Strategies for Modeling with Multiple-point Simulation Algorithms

Alexandre Boucher
Advanced Resources and Risk Technology, LLC. Sunnyvale, CA, USA
aboucher@ar2tech.com

Abstract
This talk discusses the main challenges and offers solutions when using multiple-point simulation algorithm for modeling real world reservoirs. Multiple-point simulation algorithms (mps) are a new family of geostatistical algorithms that aim at reproducing spatial patterns, such as connectivity, depicted in a training image. A training image contains the possible spatial configurations for any given geological object and relationships between objects; it replaces the traditional variogram model.

The paper addresses the definition of the training image and discusses the commonalities between the various mps algorithms. While mps algorithms are easy to use on small controlled data sets, in practice three problems are recurrent: (a) finding a training image, (b) presence of trends either in the training image or in the data and (c) large, non-repetitive objects/facies. Each of these difficulties and potential solutions will be illustrated with simple as well as complex examples.

Introduction
Multiple point simulation (mps) is a geostatistical simulation technique first developed at Stanford University in the past 10 years. Since then it had been successfully applied in the oil & gas industry to model a variety of reservoirs. An mps algorithm aims at reproducing spatial patterns, such as connectivity, that are depicted in a training image (TI). A training image contains the possible spatial configurations for any given geological object and relationships between objects.

The term multiple-point statistics is used in reference to the traditional variogram model used with kriging that only takes into consideration the average square difference between two points, i.e. a measure of two point statistics. The variogram (or equivalently the covariance function) characterizes spatial patterns between pairs of points and often fails at capturing important geological patterns such as connectivity and curvilinearity.

The mps algorithms aim at characterizing patterns using several points, typically between 20 and 100, instead of two, thus providing more realistic representation of geological patterns.
There are now many mps algorithms available e.g. SNESIM, FILTERSIM, IMPALA, Direct Sampling (Strebelle, 2002; Wu et al, 2008; Straubhaar et al, 2010; Mariethoz et al, 2010) that are currently being used in real-world applications. The goal of this paper is not to compare these algorithms but instead to take a larger view of the multiple-point formalism and offer strategies for their successful applications. The example in this paper will be using the search tree partitioning algorithm (Boucher, 2008) programmed in the SGeMS software but could have been done using others algorithms.

**Training image**

Traditionally, in kriging-based algorithm, such as the sequential Gaussian simulation (sgsim), the pattern selection is implicitly done when choosing a variogram model. Ideally, when the data are numerous enough, the variogram model can be directly inferred from measured data. The equivalence with mps algorithms is the selection of a training image. Note that direct inference of mps statistics from the data is currently not possible for reservoir (the reader is referred to Dimitrakopoulos (2010) for a promising solution to inferring mps statistics from hard data).

A training image is a repository of the patterns and their respective likelihoods. A frequent pattern appears more often in the training image than a rare one. The actual position of a pattern in the training image is (at first) irrelevant, what matters is its presence and its frequency.

To correctly choose a training image it is important to understand why the actual positions of patterns in the training image do not matter. The mps algorithms follow the sequential formalism; they simulate one point (sometimes a group of points) at a time conditional to previously simulated points and hard data present on the grid. At the beginning, the grid only contains the hard data and it is gets filled as the simulation progresses. At every uninformed grid node, the mps algorithm will extract the existing data on the grid (both hard and previously simulated) given a neighborhood (usually defined by geometrical constraints) and try to find some matching patterns in the training image. The data points found in the neighborhood is called a data event. The definition of a good match between the data event and patterns in the training images differs with each algorithm. Once matching patterns are found, the simulated values are chosen either by taking it from the best matching pattern or through a Monte Carlo simulation of the set of reasonably matching patterns. The mps algorithm searches the full training image and does not consider the locations of the matching patterns (it is possible to keep track of the patterns location and this is discussed in a later section).

Most disappointments when applying mps algorithms to real data set arise from a misunderstanding of the role of the training image. Initially, the TI was interpreted as an analog to the natural phenomenon. This dual interpretation is very convenient since analogs are rather easily available from different sources such as photography, remote sensing, physics-based simulation or expert drawing. For the modeler, it is often easier to build a graphical representation of the phenomenon as is, i.e. with the trends, instead of breaking it down into
repetitive patterns that are not location specific. Unfortunately, an analog type of TI will usually contains trends that will be lost once filtered by the mp algorithm.

