SPATIAL VS ENSEMBLE STATISTICS FOR SPATIAL SIMULATION ALGORITHMS

DONALD E. MYERS
Department of Mathematics
University of Arizona
Tucson, AZ 85721

ABSTRACT

Spatial simulation algorithms are designed to generate realizations of random functions and to reproduce statistical characteristics of the random function. In practice only a small number of realizations are generated and hence the characteristics of the realizations are more important than those of the random function. Most algorithms reproduce the statistical characteristics, such as the covariance function, mean and univariate distribution, only in an average sense. At least second order stationarity is an implicit assumption in using the algorithms. Empirical and theoretical results for convergence are given for several of the commonly used algorithms such as LU, Sequential Gaussian and Turning Bands. Simulated Annealing is an algorithm where each realization reproduces the statistical properties with large probability.

1.INTRODUCTION

The interpolation of irregularly spaced data or the estimation of spatial averages from irregularly spaced data is of considerable importance in many areas of application, Myers (1996). But all interpolation algorithms smooth the data in some sense. For example, if one considers the empirical distribution of the data, the empirical distribution of the interpolated values is generally more compact, in particular the variance is decreased. Moreover the spatial correlation is generally decreased, i.e., the interpolated function is more continuous than the empirical function determined by the data. These are not always desirable characteristics. In geostatistics the spatial data is usually considered to be a non-random sample from one realization of a random function. Interpolation corresponds to "filling in " the rest of the realization (this will likely be non-unique but only one realization will be generated). Simulation , in contrast corresponds to generating multiple realizations of the random function. Since the random function is not completely known, i.e., there is insufficient information to completely characterize the random function one can only require that certain properties be reproduced. There are three that usually are considered. First the spatial correlation should be preserved (in some sense), the univariate distribution should be preserved and the new realization should match the data values at the data locations. Two of these, preservation of the spatial correlation and the univariate distribution have two interpretations. One is theoretical and one is empirical. We discuss the differences for several commonly used algorithms and provide some numerical results to illustrate these differences. Because there are differences in the algorithms the question of equivalence is an important one and is discussed in Myers (1994). We begin by a brief review of spatial simulation algorithms.

2. WHAT ARE SPATIAL SIMULATION ALGORITHMS?

Let $Z(x)$ denote a random function where x is a point in 1, 2 or 3 space. Both for the applications and for the derivation of the algorithms it is necessary to assume some form of stationarity. This may be described in terms of moments or in terms of distribution functions. The most common

assumption is second order stationarity, i.e.,

(i) E[Z(x)] exists and is constant for all x in the region of interest

(ii) Cov[Z(x+h),Z(x)] exists for all x, x+h in the region of interest and depends only on h

The second condition then implies that the variance is constant. These two conditions do not imply anything about the univariate distributions or the finite dimensional distributions. For some of the algorithms however there is an additional implicit assumption about the (constant) univariate distribution type and in particular it is common to assume that $P[Z(x) \leq z]$ depends on z and is Gaussian. This may require a transformation (and likely then a back transformation).

Alternatively it might be assumed that for any finite number of points u, v,..., w and vector h that

$$P[Z(u) \leq z_1,....,Z(w) \leq z_m] =$$

$$P[Z(u+h) \leq z_1,....,Z(w+h) \leq z_m]$$

If the first two moments exist then they are constant and the random function will be second order stationary. This stronger assumption is often used in connection with the assumption that these joint distributions are either Gaussian or LogGaussian.

It is important to note that one does not generate an entire realization , nor even the realization restricted to a region in space. The algorithms will only generate the realizations at a finite number of points. While not essential these are usually taken to be on a regular grid. Some algorithms are more restrictive than others to the extent that they place limitations on the number of points on the grid. Simulation is widely used in both hydrology and petroleum to generate inputs for flow models. This may result in scaling problems.

In the geostatistical literature it is common to use the term "conditioning" to refer to the requirement that each generated realization match the data values at data locations but we see that it might be interpreted in a broader sense. In many applications only data is known, i.e., there are no state equations and hence the data is used to (a) model the univariate distribution , (b) estimate/model the spatial correlation structure function and finally (c), to determine the values of the generated realization at data locations. Note that this last condition may only be implicit, i.e., the data locations need not be points on the grid where simulated values are generated.

