Petroleum geostatistics for nongeostatisticians

Part 1

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.

Coordinated by M. Ray Thomasson and Lee Lawyer

This is the first of two articles intended to describe petroleum geostatistics for the nongeostatistician. There are many misconceptions about geostatistics, what it is, and what it can or can't do for the petroleum industry.

The first article defines geostatistics, examines its origins, and reviews the spatial model and the kriging interpolation algorithm. The second article describes geostatistical conditional simulation and its use for uncertainty (risk) analysis.

Earth science data exhibit spatial correlation to greater or lesser degrees. As the distance between two data points increases, the similarity between the two measurements decreases. Geostatistics is a rapidly evolving branch of applied statistics and mathematics that offers a collection of tools which quantify and model spatial variability. Spatial variability includes scales of variability (heterogeneity) and directionality within data sets.

Origins of geostatistics. The origins of geostatistics are found exclusively in the mining industry. D. G. Krige, a South African mining engineer, and H. S. Sichel, a statistician, developed a new estimation method in the early 1950s when "classical" statistics was found unsuitable for estimating disseminated ore reserves.

Georges Matheron, a French engineer, developed Krige's innovative concepts and formalized them within a single framework with his *Theory of Regionalized Variables*. Matheron, at the Centre de Geostatistique, pioneered the use of mining geostatistics in the early 1960s. The word *kriging* was coined in recognition of D. G. Krige.

It is interesting that geostatistics was not originally developed to solve interpolation problems (kriging) but to address what is called the support effect. In ore mining, this refers to the difference between the variance of average values measured from large samples and the variance of average values measured from small samples, which leads to a systematic bias in estimates. Support means the volume on which a property is measured. Most of us unwittingly overlook the support effect in the petroleum industry, especially when combining well-log measurements and seismic attributes, or core and well test permeabilities.

Editor's Note: 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.*

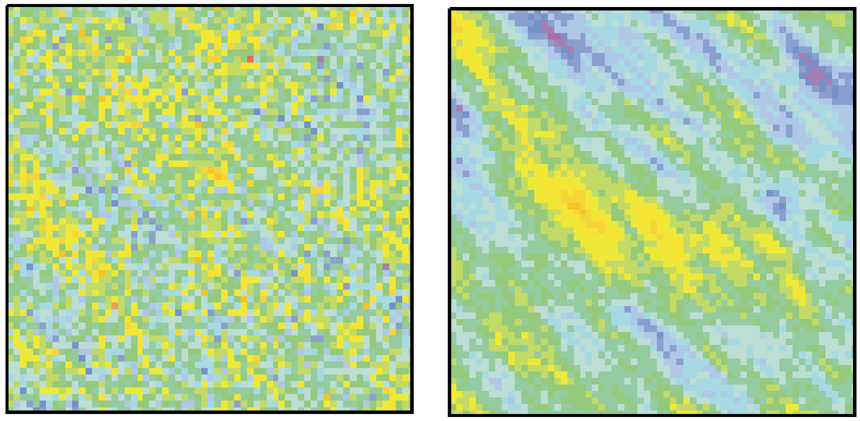

Figure 1.

By the early 1970s, kriging had proved to be very useful in the mining industry. Geostatistics was introduced to the petroleum community in the mid-1970s through its first commercial software package, BLUEPACK.

The technique spread to many other areas of earth science in the 1970s with the advent of high-speed computers. However, it was not until the mid- to late 1980s that geostatistical techniques were used to any extent in the petroleum industry, and its popularity has grown every year since.

Geostatistics in the petroleum industry. Maps and mapmaking are integral parts of reservoir characterization. A map is a numerical model of an attribute's (e.g., porosity, permeability, thickness, structure) spatial distribution. However, mapping an attribute is rarely the goal; rather, a map is used to make a prediction about the reservoir. To paraphrase Andre Journel of Stanford University, "A map is a poor model of reality if it does not depict characteristics of the real spatial distribution of those attributes that most affect how the reservoir responds."

