Petroleum geostatistics for nongeostaticians

Part 2

Richard L. Chambers, CGC, Broken Arrow, Oklahoma, U.S.
Jeffrey M. Yarus, Knowledge Reservoirs, Houston, Texas, U.S.
Kirk B. Hird, BP Amoco, Houston, Texas, U.S.

For more than a decade, stochastic, or geostatistical, modeling methods have been increasingly used to "map" spatially correlated data. Recall that kriging is a deterministic method whose function has a unique solution and that does not attempt to represent actual variability of the studied attribute. Thus, the smoothing property of kriging dismisses local detail in place of a good average. However, often the geoscientist or reservoir engineer is more interested in fine-scale details captured by the estimation variance than a map of local estimates of the mean.

Geostatistical jargon is confusing. For example, many authors often use *stochastic*, *probabilistic* and *conditional* simulation interchangeably. We consider a stochastic model to be conditional when it honors the measured data and the spatial model (variogram or covariance). But for the sake of simplicity, we also consider these terms equivalent in this article.

To many, stochastic methods are analogous to tossing a coin. They are suspicious because it is well known that the natural processes responsible for creating reservoirs are not random. In light of this, stochastic methodologies are often rejected outright. Although it is true that reservoirs are not products of random processes, it is also true that they have attributes that cause them to behave as if they were random. For example, physical and chemical processes often change reservoir characteristics from their initial state, confounding our ability to make predictions even when the processes are understood. Such

*Editor's Note: Part 1 of this article (*TLE, *May 2000) defined geostatistics, examined its origins, and reviewed the concepts of the spatial model and the kriging interpolation algorithm. This article describes geostatistical conditional simulation, also known as stochastic modeling, and its use for generating realistic maps of reservoir heterogeneity, uncertainty analysis, and economic risk analysis.The Geologic Column, which appears monthly in* TLE, *is (1) produced cooperatively by the SEG Interpretation Committee and the AAPG Geophysical Integration Committee and (2) coordinated by M. Ray Thomasson and Lee Lawyer.*