3. SPATIAL vs ENSEMBLE

There are two ways to determine whether the realizations are "conditioned" to the data. One is by the construction of the algorithm, this will mean that characteristics are preserved in an average sense, averaged over all possible realizations. The second is to use empirical statistics of the separate realizations or of a set of realizations. Since the total number of simulated values will be the product of the number of grid points times the number of realizations generated, this will likely limit either grid sizes or the number of realizations. A further question is whether the characteristics should be preserved in both fashions or only in one, in the latter case which one? This distinction is easier to understand by considering several of the common algorithms.

Although not considered in this discussion, the Simulated Annealing algorithm has some advantages if reproduction of statistical characteristics for each realization is a desirable attribute, see Deutsch (1995) and Deutsch and Cockerham (1994).

4. THREE ALGORITHMS

A. TURNING BANDS

Consider first the problem of matching the realization to the data values at data locations. Write the simulated random function in the form

$$Z_{sc}(x) = Z_K(x) + [Z_{sc}(x) - Z_K(x)] \quad (1)$$

If $Z_K(x)$ is an exact interpolator, e.g., simple or ordinary kriging, and $[Z_{sc}(x) - Z_K(x)]$ is a mean zero random function whose spatial correlation function is the prescribed one then $Z_{sc}(x)$ is a random function with the desired characteristics. This decomposition is analogous to using a regression equation for prediction. While the choice of algorithm for simulating $[Z_{sc}(x) - Z_K(x)]$ is not crucial in this decomposition, some algorithms allow incorporating the conditioning into the algorithm. One of the earliest algorithms was known as Turning Bands, Journel (1974), strictly speaking it is not a simulation algorithm but rather it is an algorithm for generating simulations in k-space by generating multiple simulations in 1-space. Let X(t) be a random function defined in 1-space and set

$$W(x) = \int X(\langle x,s \rangle) dP(s) \quad (2)$$

where dP(s) is a probability measure on the sphere, $\langle x,s \rangle$ is the inner product of x, s. It is easy to see that the covariance function of W(x) is given as a convolution of the covariance of X(t), for k=3 this is particularly easy to invert. The case of k=2 is discussed in Booker (1985), Mantoglou and Wilson (1982. To obtain the algorithm this representation must be discretized. Let $s_1,...,s_m$ are equally spaced directions and set

$$W(x) \approx \sum_{[i = 1,...,m]} a_i X(\langle s_i, x \rangle) \quad (3)$$

where the a_i's are of the form $1/m^{0.5}$. X(t) can simulated using a Box-Jenkins Moving Average.

There are certain properties or characteristics that can be seen imediately, some of which are a consequence of the discretization. First of all, the integral is replaced by a finite sum but the relationship between the covariance of W(x) and the covariance of X(t) is obtained from the integ relationship, some adjustments are necessary ensure that the variance comes out rig Particularly when k=3, geometry intervenes places some restrictions on the values of m, i.e., maximum number of equally spaced directio This restriction does not occur when k= Ultimately the simulated value of W(x) (for a fi x) is a linear combination of uncorrelated rand numbers. The Central Limit Theorem implies t at least in a limit sense that the distribution, ag for a fixed x, should be approximately Gaussi Hence it may be desirable that W(x) has a Gauss distribution then the random numbers are dra from a Gaussian distribution. Note that t distributional property will be preserved in theoretical or ensemble sense. It does not imp anything about the empirical spatial distribution any one realization. Those distributions are clea affected by the pattern of locations on the grid, t number of grid locations, the mesh of the gr Finally because the algorithm is both cpu a memory intensive, "bands" are used and these w produce some effects graphically.