In that regard the training image will always be visually misleading. Someone looking at a training image does not extract the same patterns as an mp algorithm does. The human observer sees a geological system and expects to get back a similar system that has been anchored on data. The algorithm sees a set of imbricated patterns and tries to set them together through a randomization process. Generating unconditional simulations (that is simulations without hard or soft data) is the best way for the human to share the perspective of the algorithm. The output from the unconditional simulations is the actual repository of patterns that the algorithm extracted from the training image. On a complex case this can be discouraging at first but should simply serve as a milestone in the modeling process. It is also useful at refining other parameters that serve to filter the training image such as the size and shape of the search neighborhood.

The rest of this paper will discuss two recurrent issues with solutions in applying mps to real world applications: (a) presence of trends either in the training image or in the data and (b) large, non-repetitive objects/facies. Most academic papers do not address these issues since they choose the training image that works well instead of having one dictated by the geological environment.

Trends

For a training image, trends refer to (a) a gradual variation in proportions of facies and (b) a gradual change in the patterns themselves (without necessarily a change in proportion). Few thick channels that evolve into numerous thinner one is an example of the second type of trends.

From the formalism described above recall that the mps algorithm filters out the actual position of the patterns when matching the data events, consequently trends in the training image are typically not be reproduced in the simulation grid. Fortunately, there are now several techniques aimed at integrating trends from the training image to the simulation grid.

A common approach is to use a probability field to control the location of simulated categories (Strebelle, 2002; Harding et al., 2004). The probability field approach uses a single complex TI but encourages some facies to be simulated at certain locations by providing a pre-defined high local probability of occurrence. The trends are codified into probability maps and the facies are forced into places at the simulation stage. This technique only takes into account trends in proportion not on the type of patterns. It requires creating the probability maps and then enforcing them with a probability integration techniques; a demanding task to ensure the reproduction of the trends. A major drawback is that tweaking the probability (often done with the tau-model or by increasing the probabilities close to zero or one) can distort other probability fields obtained through proper geophysical calibration.
Another approach is to divide the simulation grid into regions and to simulate each region with a different TI, see Wu et al. (2008) for an example. A TI can be rotated or scaled differently for different regions or different TIs can be used for different regions. A major issue with the region approach is that there is no guarantee that the different TIs are compatible with one another. This can manifest as discontinuities at the borders across regions since there is no model informing the transitional patterns between regions. From a modeler’s perspective the region approach is awkward. The modeler does not consider one “realistic” TI but a series of images that have the right patterns for each zone. These TIs may not be related to each other and may not easily refer to a unique geological concept.

A better approach, that most recent algorithms have embraced, is to use a realistic complex training image coupled with fields that model the trends. These fields can be (paleo) elevation, channel average thickness, gradient, distance from sedimentary sources, age, or any other information that is relevant to shape patterns in the training image. These trends functions must be defined on both the training image and on the simulation grid. When searching for matching patterns in the training image the algorithm must then weight both the trend functions and the usual data event. This technique (used in IMPALA, Direct Sampling, Search tree partitioning (Boucher, 2008), and Chugunova and Hu (2008)) is, in this author’s opinion, the best option at using mps for real world applications.

More importantly, it allows the geologist to choose a training image based on the geological environment instead of basing it on what an algorithm basic capabilities. The job of the modeler is then to build these trend functions, either from a geometric or geological perspective, in both the training image and simulation grid. It also does not interfere with probability fields derived from geophysical attributes.

Figure 1 shows a complex training image with multiple trends of proportions (e.g. some facies are only present in some part of the grid), of shapes (the orientation of the blue planes gradually varies in space) and of interactions (the orange dots never touch the blue planes and the red lines always touch the blue planes). The intensity follows the trend function (view A). For the simulations (shown in Figure 2), the first step is to create the trend function (top row). The trend function serves at finding the relevant patterns in the training image given the intensity of the trend function. Two views of three simulations are shown on rows one to four. Note the preservation of the angles, the proportions and the interactions between the facies. The simulations have been created with the search tree partitioning approach.
Figure 1 Training image with trend function. Note that (1) the orientation of the blue planes varies with the trend, (2) the proportion of the red lines is higher in the middle, (3) the orange dot increases in proportion with the trend, (4) the green ellipsoid are only present on the right hand side of the image and (5) only the blue planes are present on the left hand side. The background facies is omitted for better viewing.

Figure 2 Application of the training image and trend function of Figure 1 to a simulation grid. Note that this grid is twice the size of the training image grid in Figure 1.
Non-repetitive training images

The second problematic case is when the geological environment is made of massive solid shapes that do not repeat. Take for instance a single carbonate mound or a channel complex. The location of the mound or the channel complex is at least partially known from geophysics but one would like to simulate its contours. This is a case when the most likely training image is already defined on the simulation grid, for instance from a geophysical interpretation. The point of the simulation is now to perturb the interpreted shapes for uncertainty characterization.