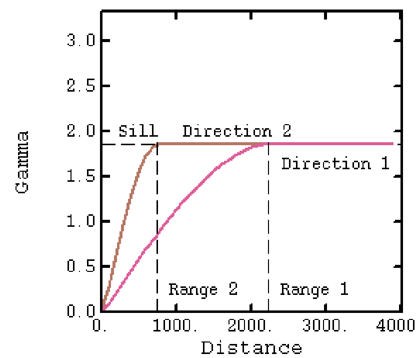

Figure 2. Anisotropic variogram showing a short-scale correlation range of 800 m and a long-scale range of 2200 m.

The enormous up-front investments for developing heterogeneous fields and the desire to increase ultimate recovery have spurred oil companies to use innovative reservoir characterization techniques. Geostatistics is one of many new technologies often incorporated into the process. For more than a decade, geostatistical techniques, especially when incorporating 3-D seismic data, have been an accepted technology to characterize petroleum reservoirs.

Geostatistical application necessitates and facilitates cooperation between geoscientists and reservoir engineers, allowing each discipline to contribute fully. This is quite different from the past, because the mathematical formalization was often left to the reservoir engineer. Thus, part of the geostatistical philosophy is to ensure that geologic reality does not get lost during reservoir model building.

Geostatistics attempts to improve predictions by developing a different type of quantitative model. The goal is to construct a more realistic model of reservoir heterogeneity using methods that do not average important reservoir properties. Like the traditional deterministic approach, it preserves indisputable "hard" data where they are known and interpretative "soft" data where they are informative. However, unlike the deterministic approach, geostatistics provides numerous plausible results. The degree to which the various models differ is a reflection of the unknown or a measurement of the "uncertainty." Some outcomes may challenge prevailing geologic wisdom and will almost certainly provide a range of economic scenarios, from optimistic to pessimistic. Having more than one result to analyze changes the paradigm of traditional reservoir analysis and may require multiple reservoir flow simulations. However, the benefits outweigh the additional time and cost.

Again, to paraphrase Journel, "... it is better to have a model of uncertainty than an illusion of reality."

Basic elements of a geostatistical study. Once initial data sets are prepared, quality controlled, and loaded into the geostatistical software, a typical work flow, with iterations, might be: (1) data mining; (2) spatial continuity analysis and modeling; (3) search ellipse design; (4) model crossvalidation; (5) kriging; (6) conditional simulation; (7) model uncertainty assessment.

The first five steps are discussed in this article. Topics 6 and 7 will be described next month.

Data mining. An early and fundamental step in any science starts at the descriptive stage. Until facts are accurately gathered and described, an analysis of their causes is premature. Because statistics generally deals with quantities of data, not with a single datum, we need some means to deal with the data in a manageable form. Thus, much of statistics deals with ways of describing the data and understanding relationships between pairs of variables. *Data speak most clearly when they are organized* (Isaaks and Srivastava, 1989).

Because there is no one set of prescribed steps in data mining, you should follow your instincts in explaining anomalies in the data set. By using various tools, you gain clearer understanding of your data and also discover possible sources of errors. Errors are easily overlooked, especially in large data sets and when computers are involved, because we simply become detached from our data. Thorough analysis fosters an intimate understanding of the data that can flag spurious results.

The petroleum industry has a classic dilemma that can be summarized as:

- very few direct "hard" observations (well data) are available
- "soft" data (e.g., seismic, well tests) are only indirectly related to the "hard" data (e.g., core, logs)
- few observations lead to much uncertainty
- it is still necessary to make predictions about the reservoir

Because there is no economic way to improve the ratio between hard and soft data, the use of geostatistical techniques is inevitable. Before discussing some of these techniques, here is some background concerning the "classical" approach.

Classical statistical data analysis includes data posting, computation of means and variances, making scatterplots to investigate the relationship between two variables, and identification of subpopulations and potential outliers.