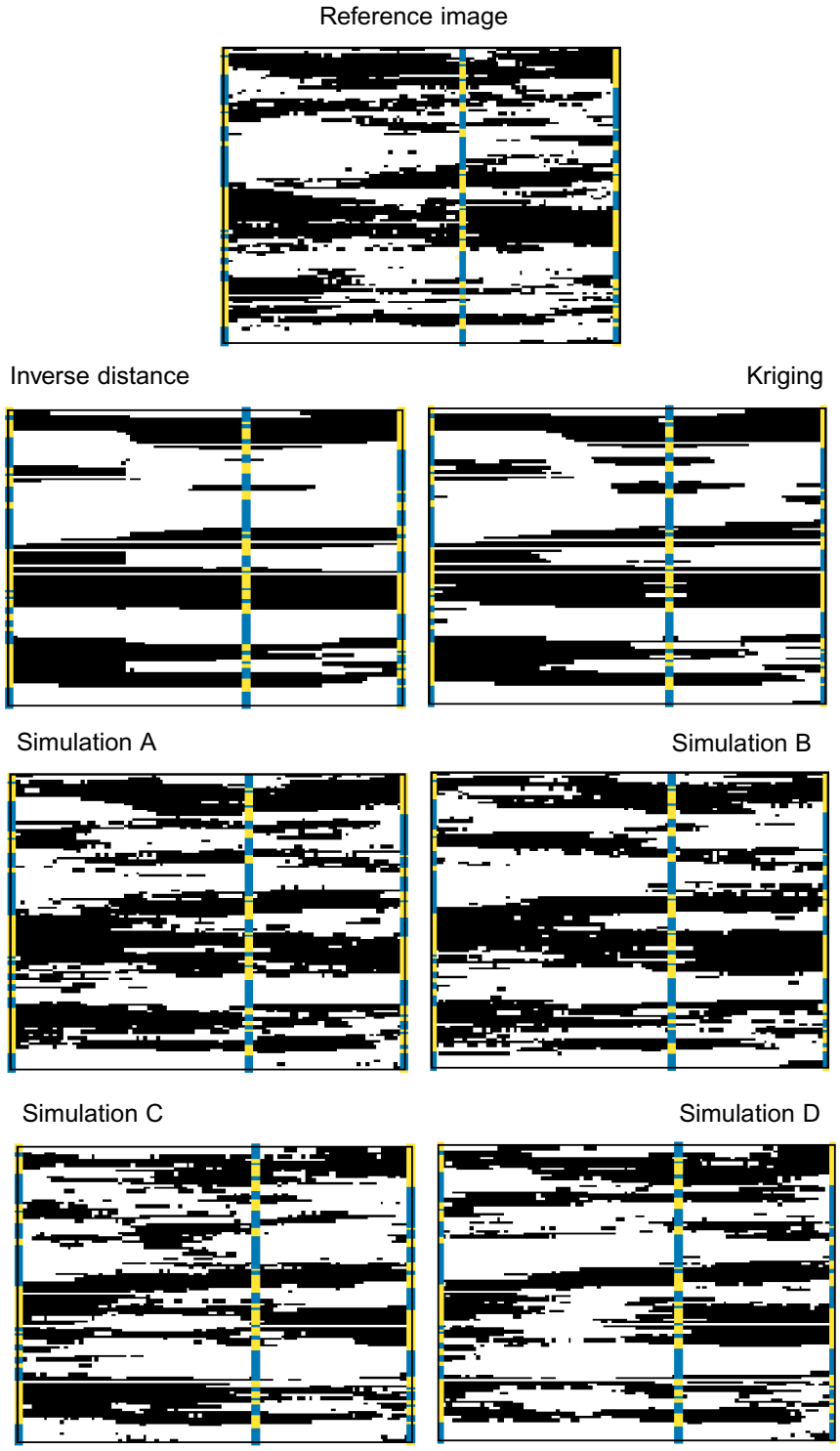

Figure 1. Distribution of sand (white) and shale (black). Vertical sampling = 5 m and lateral resolution = 100 m in reference image. Three "wells" penetrate the section. Sand along the well bore is blue and shale is yellow. Inverse distance, kriging, and conditional simulations A-D are based on well data.

changes, though, result in behavior that can be captured using stochastic principles.

Additionally, like the traditional deterministic approach, stochastic modeling preserves "hard" data where known and interpretative "soft" data where informative. But unlike the deterministic approach, it provides geoscientists and reservoir engineers many equally probable models called realizations. The kriged solution is the average of numerous realizations, and the variability among different outcomes is simply a measure of uncertainty at any location. Thus, the standard deviation of all values simulated at a fixed location is a quantification of its uncertainty.

What do we want? Before describing various simulation methods, it is useful to ask what it is that we want from a stochastic modeling effort. The method we choose depends on the goal and—to a great extent—the types of available data. Not all conditional simulation studies need a Cadillac when a Volkswagen will do.

A number of reasons exist for performing stochastic simulation; four important ones are: (1) capturing heterogeneity; (2) simulating facies or rock properties or both; (3) honoring and integrating complex information; (4) assessing uncertainty.

Capturing heterogeneity. A good model of heterogeneity implies better understanding of connectivity between permeable and nonpermeable zones, resulting in better sweep efficiencies and production forecasts. Over the past decade, it has become increasingly apparent that reservoir performance predictions are more accurate when based on models that reflect possible reservoir heterogeneity. We are painfully aware of countless examples of failed predictions due to overly simplistic models.

Although detailed models of heterogeneity are not necessarily required for volumetric calculations, heterogeneity does have great impact on flow characteristics. Each realization should have about the same net-to-gross ratio, although generally somewhat more conservative than simple deterministic models. However, because stochastic simulation preserves the variance from observed data, the realizations are "rougher" in appearance, showing more variability numerically in the interwell space and thus emulating more realistic conditions. Although each stochastic realization results in a different picture of fluid flow, each realization generally provides more realistic information about the actual flow behavior than conventional deterministic models.