In such cases the focus should be on the patterns of the contacts between the facies. The assumption is that the geological uncertainty decreases away from the interpreted contacts. For example, the central part of channel complex may be known with certainty; it is its extent that is uncertain. The relevant patterns are those seen when transitioning from one geological facies to another; this is also where the patterns are repeating. The mps algorithm then learns the shapes of these contacts instead of the shape of the body itself. This can be done by defining a zone of uncertainty around the facies contacts; this zone of uncertainty serves both as a repository of patterns and as simulation domain. The larger the uncertainty zone the more dissimilar the realizations. The definition of the uncertainty zone would take into consideration the geologists understanding of the deposit and their ability to model the contact zone.

Consider Figure 3 that shows such case with a carbonate mound. Note that the geological objects do not repeat; there is only one inner core (red), one shell (yellow) and one set of debris (gray). In this case, the repository of patterns and the simulation zone is restricted to regions in the training images that contain contacts between facies. The extent of these regions can be increased or decreased depending of the certainty of the original position and/or shape of the base geological interpretation. This assumes that the relative position of each part of the mound is known (for instance from geophysics).

Building training images

Now that algorithms and techniques to handle many types of training images are available, one still needs to find a relevant training image. Most academic research has focused on developing faster and better simulation algorithms, but very few aim at tackling the challenge of obtaining the training image. This is unfortunate as I would argue the main difficulty in using mps algorithms are not the mps algorithms per se but building the training image. There is a huge difference between taking an appealing training image and generating simulations versus starting with a data set and having to find a relevant training image.

At this time, object based algorithms are the best sources of 3D Tis since they need not be conditional to data. A TI must only correctly represent the geo-objects and the spatial relationship between the geo-objects. Most current object-based simulation programs are specific to a geological environment with most of the developments targeted for fluvial systems.
such as fluvsim (Deutsch and Wang, 1996; Deutsch and Tran, 2002 or more recently with event-based fluvial model (Pyrcz et al., 2009) which can reproduce a wide variety of fluvial systems.

Figure 3 Simulation of a carbonate mound by re-simulating the contact between the facies. Each column represents a different view of the mound which consists of an inner structure (red), an outer shell (yellow) and debris (gray). The first row is the training image, created using the SGeMS ti-generator and the subsequent rows show 5 simulations.
The new SGeMS Object Simulator (Boucher et al., 2010) aims at handling many types of geological objects by overcoming the rigidity of predefined geological objects in the current training image generators. It allows the modeler to generate geological objects with complex geometries and with relevant interactions between these objects.

First, a geological object (geo-object) is built either from pre-coded shapes, such as ellipsoid, cuboid, kernels, or user-defined. These basic shapes are then assembled with geometrical operations (difference, union and intersection) to create new complex shapes. For added flexibility, any shape can be translated, rotated and sheared. A second set of parameters control the interactions between the geological objects. Each parameter used in the training image construction (e.g. size, rotation angles, number of stacks) can vary in space. This locally varying parameterization allows the representation of trends in geological body geometry, interactions, and locations.

The defined geo-objects are then positioned on a grid given a set of constraints through a simulation event that controls the target number of geo-objects (stopping criterion), the interaction between geo-objects (interactions rules) and potential preferential locations (positioning and spatial intensity field). For common geometries such as fluvial and carbonate systems, specialized object are created to ease the modeling. The training images found in this paper have been created with this algorithm.

Specifically, the training image seen in Figure 3 has been generated by combining one half sphere (the inner core), one East-West sheared half-ellipsoid (the shell) and one asymmetric and East-West sheared bell curve-shaped volume (parameterized by a radial kernel). Finally the inner core supersedes both the outer shell and the debris; the outer shell supersedes the debris.

**Conclusion**

Multiple-point algorithms are very appealing for modeling geology. They are a great communication tool between the geologist and modelers since the geological concepts are passed as an image instead of a mathematical function (e.g. spherical variogram). They do offer the promise of more realistic reservoirs; however they do not ease the modeling process. Choosing the training image is much harder than selecting a variogram model mainly because the lack of tools to create TI and because of higher expectations.

The training image is a bridge between the geological knowledge about the reservoir and the numerical model. The mp algorithms do not understand the geological concept behind the a training image, it is the modelers responsibility to provide that information, preferably with trend function, such that the geological concept is correctly reproduced on the simulation grid. Finally, there is no point of designing a complex “realistic” training image if the algorithm to be used cannot handle it or if the user lacks the skill to integrate that information into the algorithm.
References


Mariethoz, G., P. Renard, and J. Straubhaar (2010), Direct sampling method to perform multiple- point geostatistical simulations, Water Resources Research., 46(11)