Histograms, graphical representations of the data distribution of a single variable, record how often values fall within specified intervals or classes. A bar depicts each class, and its height is proportional to the number of values within that class. The histogram shape informs us about the distribution of the data values. Ideally, we like to see a bell-shaped, symmetrical distribution around the mean value. This is referred to as a normal, or Gaussian, distribution and has a predictable shape based on the data mean and variance. Many statistical and geostatistical methods assume such a data model. If the shape is skewed to either side of the mean, then often it is necessary to adjust the shape by transforming the data into Gaussian form. Complex histograms may indicate mixing of multiple distributions. Categorization of the data (e.g., by facies) often identifies the underlying distributions.

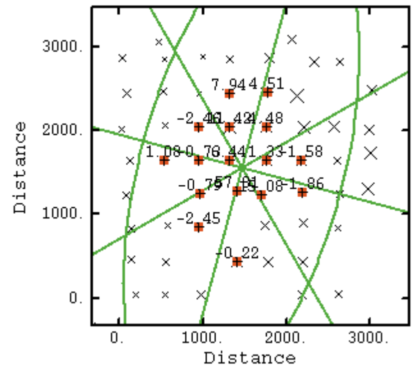

Figure 3. An anisotropic search ellipse with eight sectors and a maximum of two data points per sector. The minor axis has a length of 1000 m. The major axis (N15E) has a length of 4000 m. The center of the ellipse is the target grid node for estimation. There are 55 sample points (x) in the study area. Weights (in %) are shown for data control points used for estimation at the target point.

Spatial continuity analysis and modeling. Variables of interest in the petroleum industry (e.g., porosity, permeability, saturation, sand/shale volumes, etc.) are the product of a vast number of complex physical and chemical processes. These processes superimpose a spatial pattern on reservoir rock properties, and it is important to understand the scales and directional aspects of these features for efficient hydrocarbon production. The spatial component makes these variables complicated, and we are forced to admit uncertainty about their distribution between wells. Because deterministic models do not handle uncertainties associated with these variables, a geostatistical approach is used because its foundation is probabilistic theory (covariance models) that recognizes these inevitable uncertainties.

Consider the two images in Figure 1. The image on the left has a nearly random appearance but does show some preferential alignment from northwest to southeast. The right image has a higher degree of spatial continuity and anisotropy, also aligned northwest to southeast. Visually, the images appear quite different, but their mean values and variances are identical. Classical statistical analysis cannot properly address the spatial continuity and directionality inherent in earth science data. Thus, we require a model describing the continuity, anisotropy, and azimuthal properties in the data. The spatially correlated variable and its mathematical expression form the foundation of geostatistics.

The spatial model. Spatial continuity analysis quantifies the variability of sample properties with respect to distance and direction (geographic location is considered only if the data exhibit a trend, a property known as nonstationarity).

Quantifying spatial information involves comparing data values at one location with values of the same attribute at other locations. For example, two wells in close proximity are more likely to have similar reservoir properties than two wells farther apart. The key question—what we want to know—is *what measured values tell us about reservoir properties at unsampled locations*.

Spatial continuity or spatial correlation analysis includes two main steps:

1) Compute experimental measures of spatial continuity, accounting for anisotropy and azimuthal directions (e.g., variogram or covariance).
2) Model the experimental variogram or covariance for use in mapping.

If data are sampled on a regular grid, the calculation search strategy is simple. Unfortunately, well data rarely form a neat regular array; therefore, to extract as much information from the data as possible, we search for data within a bin rather than searching along a simple vector. Spatial continuity is often very frustrating and sometimes seemingly hopeless; remember that it is an iterative process and much is learned about the data in the process of executing the process. The two common measures of spatial continuity are the variogram and the covariance.

Variogram. For each azimuth and lag (separation) distance studied, all measured values can be spatially correlated and expressed as a statistical value known as the variogram (γ):

$$\gamma_{(h)} = \frac{\sum\left[Z_{(xi)} - Z_{(xi+h)}\right]^2}{2n}$$

where Z(xi) = the sample value at location xi; Z(xi+h) = the sample value at location xi + h, h = the lag distance, and n = the number of data pairs.