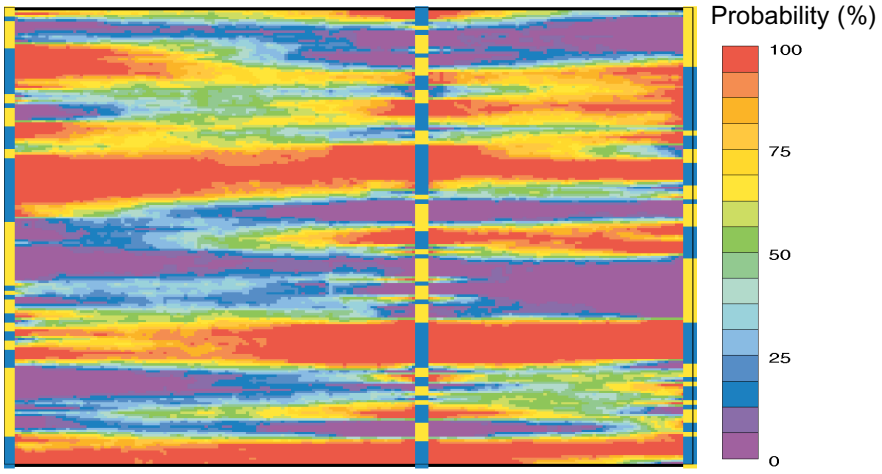

Figure 2. Uncertainty map showing the probability of occurrence of sand computed from 50 simulations based on the three wells in Figure 1.

The reference image in Figure 1 shows distribution of sand (white) and shale (black) with vertical sampling of 5 m and lateral resolution of 100 m. Three "wells" penetrate the section, with sand (blue) and shale (yellow) noted along the well bore. The images of inverse distance, kriging, and four conditional simulations are based on data from the three wells; each image is different. The reference image shows more heterogeneity with a net-to-gross (N/G) ratio of 46.3%. Inverse distance (N/G 44.5%) and kriging (N/G 46.1%) show much more lateral connectivity of shales—not unexpected because of the smoothing property of interpolation methods. Four of 50 simulations, also based on three wells, look similar globally but different locally and have the "look and feel" of the reference image. N/G ratios range from a low of 42.9% to a high of 50.8% but average 46.3%. Although many stochastic realizations are created to understand and quantify uncertainty, quite often only a single realization is used for flow simulation and performance prediction.

Simulating facies or rock properties (or both). Modelers generally follow a multistep approach when constructing a stochastic reservoir model. The reservoir architecture, usually the first priority, consists of the overall structural elements—faults, top and base of reservoir, etc. The second step is identifying different geologic units, based on sequence stratigraphic principles, and defining internal bedding geometries (onlap, offlap, etc.) for each unit. The next step typically involves modeling spatial distribution of depositional facies (e.g. eolian, deepwater fan, channels, etc.), or log-derived facies (lithofacies). Depositional facies information generally provides the greatest clarity in terms of spatial geometries, but this is not always available. Lithofacies are more easily obtainable but are not guaranteed to honor actual depositional facies boundaries. However, the two are usually strongly related, and it is common at this point to group facies exhibiting similar petrophysical and saturation properties into more general units called lithotypes. The lithotypes are then modeled and the results should reflect the spatial arrangement of the flow units. The final step in building the static (high-resolution) geologic model is to populate the lithotypes with rock and fluid properties. The important difference between modeling facies and modeling rock properties is that the former is a categorical variable and the latter are continuous variables.

Honoring secondary data. Stochastic methods, such as kriging also allow us to incorporate a broad range of information that most conventional methods cannot accommodate. Many are interested in the stochastic simulation not for its range of plausible outcomes but for its ability to simultaneously integrate additional "soft" data (e.g., seismic or well tests) because it improves reliability away from control points where only secondary forms of data are available.

Assessing uncertainty. Anyone who

forecasts reservoir performance understands that there is always uncertainty in the reservoir model. Performance forecasts or volumetric predictions are often based on a "best-case" model. However, a reservoir asset team may also be interested in other models, such as the "pessimistic" and "optimistic" cases. A minimum of three models allows the team to assess whether the development plan, based on the best-case scenario, is flexible enough to handle a range of uncertainty.

