Short Note:
Some Implementation Aspects of Multiple-Point Simulation

Steven Lyster[1], Clayton V. Deutsch[1], and Julián M. Ortiz[2]

[1]Department of Civil & Environmental Engineering
University of Alberta

[2]Departmento de Ingeniería de Minas
Universidad de Chile

Abstract

Variogram-based simulation techniques for modeling of geologic bodies cannot reproduce complex structures featuring curvilinearity or intricate relations between different facies. Multiple-point statistics attempt to better model these complicated bodies. Implementation of multiple-point statistics in simulation algorithms is difficult. This note discusses some issues commonly encountered when attempts are made at honouring multiple-point statistics.

Multiple-Point Statistics

Traditional two-point statistics such as the variogram cannot fully characterize complex geological shapes and structures. Kriging and stochastic simulation using variograms are satisfactory for many applications; however, curvilinearity and complex interactions between different facies are not well represented by variogram-based geostatistical methods. Multiple-point (MP) statistics attempt to better represent these complex features. MP statistics summarize the joint distribution of three or more points simultaneously.

The inference of MP statistics is difficult due to the large amount of information they contain. Often, MP statistics are inferred from a *training image* [1,5,8,9]; this is a realization developed from highly sampled areas, object-based modeling, outcrop mapping, or some other source that produces a high-resolution, exhaustive model at the same scale as the simulation. The training image (TI) is deemed fully representative of the area being simulated and must contain all relevant geological features. The use of a TI is not required for MP simulation; the statistics can come from any source. Nevertheless, a TI is the most convenient way of deriving the MP statistics as most desired statistics can be extracted directly with no need to fit them with positive definite models. This approach also ensures that the statistics used are consistent since at least one realization exists that honours them all [4] and from which consistent statistics of different dimensionality can be inferred or extracted.

One of the most common statistics used in this context is the MP histogram [3]. The different facies are assigned a value $k=1,\dots,K$. Any arrangement of N points can then be given a unique one-dimensional index j using the equation [4]:

$$j = 1 + \sum_{n=1}^{N} \big(k(n) - 1\big) \times K^{n-1} \qquad (1)$$

where k(n) is the category at the nth point in the MP configuration. This one-dimensional indexing results in K^N different classes that could be represented. The *N*-point histogram is quite robust in characterizing geological phenomena. Such a MP histogram contains all lower-order statistics within the same configuration, such as the *N*-1 and *N*-2 point histograms. While this is quite a powerful tool, the huge amount of information contained can cause problems for deriving the statistics as well as reproducing them.

MP statistics are often used to derive conditional probabilities from Bayes' relation:

$$Prob\{A_k = 1 \mid D = 1\} = \frac{Prob\{A_k = 1, D = 1\}}{Prob\{D = 1\}} \tag{2}$$

where $A_k=1$ denotes the occurrence of facies k at the point being simulated (point A) and D is the MP configuration around the point. Therefore $Prob\{A_k\}$ is the probability of facies k occurring at point A and the left side of equation (2) is the probability of facies k occurring at point A given configuration D.

Implementation of a robust method of using MP statistics in geostatistical simulation is quite difficult. Several different approaches have been proposed: single and extended normal equations (Guardiano and Srivastava, 1992 [5], Strebelle and Journel, 2000 [8] and 2001 [9]), a neural network iterative approach (Caers and Journel, 1998 [1]), and simulated annealing (Deutsch, 1992 [2]). Each of these methodologies has its own issues.

Implementation Aspects

All methods for reproducing MP statistics in a simulated realization must overcome certain obstacles. Some of the different facets of MP simulation that will be discussed here are: iterative approaches and their particular challenges, post-processing, utilization of multiple grids, and the problem of dimensionality.

Iterative Approaches

Neural networks (as presented by Caers and Journel, 1998 to store the information for MP simulation) and simulated annealing (as presented by Deutsch, 1992 to perturb values toward MP statistics reproduction) are both iterative approaches to reproducing MP statistics. Each of these methods visits a node of the simulated field, updates it based on some criteria, then moves on to the next node until all have been visited. Once every node has been visited the process is repeated until some stopping criterion is met.

The positive aspects of an iterative approach are flexibility to deal with many different MP statistics, lower-order statistics, secondary data and other complex data types. These methods are very robust in their implementation and honouring of statistics. The main drawback of these iterative approaches is the long amount of time often required for large realizations. These processes can be CPU intensive and there is no guarantee of satisfactory convergence to the target statistics. The decision of when to end the simulation is a difficult decision to make and can have an impact on the quality of simulated realizations and time for simulation.

Solving the problem of high time requirements could be quite simple; as computers are becoming increasingly fast it is becoming more feasible to use iterative approaches to MP simulation. Currently small grids with simple MP statistics do not take more than a few minutes to simulate. If CPU speed and RAM continue to increase at their present rate, within a few years it should be possible to produce large three-dimensional realizations of complex geologic structures in a reasonable time (a few hours).

The indexing system for MP histograms could be improved to allow faster simulations. An approach using search trees has already been implemented [8,9]; this will be discussed further in the "dimensionality" section. Other ways to improve speed of simulation for iterative approaches could be smarter perturbation mechanisms; better pseudorandom paths, perhaps visiting crucial nodes more or less often; and post-processing of previously-generated images.

Post-Processing

Iterative approaches can be used to take an initial image from another source such as sequential indicator simulation or a previous MP simulation, and post-process the image to better match the desired statistics. The process could be much faster than starting from a random image as some features are already contained in the initial image. Post-processing is a potentially powerful tool in that it allows realizations that already contain some information to be improved, cleaned and/or modified.

One downside of post-processing is that it adds complexity to the simulation. If the MP post-processing is allowed to change the initial image too much (for example, setting the initial temperature too high in simulated annealing), the original statistics honoured may be lost. However if not enough change is allowed (setting the initial temperature too low), the MP statistics may not be reproduced adequately. This could be overcome by an experienced user carefully setting the parameters for the algorithm used. The added complexity and user input must be balanced against the time gains and improved honouring of statistics, if any.

As an alternative simulation technique, one facies type at a time could be simulated then the image post-processed to add another and another as required. This sequential or stepwise approach could be used to simplify the MP statistics in any single simulation. Rather than check the relations between all facies, only one category vs. all others must be examined. While this is more complex for the overall simulation, each step has simpler statistics and should run faster.

Multiple Grids

Many MP simulation methods have difficulty capturing long-range continuity of geologic features. Large N-point pattern sizes for histograms or conditional probabilities are impossibly large or sparse, which leads to inference problems or artefacts in the simulated realizations. The short-range continuity is matched well.

A multiple grid approach is one technique to better represent longer-range connectivity. This involves simulating a coarse grid in the first pass (using corresponding widely-spaced MP statistics) then freezing these simulated points as conditioning data and simulating on a finer grid. This can be repeated as many times as needed to get to the appropriate density of nodes for the final grid. The main difficulty of this approach is that a very large training image is required to be able to infer the MP statistics for a coarsened grid.

The advantages of the multiple grid methodology are that it helps enforce long range connectivity and is easy to implement; simply repeat the simulation process with different statistics on each new grid. One disadvantage is that the multiple grid size must "fit" within the desired fine grid. Another is the difficulty in honouring conditioning data. If the conditioning data does not fall directly on one of the coarse nodes, then we must decide whether or not to move it to the nearest grid node despite the distance between them. There is also a chance of freezing the finer MP statistics by using coarse grids as conditioning for the subsequent steps in the simulation process. These considerations must be taken into account.

The question of how many grids to use is an issue that must be considered. Too many and the simulation will be unnecessarily slow. Also, the earlier grids may be very coarse and this could interfere with the later reproduction of fine structures. Too few grids will reduce the benefits of the multiple grid approach. The spacing of the grids must also be considered; optimal spacing of the grids is dependent on the range of the geologic structures being simulated, the available information from which to infer statistics, and how widely spaced the conditioning data to be honoured are.

Dimensionality

MP statistics potentially contain a huge amount of information and may be difficult to infer. A MP histogram with K=5 different facies and N=9 points would have almost 2 million classes. Inferring this histogram from a TI with only a few hundred thousand points could lead to many uninformed or poorly informed classes. Better methods of generating a TI for inference of statistics are a possible way to get around this problem. Simulation methods accounting for classes with zero occurrences as special situations could be another.

Storage of the huge number of classes is also an issue. Search trees [8,9] can be used to reduce the memory requirements for storage of the MP statistics and to improve speed. This and other improved storage methods for MP statistics need to be investigated further. Different methods for indexing the classes in a MP histogram could be investigated, individual occurrences of classes could be stored rather than frequencies or occurrences of each class, or simulation could proceed in a binary stepwise fashion as mentioned under "post-processing" to simplify the MP relations. Waiting for storage capacity of computers to increase is another possibility that may be considered in the long term.

The large number of classes that are represented can also cause problems when most of the simulated grid has been populated; it may not be possible to select a facies that correctly matches the MP statistics based on the combination of facies around the point being populated.

Simulated Annealing Aspects

Simulated annealing (SA) has its own particular challenges and strengths. The discussion that follows deals with selection of the objective function for MP SA, the convergence criterion and when to stop the simulation, and CPU time issues and possible ways to resolve this.

Selection of Objective Function

A typical objective function for SA involves the sum of squared differences between target and simulated statistics; for MP simulation this could be the sum of the squared differences of histogram frequencies, given by the equation:

$$O = \sum_{j=1}^{K^N} \left[f_j - f_j^* \right]^2 \tag{3}$$

where f_j is the frequency of class j in the TI and f_j^* is the frequency in the simulated realization. Alternatively, other MP statistics such as n-point connectivity functions [6] or MP covariances could be used.

More components could be added to the objective function to make the simulation more robust, such as a two-point variogram function for long-range connectivity or additional MP statistics. Weighting of these different components could be determined by an equation such as:

$$\omega_i = 1 / \overline{\Delta O} \tag{4}$$

where $\overline{\Delta O}$ is the average change in the objective function over some number of perturbations, perhaps 10000 or one for every simulated node. Which components for the objective function are needed or would be the most relevant is a decision that must be made. The upside is that many different objective functions could be included if this is deemed necessary. The weighting for the objective function components may need change throughout the simulation, as the average ΔO values will likely vary from beginning to end.

Including a covariance function or variogram as a component objective function requires some special consideration; these or any other two-point statistics must be omnidirectional so as to prevent honouring of the statistics in on or a few directions while not the others. This requirement may limit the range of the variogram function used.

Checking for Convergence

Due to the huge number of classes that may be present in MP simulation convergence of the objective function to some arbitrary low value is very unlikely to occur in a reasonable amount of time. Therefore, the number of times the maximum number of perturbations is reached before a temperature change should be the stopping criterion for SA when using MP statistics. The proportion of accepted perturbations should decrease as the simulation proceeds and the temperature is lowered. The acceptance dropping below a certain threshold, say 1-10%, could be used as an alternative stopping criterion. These approaches to the SA schedule would reduce the time required.

The question of when to end the simulation must balance time requirement and honouring of statistics. Longer simulations will probably match the desired objective function better, but there are diminishing returns when a simulation is carried on longer. Also, in the case of post-processing, if a SA algorithm is allowed to run longer it runs the risk of distorting any statistics that were honoured in the initial image.

CPU Time

SA is often a CPU-intensive process. In the case of MP simulation using a histogram there are N terms required to locally update the objective function after every perturbation. Updating the many classes of the histogram can be sluggish when the histogram is large.

Checking for acceptance against the Boltzmann probability distribution [2] could be computationally expensive. While checking once is almost instantaneous, calculating the probability of acceptance for a perturbation then checking it and repeating millions of times is quite slow.

Counting the total perturbations performed and checking for the conditions to lower the temperature are unnecessary if convergence is not being checked. When many perturbations for every grid node are required, eliminating any computational processes can have a positive affect on speed. Lean code is therefore much preferred in SA algorithms.

Other possibilities for reducing CPU time have already been discussed: Waiting for CPU speed to catch up to requirements; Improved storing methods for MP histograms; Post-processing; and Stepwise simulation of one facies at a time. Smarter perturbation mechanisms are another possibility, such as selecting a new facies for a perturbation based on which would lower the objective function the most rather than choosing randomly.

Discussion

The use of MP statistics in geostatistical simulation is an emerging area of research. The complexity of MP simulation makes further research in this area necessary if it is to someday supplement traditional two-point estimation methods such as kriging and stochastic simulation.

A variety of resource estimation methods are necessary for the many different circumstances that may be encountered in practice. Different deterministic and iterative approaches to MP simulation will likely have their own places even as different flavours of kriging and stochastic simulation are used today.

References

1. J. Caers and A.G. Journel. Stochastic Reservoir Simulation Using Neural Networks Trained on Outcrop Data. SPE 49026. In *Proceedings of SPE ATCE*, New Orleans, LA, October 1998.
2. C.V. Deutsch. *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data*. PhD thesis, Stanford University, Stanford, CA, 1992.
3. C.V. Deutsch and E. Gringarten. Accounting for multiple-point continuity in geostatistical modeling. In *6th International Geostatistics Congress*, Cape Town, South Africa, April 2000. Geostatistics Association of Southern Africa.
4. C.V. Deutsch. *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 2002.
5. F.B. Guardiano and R.M. Srivastava. Multivariate geostatistics: Beyond bivariate moments. In A. Soares, editor, *Geostatistics Troia '92*, volume 1, pages 133-144. Kluwer, 1993.
6. J.M. Ortiz. *Characterization of High Order Correlation for Enhanced Indicator Simulation*. PhD thesis, University of Alberta, Edmonton, AB, 2003.
7. J.M. Ortiz. Selected aspects of multiple-point statistics. In *Centre for Computational Geostatistics*, Volume 5, Edmonton, AB, 2003.
8. S. Strebelle and A.G. Journel. Sequential simulation drawing structures from training images. In *6th International Geostatistics Congress*, Cape Town, South Africa, April 2000. Geostatistics Association of Southern Africa.
9. S. Strebelle and A.G. Journel. Reservoir Modeling Using Multiple-Point Statistics. SPE 71324. In *Proceedings of SPE ATCE*, New Orleans, LA, October 2001.