The (semi)variogram correlation term is a measure of dissimilarity, or increasing variance as a function of distance. The variogram is the sum of the squared differences of all data pairs falling within a bin (lag) divided by twice the number of pairs found for that lag. Computing and plotting gamma as a function of increasing lag distance, *h*, results in a plot of the experimental variogram.

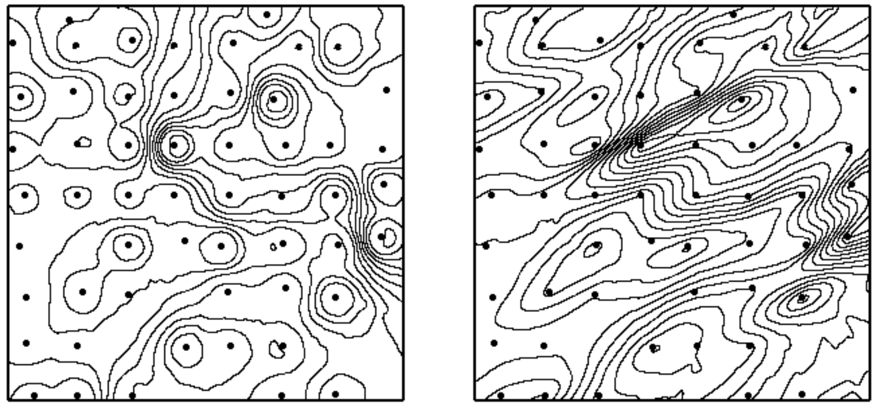

Figure 4. The difference between inverse weighted difference interpolation (left) and kriging (right) using a 3:1 anisotropic variogram model oriented N60E. The neighborhood search ellipse is identical for both.

Covariance. The covariance, another measure of spatial dependence, is derived from the variogram. The variogram increases with separation distance, but the covariance decreases. Thus, the covariance is a measure of similarity (correlation).

The relationship, where $C_{(o)}$ is the variogram sill,

$$C_{(h)} = C_{(o)} - \gamma_{(h)}$$

confirms the above statement that the covariance behaves inversely with the variogram.

Computing the covariance for increasing lags (double, triple, etc.) results in a plot showing decreasing covariance (correlation) with distance. The correlation scale length is determined when the covariance value reaches zero (no correlation).

Anatomy of an anisotropic correlation model. With increasing distance, $\gamma_{(h)}$ (or $C_{(h)}$) tends to reach a constant value, known as the sill (dashed horizontal line in Figure 2). For a variogram, the sill is the variance (σ^2) of measured data. The distance at which the sill is reached by the variogram is called the *range* or *correlation length*. The covariance reaches its range when $C_{(h)} = 0$. The range for a variogram and the covariance should be the same for a given set of search parameters. The *sill* and *range* are useful properties when one compares directional trends in the data. Often the variogram and covariance show a discontinuity at the origin, termed the *nugget effect*. The nugget effect is considered random noise and may represent short-scale variability, measurement error, sample rate, etc.

Spatial continuity analysis is one of the most important steps in a geostatistical study, because it strongly influences the kriging and conditional simulation results and associated uncertainties.

Kriging and conditional simulation applications require knowledge of the correlation function for all possible distances and azimuths. This requires a model of the experimental variogram (covariance) in order to know the variance (covariance) at any location, not just along specific interdistance vectors corresponding to angular/distance classes. Spatial modeling is not curve fitting, in the least-squares sense, because the selected model must ensure that the kriging variance is ≥ 0—a condition not necessarily satisfied by least-squares or other fitting methodologies.

Search ellipse design. Because computers are used in mapping, we must instruct the program how to gather and use control points during interpolation. Most familiar with computer mapping know that this involves designing a search ellipse or neighborhood. We must specify the length of the search radius, the number of sectors (typically four or eight), and the number of data points per sector. