the Central Limit Theorem, there are only a few distribution types for which the sum of independent random variables has the same distribution type as the summands. These include the Normal and the Gamma, hence it is common to utilize a transformation to transform the original data frequency distribution to Normal and then to re-transform the simulated values back. Since the original data set is finite, the empirical distribution will be discrete and

hence some smoothing is needed to transform to Normal, the transformation is not uniquely determined but requires some user choices. In addition the covariance used in the simulation step reflects the spatial structure of the transformed data and in general there is no assurance that the spatial structure of the original data is preserved in the process. A desire to simplify the algorithms gave rise to an emphasis on the LU decomposition and sequential gaussian methods both of which use normality more explicitly. As the need for application to non-continuous variables and also for simulation on much larger grids increased, other algorithms were introduced. The choice between the algorithms was and is largely driven by personal preference or practical aspects of the application. Underlying this diversity of algorithms was an implicit but never stated assumption that there was some form of equivalence and hence the difference was only computational. This might be interpreted as meaning that if one considers the set of all possible realizations that could be generated (for a given set of parameter assumptions) was the same for all methods or at least approximately so. Since in practice one only generates a small number of realizations, i.e., the values at a finite number of points for each of a small number of realizations there is another implicit assumption, that for the particular purpose or objective the conclusions were essentially the same irrespective of which small set of realizations is used. Neither of these implicit assumptions has really been tested or even considered, most users do not use multiple algorithms and make comparisons nor do they generate multiple finite sets of realizations to compare between the sets.

Characterizing or defining "equivalence" is non-trivial. One very strong perception would be that if one considers the totality of possible realizations from one algorithms vs another then the two are equivalent if the two sets are the same. Unfortunately the algorithms in general only depend on second moment properties and hence do not uniquely determine a random function. One might also consider equivalence in a numerical sense, two algorithms are equivalent if for any finite set of grid points, the same sets of values are generated by both algorithms. While equivalence is principally a theoretical question, if two algorithms are not equivalent then one must ask how to decide between algorithms and to what extent is the validity of the conclusions related to the choice of the algorithm.

All of the simulation techniques depend in a critical way on the use of a random number generators. It is well-known that such numbers are in fact not really random but by knowing the "seed" the sequence is completely predictable. However there are more serious problems, there are many different random generator algorithms and they are not equivalent nor do they have the same characteristics. Some are highly dependent on hardware and all depend on the word length in memory. A highly touted "portable" random number generator has recently been shown not to be machine independent as earlier believed. As yet there has been little study of the effect of different random number generators on the various simulation algorithms.

REFERENCES

- Alabert, F. 1987. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Math. Geology* 19, 369-386
- Bierkens, M.A.P., and Weerts, H.J.T., 1994, Application of indicator simulation to modelling the lithological properties of a complex confining layer. *Geoderma* 62, 265-284
- Booker, P.I. 1985. Two dimensional simulation by Turning Bands. *Math. Geology* 17, 81-90
- Dalle-Gupta, A., Lake, L. and Pope, G.A., 1995, Characterizing heterogeneous permeable media with spatial statistics and tracer data using sequential simulated annealing. *Math. Geology* 27, 763-788
- Davis, M.W. 1987. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Math. Geology* 19, 91-98
- Davis, M.W. 1987. Generating large stochastic simulations-The matrix polynomial approximation approach. *Math. Geology* 19, 99-108
- Dietrich, C.R. and Newsam, G.N., 1995, Efficient generation of conditional simulations by Chebychev matrix polynomial approximations to the symmetric square root of the covariance matrix. *Math. Geology* 27, 207-228
- Dimitrakopoulos, R. 1990. Conditional simulation of intrinsic random functions of order k. *Math. Geology* 361-380
- Deutsch, C.V., 1995, Introduction to the application of simulated annealing for geostatistical applications. *GEOSTATISTICS Newsletter* 7, 8-14
- Deutsch, C.V. and Cockerham, P.W., 1994, Practical considerations in the application of simulated annealing to stochastic simulation. *Math. Geology* 26, 67-82
- Easley, D. H., Borgman, L.E. and Weber, D., 1991, Monitoring well placement using conditional simulation of hydraulic head. *Math. Geology* 23, 1059-1080
- Gutjahr, A., Bullard, B. and Hatch, S., Joint conditional simulations and flow modeling. in *Geostatistics for the Next Century*, R. Dimitrakopoulos (ed), Kluwer Academic Pub. Dordrecht 185-196
- Journel, A.G. 1974 Geostatistics for conditional simulation of ore bodies. *Economic Geology* 69
- Journel, A.G. and Isaaks, E.H. 1984. Conditional indicator simulation: Application to a Saskatchewan uranium deposit. *Math Geology* 16, 685-718
- Mantoglou, A. and Wilson, J.L. 1982. The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resources Res.* 18, 1379-1384
- Myers, D.E., 1989, Vector conditional simulation. in *Geostatistics*, M. Armstrong (ed), Kluwer Academic Publishers, Dordrecht, 283-292