Stochastic simulation lends itself to Monte Carlo risk analysis because the methods offer many models consistent with the input data. One critical aspect is the belief in some "space of uncertainty" that has been sampled unbiasedly and adequately by a set of realizations. Results can then be summarized as a probability distribution rather than a simple selection of some limited number of plausible realizations from a larger set.

Stochastic simulation is not an estimation method like kriging. The value simulated at any fixed grid node represents a value drawn from a probability distribution. This distribution is predetermined and is based on information from the control data within the search ellipse, informative "soft" data (if used), and the spatial model. Thus, 100 stochastic realizations yield 100 simulated values per grid node.

The questions now become (1) *What do I do with all these simulated images?* and (2) *Which is correct?*

In answer to the second, any of the simulated images is, technically, equally probable. However, just because each realization is equally probable does not mean each is geologically acceptable. Each simulated image should be examined to determine whether it is a reasonable representation of what is known about the reservoir. If not, it should be discarded.

The first question is answered with an example. Let us say that a geostatistical study used collocated kriging for porosity with seismic acoustic impedance as the secondary attribute. How reliable is the map? In other words, what level of confidence is associated with the results? The same question could be asked about a map created with the inverse-distance method or any traditional interpolation method (including hand-contouring).

Because stochastic modeling generates independent realizations, the numerous outcomes are often post-processed to quantify uncertainty. Possible maps generated from a suite of simulated images include:

- Mean: This map is the average of *n* conditional simulations. At each cell, the program computes the average value based on the values at that location from all simulations. When *n* is large, the map converges to the kriged solution.
- Standard deviation: A map of the standard deviation at each grid cell computed from all input maps. This measures the standard error and is used to analyze uncertainty.
- Uncertainty or risk: This map displays the probability of meeting or exceeding a user-specified threshold at each grid cell. Grid cell values range from 0 to 100%.
- Isoprobability: These maps display the attribute value at a constant probability threshold.

Figure 2 is an uncertainty map for the occurrence of sand computed from 50 simulations based on the three wells in Figure 1. Remember that blue along the well bore indicates sand and that yellow is shale. Thus, the most likely locations of porous rock can be readily identified and locally validated by well-bore information.

Stochastic simulation. Stochastic simulation is a Monte Carlo technique designed to honor measured data; closely reproduce the data histogram; honor the spatial model; be consistent with secondary data; and assess uncertainty in the reservoir model.

Several simulation methods are available. The choice depends on goals and data types and availability. Commonly used methods are turning bands, sequential simulation, simulated annealing, probability field, matrix decomposition, and Boolean (or object-based models such as marked-point process).

Turning bands. This, one of the earliest simulation methods, first creates smooth kriged models along a set of random lines on top of a regular grid. Next, residual values are added to each value along a given line to reproduce the original data variance. The residual values are added through a nonconditional simulation step that uses the same histogram and spatial model as in kriging but does not use the actual data values at the wells. The number of lines is controlled by the modeler and the nodes between lines are linearly interpolated—the more lines, the less interpolation. The final model honors the original data, the spatial model, and the original variance and has an appropriate level of spatial heterogeneity.

Sequential simulation. Three sequential simulation procedures use the same basic algorithm for different data types.

- Sequential Gaussian simulation (SGS) simulates continuous variables, such as petrophysical properties. The procedure is essentially the same as kriging, with the addition of a bias.
- Categorical sequential indicator simulation (SIS) simulates discrete variables. It is possible to create a grid of 0s and 1s, using the same methodology as SGS, which represent "lithofacies" (pay/nonpay, or sand/shale).
- Bayesian sequential indicator simulation, a newer form of SIS, allows direct integration of seismic attributes using a combination of classification and indicator methods.

The general process is:

1) Select at random a node not yet simulated in the grid.
2) Use kriging to compute a local conditional probability distribution function (lcpd), with zero mean and unit variance. Computing the lcpd depends on the simulation method used.
3) Draw at random a single value, zi from the lcpd, whose maximum spread is two standard deviations about the mean, mi.
4) Create a newly simulated value ZSi* = mi + zi.
5) Include the newly simulated value ZSi* in the set of conditioning data. This ensures that closely spaced values have the correct short-scale correlation.
6) Repeat until all grid nodes have a simulated value.

The order in which grid nodes are randomly simulated influences the cumulative feedback effect on the outcome. The selection process is random but repeatable. For each simulation, grid nodes are shuffled into an order defined by a random seed value. Each random seed corresponds to a unique order of the grid nodes and different random seed values

produce different paths through the grid. Although the total possible number of orderings is very large, each random path is uniquely identified and repeatable.

Simulated annealing. This method was borrowed from metallurgy. When two pieces of metal are fused, the zone of attachment is heated to a point at which the atomic structure can be rearranged. As the metal cools, a bond is formed where the two pieces of metal are joined. As applied to stochastic modeling, the idea is to produce an initial starting realization, introduce some particular conditions (new pieces of metal to be fused), "heat" it up, and "cool" it down. The result is to rearrange the pixels to match the particular conditions introduced. The method constructs the reservoir model via iterative trial and error and does not use an explicit random function model. The simulated image is formulated as an optimization process.

For example, our desired result might be an image of a sand/shale model with a 70% net-to-gross ratio, an average shale length of 60 m, and average shale thickness of 10 m. The starting image has pixels arranged randomly with sand and shale in the correct global proportion, but net-to-gross is incorrect due to the random assignment of the sand and shale. In addition, the average shale length and width are too short. During the computation, the annealing algorithm attempts to modify the initial image by swapping information from node to node until the final image matches the statistics of the input data. This produces excellent results but can be inefficient; i.e., millions of perturbations may be required to arrive at the desired image. However, these methods are becoming more attractive because of the availability of faster computers with more memory.

Probability field. This method is an alternative to the sequential simulation methods described earlier. In sequential simulation, the value drawn from the local cumulative probability distribution (lcpd) at a particular grid node is treated as hard data and included as local conditioning data. This ensures that closely spaced values have the correct short-scale correlation. Otherwise, the simulated image would contain too much short-scale (high-frequency noise) variability. The idea behind probability field, or *P*-field, simulation is to increase the efficiency of computing the lcpd from the original well data only. *P*-field simulation gets around the problem of too much short-scale variability by controlling the sampling of the distributions rather than controlling the distributions as in sequential simulation. The method decouples computation of the lpcd and its sampling. Subsequent realizations don't need to recompute the lpcd, greatly speeding up computation time.

Matrix decomposition. Some simulation techniques involve matrix decomposition. L-U decomposition, for example, uses different outcomes created by multiplying vectors of random numbers by a precalculated matrix derived from spatial continuity information (typically from a variogram or correlogram). Matrix methods can be viewed as sequential simulation because multiplication across the rows of the precalculated matrix and down the column vector of the random numbers can be construed as a sequential process in which the value of each successive node depends on the values of previously simulated nodes. Only hundreds of nodes can be simulated at a time using matrix decomposition methods. Large models are simulated by splicing smaller models or by a moving-window approach.

Boolean (or object-based) techniques. These methods create reservoir models based on objects (groupings of pixels) that have a genetic significance rather than having been built up from one elementary node or pixel at a time. To use such methods, you need to select a basic shape for each depositional facies that describes its geometry. For example, you might want to model sand channels that look like half ellipses in cross-section, or deltas as triangular wedges in map view. You must specify the proportions of the shapes in the final model and choose a distribution for parameters that describe the shapes. Some algorithms have rules describing how geobodies are positioned relative to each other. For example, can they cross like braided streams or attach like splays and channels? Do the objects repulse or attract, or must there be a minimum distance between shapes?