B.SEQUENTIAL GAUSSIAN

This algorithm is based on several we known properties of the multivariate Gaussi distribution and the Simple Kriging estimat Recall that the Simple kriging estimator is of t form

$$Z^*_{SK}(x) = m + \sum_{[i = 1,...,n]} \lambda_i [Z(x_i) - m] \quad (4)$$

$$= m + (Z-m)C^{-1}C_x$$

where

$$\sum_{[i = 1,...,n]} \lambda_i C(x_i - x_j) = C(x - x_j) \quad (5)$$

$$j = 1,..., n$$

and the Kriging variance is given by

$$\sigma^2_{SK} = C(0) - C_x{}^T C^{-1} C_x \qquad (6)$$

The second form of eq (4) is of course just the conditional expectation and and eq (6) is just the conditional variance. At a non-data location then one simulates a Gaussian distribution whose mean is given by eq (4) and whose variance is given by eq (6). The $Z(x_i)$'s are the data values. Since the Simple Kriging estimator is exact the desired conditioning is incorporated into the algorithm and need not be done separately. The analogy with using a regression equation for prediction is even stronger in this case. Note however that there is an implicit assumption of a multivariate Gaussian distribution. In practice the algorithm will sequentially traverse the set of grid points simulating each in turn and treating previously simulated values as "data" for the subsequent conditioning. Note again that the preservation of the univariate distribution and the spatial correlation is in a theoretical sense. The algorithm does not ensure that the empirical properties of any one realization match those of the model. While one repeatedly uses a Gaussian distribution to generate the random numbers, the mean and the variance are changing from one grid point to another hence viewed as a data set the generated values need not be Gaussian. This algorithm has both advantages and disadvantages. First of all, it is dimension independent and secondly the sequential nature of the algorithm reduces memory demands. The user or the programmer must make choices however that will affect the results. While on the average, the particular sequencing of the grid points will have no effect on the simulations the actual application may. It would be possible to use a unique neighborhood for the Simple Kriging but in practice a moving neighborhood is used instead, the parameters of this neighborhood can significantly affect the results, i.e., the empirical results. As noted above a multivariate Gaussian distribution is implicit in the algorithm and hence transformations may be necessary. Univariate transformations are easy to construct but multivariate transformations that would be consistent across the number of variables would be much more difficult. A variation on Sequential Gaussian uses the Fourier Transform to remove the spatial dependence, for a discussion of this approach see Easley, Borgman and Weber (1991).

C. L-U DECOMPOSITION

Let $C_{ij} = \mathrm{Cov}[Z(x_i), Z(x_j)]$ and set

$$\underline{C} = \begin{matrix} C_{11} & \cdots & C_{1n} \\ \cdots & \cdots & \cdots \\ C_{n1} & \cdots & C_{nn} \end{matrix}$$

and $\underline{C} = LU$, the Cholesky decomposition of $\underline{C}$. The let Y be a vector of uncorrelated random numbers, YU will be a vector of correlated random numbers. The covariance matrix will be reproduced, in a theoretical/ensemble sense. Again, if the random numbers are Gaussian then the resulting distributions will be Gaussian, for the random function. Conditioning by the use of Simple Kriging is easily incorporated by partitioning the matrix $\underline{C}$ into blocks, one being the covariances between data locations, one being the covariances between grid locations and the other two being covariances between pairs of locations one from each set. Because the entire set of grid points is simulated at one time the size of the matrix $\underline{C}$ is a restriction. For ordinary computers, 1000 x 1000 is an upper bound, Davis (1987a). The square root of $\underline{C}$ can also be used and this can be approximated by an orthogonal polynomial, Davis (1987b). Dietrich and Newsam (1995). As pointed out in Myers (1989) the univariate distributions can not be the same in these two cases even though the same distribution is used for the random number generator. This algorithm is easily extended to the vector valued case as shown in Myers (1989). The use of Ordinary Kriging in

lieu of Simple Kriging is not easily incorporated.

4. NUMERICAL RESULTS

Numerical results were obtained using a software package called ISATIS, which was furnished to the author courtesy of GEOMATH. The data is a topography i.e., elevation data set that comes with the software. Figure 1 shows the pattern of data locations, Figure 2 shows a plot of data values versus the east-west coordinate and Figure 3 a plot of data values vs the north-south coordinate. Figure 2 was interpreted as indicating a non-stationarity. A linear trend surface was fitted to the data and residuals computed. Figure 4 shows the histogram of the data values and Figure 5 the histogram of the residuals. Figures 6 and 7 show the sample variograms (all directional) for the original data and the residuals. Only the residuals were used for later computations. Three grids were generated overlaying the region in Figure 1; a 5 x 5 grid, For the Turning Bands simulations, 15 directions were used. For the conditioning, i.e., the kriging, an isotropic variogram was used with a search neighborhood radius of 200 m, requiring a minimum of one data location inside a neighborhood and using a maximum of 10. Multiple realizations were generated on each grid using the same variogram model and using all three algorithms. The multiple realizations were obtained by updating the starting seed for the random number generator.

