Geostatistical Simulation : Thoughts and Questions

Donald E. Myers
Department of Mathematics
University of Arizona
Tucson, AZ 85721

Introduction

There can hardly be any doubt about the importance and usefulness of simulation as a tool in geostatistics, particularly conditional simulation. Beginning with the early work of Journel which popularized its use in mine planning and his later work in applying it to reservoir modeling, the extensive activity of Gelhar's group at MIT in applying simulation to problems in hydrology as well as many others who have extended the use of simulation including that of Gotway, Zimmerman and Zimmerman for evaluating variogram estimators, it has become as well known as the various kriging methods. However there may be a lack of clarity about what is meant by conditional simulation at least in some practical senses. Most, if not all, of geostatistics is based on the use of a random function model. The random function may be thought of as a random variable whose "values" are functions, i.e., realizations. In real applications, neither the random function nor even one realization is completely known and the data is considered to be a non–random sample from one realization of the random function although it is not known which realization. Even a countably infinite sample would be insufficient to uniquely characterize the realization. Fortunately kriging (in its various forms) is invariant with respect to the choice of the realization, as shown by Matheron. The disadvantage of kriging is that it smoothes the data whereas simulation is intended to retain and exhibit the spatial variability.

Conditional simulation then is a process for generating the remainder (or more precisely, a part of the remainder) of one or more of the realizations of the random function from which the data could be considered a sample. To generate an entire realization it would be necessary to obtain at least an implicit analytic representation for that particular realization. In practice however the real objective is merely to generate the values of the realization at a finite number of points, usually taken to be on a grid. The "process" is itself realized by an algorithm and then the algorithm is implemented by a program, i.e., encoded in one of the standard programming languages and executed on a particular computer. Several ideas are implicit in the use of conditional simulation in geostatistics but they do not seem to receive much attention in the literature. One such idea is that of the "equivalence" of different algorithms and a second is the equivalence of different programs for the same algorithm and finally an insensitivity to hardware. Many of the early papers were devoted to showing that a "new" algorithm or a variation on a particular algorithm required less CPU time or less memory than another algorithm. Less CPU time or less memory was (and still is) considered a desirable attribute and hence one algorithm may be judged "better" by virtue of one or both of these characteristics. How should we compare algorithms and does it make any difference which algorithm we use? Do CPU time and memory requirements provide adequate criteria for comparing algorithms and programs?

Equivalence

While it is a characterization that is difficult to use in any practical sense, one way to characterize a random function is to consider the set of all possible realizations together with a probability distribution on those realizations. In geostatistics the realizations are considered to be "equally likely". Two random functions would be "identical" if they had the same set of realizations and the same probability distribution. It would be tempting to say that they are exactly the same but in the same sense that we can have identically distributed but distinct random variables we could have "identical" but distinct random functions. This suggests considering two algorithms equivalent if they generate the same set of realizations although not necessarily in the same order. Since it is not possible to generate all possible realizations using a particular algorithm nor even to generate all of one realization, are there practical ways to test for the equivalence of two algorithms? Does it matter if we use two non–equivalent algorithms? Are there some applications of conditional simulation for

h it doesn't matter and others for which it does? If how do we distinguish them? If two algorithms ere equivalent in the sense above but produced the realizations in reverse order how would we detect the equivalence if only a finite number of realizations are generated?