Once the distribution of parameters and position rules are chosen, the remaining steps are:

1) Fill the reservoir model background with some lithofacies (e.g., shale).
2) Randomly select a starting point in the model.
3 Randomly select one lithofacies shape and draw it with appropriate size, anisotropy, and orientation.
4) Check if the shape conflicts with any conditioning data (e.g., well data) or with other previously simulated shapes. If not, keep the shape; otherwise, reject it and go back to the previous step.
5) Check to see if the global proportions are correct; if not, return to step 2.
6) Simulate petrophysical properties within the geobodies using more classical geostatistical methods.

If control data must be honored, this step is typically completed before the interwell region is simulated. Care must be taken to ensure that there are no conflicts with known stratigraphic and lithologic sequences in the wells.

Boolean or object-based techniques are of current interest in the petroleum industry—a number of research/academic institutions and commercial vendors are working on implementation algorithms. This approach to modeling is particularly satisfying to geologists because the objects created are based on the statistics of shapes and facies relationships that have been measured and because images from the ensuing depositional facies model look realistic.

In the past, Boolean-type algorithms could not always honor all conditioning data, because the algorithms were not strict simulators of shape. However, new technology has greatly alleviated this problem. The number of input parameters for Boolean modeling can be large, and this is sometimes a criticism. Nevertheless, most geologists are very familiar with the type of information required, and although this lends a certain amount of determinism to the model, it is practical—particularly when the depositional systems that comprise the reservoir are understood.

Conclusions. Geostatistics is a rapidly evolving branch of applied statistics and mathematics aimed at quantifying and modeling spatial variability. Spatial variability includes scales of heterogeneity and direction within data sets. Geostatistics is a powerful formalism that must be applied with care.

Hopefully, this article has helped readers understand that stochastic modeling algorithms are more than a simple coin-tossing experiment for simulating interwell space. Geostatistics treats deterministic information as such and what is unknown as probabilistic. Variability between realizations is a measure of uncertainty (remembering of course, that there is also uncertainty in the control data and in the spatial model applied during the simulation).

The choice of a geostatistical method depends on the project goal, data types and availability, depositional environment, and the scale at which the model must be used.

In 1998, Dubrule presented some thoughts about the future of stochastic modeling. Following are some of our thoughts which expand his ideas:

- Better geologically based rules for different environments: We need more effort toward improving and quantifying rules for geologic input to geostatistical modeling. As geologists become more aware of and comfortable with stochastic modeling techniques, geology will drive the algorithms rather than the opposite.
- Better use of indirect or "soft" information: Even though techniques exist to constrain geostatistical models by seismic attributes and well information, the results are not always accepted, because assumptions for their use are not always clear. Thus, techniques will be developed to help us better understand the interrelationships between petrophysical properties and seismic attributes, and to combine quantitatively static and dynamic data.
- Better quantification of uncertainty: Uncertainties exist at all levels, from input data to model assumptions and model parameters. Techniques will be developed to help us better quantify and use these uncertainties at all phases of modeling to provide better models for economic risk analysis.
- More effort on developing stochastic structural modeling methods: Quite often, geologic surprises in the fluid-flow modeling could be related to poor use of fine-scaled structural features such as subseismic faults. Several modeling packages are now on the market for modeling and simulating fault and fractures patterns. However, it is still difficult to incorporate these results into the overall high-resolution geologic model and finally in the fluid-flow simulation, especially as related to the latter, because flow-simulator software requires so much upscaling (coarsening) of the geologic model that fine-scale structural and stratigraphic features are often lost. Future developments in parallel processing for flow simulation should help minimize the need for so much upscaling.
- Integration of geostatistical modeling tools in 3-D "earth-modeling" tools: Great strides have been made in developing tools which integrate structural, stratigraphic, sedimentological, and petrophysical information into a consistent 3-D "earth model." Geostatistical models are becoming a tool kit within earth-modeling software.

Suggestions for further reading. *Geostatistics in Petroleum Geology* by Dubrule (AAPG Continuing Education Course Notes No. 38, 1998). *Geostatistics and Petroleum Geology* by Hohn (Kluwer Academic Publishers, 1999). *Stochastic Modeling and Geostatistics*, edited by Yarus and Chambers (AAPG Computer Applications in Geology, No. 3, 1994). TLE

Corresponding author: geodoc547@aol.com