Statistics

Multiple statistics were computed for each realization or each set of realizations. First a (spatial) histogram was computed to compare with the distribution type stipulated in the algorithm, for fixed grid size histograms were compared between realizations to determine the extent to which the histogram shape remains constant. Although only a small number of realizations were generated, histograms across realizations can be computed for each grid point again to compare with the stipulated model and also to verify stationarity in the algorithm. There was a marked difference between algorithms when reviewing the histograms for separate realizations, the histograms for the Turning Bands simulations were much more compact than for the Sequential Gaussian.

Sample (spatial) variograms were computed for each realization for each grid size and as might be expect some effect of grid mesh is seen both because of the discrepancy in the numbe of grid points and also in the minimum interlocation distances. This relates in part to the relationship between the range of the variogram and the grid mesh as well as the radius of the search neighborhood used in kriging. The sample variograms for the Turning Bands simulations were less like the stipulated model than for the Sequential Gaussian simulations.

Another way of comparing realizations is to compute sample cross-variograms for a pair of realizations. This provides a characterization of how two realizations differ "at a distance". The cross-variograms for pairs of Sequential Gaussian simulated realizations had a more pronounced sill than for the those generated using the Turning Bands Algorithm.

5. SUMMARY

The numerical results confirm what might be expected, namely that the statistical properties of separate realizations can differ significantly from those of the model used in the algorithm. In addition the degree of reproduction of these statistical properties varies from one algorithm to another. The distinction between spatial and ensemble statistics for simulations is one of the reasons for the interest in another algorithm (not implemented in ISATIS)

called Simulated Annealing. In this algorithm both the univariate empirical distribution and the spatial correlation, e.g., the sample variogram, are forced to match or be very close to the model. The algorithm involves an iterative process to achieve this.

REFERENCES

Alabert, F. (1987), "The practice of fast conditional simulations through the LU decomposition of the covariance matrix," **Math. Geology** Vol. 19, pp. 369-386

Booker, P.I. (1985), "Two dimensional simulation by Turning Bands". **Math. Geology** Vol. 17, pp. 81-90

Davis, M.W. (1987a), " Production of conditional simulations via the LU triangular decomposition of the covariance matrix," **Math. Geology** Vol. 19, pp 91-98

Davis, M.W. (1987b), "Generating large stochastic simulations-The matrix polynomial approximation approach," **Math. Geology** Vol 19, pp. 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** Vol.27, pp. 207-228

Deutsch, C.V. (1995), "Introduction to the application of simulated annealing for geostatistical applications," **GEOSTATISTICS Newsletter** Vol. 7, pp. 8-14

Deutsch, C.V. and Cockerham, P.W., (1994), "Practical considerations in the application of simulated annealing to stochastic simulation," **Math. Geology** Vol. 26, pp. 67-82

Easley, D. H., Borgman, L.E. and Weber, D.(1991), "Monitoring well placement using conditional simulation of hydraulic head," **Math. Geology** Vol. 23, pp. 1059-1080

Journel, A.G. (1974), " Geostatistics for conditional simulation of ore bodies,". **Economic Geology** Vol 69

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.** Vol. 18, pp. 1379-1384

Myers, D.E., (1989), "Vector conditional simulation" in **Geostatistics**, M. Armstrong (ed), Kluwer Academic Publishers, Dordrecht, 283-292

Myers,D.E. (1994), "Geostatistical Simulation: Thoughts and Questions," **GEOSTATISTICS Newsletter**, Vol 7,

Myers, D.E., (1996), "Choosing and Using Simulation Algorithms," in **Proceedings of Spatial Accuracy Assessment in Natural Resources and the Environmental Sciences** (second international symposium, Ft. Collins, CO, 21-23 May 1996), H. Todd Mowrer, R.L. Czaplewski and R.H. Hamre (eds), USDA, Technical Report RM-GTR-277, pp. 23-29