Invariance

All of the methods, e.g., algorithms, that have been proposed so far are based on preserving several properties of "the" random function. First, the "data" is honored. Secondly, the spatial correlation structure is preserved (which usually means that the second moment is preserved). Thirdly, the mean is preserved and finally it is usually considered desirable to preserve the marginal distribution. With perhaps the exception of sequential Gaussian simulation (SGS) none of the algorithms utilize nor provide any information about higher order moments nor any information about joint distributions. What do these three conditions actually imply? Honoring the data: This property is easily described. The values of the (simulated) realization at data locations are exactly the data values at the respective locations.Preservation of the spatial correlation structure: The standard algorithms (Turning Bands, L–U decomposition, sequential Gaussian) preserve the variogram or covariance on the average. The average is with respect to the set of all possible realizations. The sample variogram (or other estimator) is a form of a spatial average and it is well–known that its use implies a form of ergodicity. It is common in evaluating a proposed algorithm (or even in an application of a known method) to not only compute the sample variogram for each realization but also to show the average of the sample variograms across realizations. Because nearly all of the standard methods are relatively computer intensive in one way or another it is rare to generate more than a small number of realizations especially if the number of grid locations is large which it often is (small is a relative term and is steadily increasing as computational power increases). Without some form of multivariate distribution assumptions it is difficult to model the behavior of the sample variogram and hence to model the behavior of the average sample variogram as the number of realizations gets large. Davis and Borgman obtained both theoretical asymptotic and some practical results but they are difficult to use in any specific application. How rapidly does the average sample variogram (or other spatial structure function) converge to the "true" spatial structure function (the one used to generate the realizations)? How rapidly should it converge and how do the convergence rates compare for different algorithms? Should the convergence rate be used as a criterion for comparing algorithms? How do we distinguish between the properties of the variogram/covariance estimator and those of the simulation algorithm? Since only second moment properties are used to characterize the random function for which new realizations are to be generated, do different algorithms generate realizations of different random functions (which have the same second moment properties) or do they merely generate different realizations of the same random function? If the simulated values were to be used for kriging then the question of "different" random functions would be of no importance, for other applications it is not so obvious. As is well known, the sample variogram is not a particularly well– behaved estimator (even though it is theoretically unbiased). If simulated data is used to evaluate variogram estimators or the behavior of the sample variogram then there is an assumption that the simulation method is without flaws, conversely if the sample variogram (or other estimator) is used to evaluate the behavior of the simulation technique then there is an implicit assumption that the estimator is well–behaved. These assumptions can not be independently checked. Hence it is important to recognize that when simulated data is used to test some other algorithm that there are two sources of error or variability which will be difficult to separate. It might be better to utilize multiple simulation algorithms. Simulated Annealing has recently attracted attention as a conditional simulation method that is "better" than other methods. The algorithm depends on forcing the estimated spatial correlation function (computed from one realization) to be close to the theoretical. However Simulated Annealing was originally introduced as an optimization method, not a simulation method. While a random function could be such that for all realizations the estimated spatial correlation function would coincide with or very closely match the theoretical function, that is a severe restriction on the random function

it overlooks a crucial point. The user must still decide how "close" the sample variogram must be to the theoretical model, i.e., when is the algorithm stopped. How does this choice affect the generated realization and how much variability is there between realizations? The spatial correlation estimator is very much dependent on the data location pattern, in the case of a grid it would depend on the grid spacing and orientation. This is true independently of the method used to generate the realization. There will be an interaction between the range of the variogram/covariance and the grid mesh used in generating new realizations. For simulated annealing to be an appropriate simulation algorithm then its application should be independent of the grid mesh. This issue seems not to have been addressed as yet. Preservation of the marginal distribution: The Turning Bands and L–U methods both generate the simulated values as linear combinations of uncorrelated random numbers. As has been noted by many authors it is preferable to begin with a standard Normal marginal distribution since such a distribution is preserved and not distorted by the Central Limit Theorem effect. The marginal distribution that is to be preserved is actually a spatial distribution and the extent to which it reflects the marginal distribution of the random function depends on an assumption of (strong) stationarity. The marginal distribution(s) of the simulated realizations are likewise spatial distributions and again an assumption of stationarity is needed. Similar disparities and assumptions are involved when considering the invariance of the first moment. The L–U method depends on the factorization of the covariance matrix but the square root has the same property. However it has been shown that the marginal distribution can not be the same for both L–U and square root factorizations. In order to begin with a standard Normal marginal distribution, it is necessary to transform the data. In particular this means that the original distribution must be determined but a finite data set will not uniquely determine the distribution. Using the spatial distribution to determine the marginal distribution implies an assumption of stationarity. After simulation the simulated values must be re–transformed, this re–transformation is not uniquely determined nor can it be uniquely modeled. The inverse transformation may not be one–to–one.

Summary

Conditional simulation is an extremely useful tool in geostatistics but the relationship between different algorithms needs to be re–examined. In the past there seems to have been an implicit assumption that any two methods were equivalent (but without an explicit understanding of what equivalence meant) and hence only computational discrepancies were important. Comparisons should be based on more than discrepancies in computer time or the size of the grids that can be simulated.

A Probabilistic Model for a Partially Explored Mining Districts

Mark Springett
4540 Keel Court
Boulder, CO 80301–3113

Sometime ago I was asked to produce a rationalised estimate of the benefits of continuing exploration for gold on a property position where several significant individual deposits had been identified, drilled and evaluated. A larger resource base was needed to amortise the estimated capital expenditure and a viable estimate of the unknown was needed in a very short period of time.
I referred back to an amusing paper published in 1977 in " Mathematical Geology" (Rowlands et al. 1977) and concluded that their application of Zipf's Law was appropriate to my immediate problem. The two Australian authors had applied the Law (rule is more correct) to firstly the size and ranking of deposits in the Zambian Copperbelt and also to the Western Australian gold province with the objective of quantifying the number and size of as–yet undiscovered deposits within the partially explored districts.

Zipf's 'Law' is a limiting case of the better known Pareto Distribution and can be non–mathematically described as a series where the biggest is twice as big as second larges, three times as big as the third largest and so on with decreasing size. If we have a number of known deposits of reasonabley well defined size in our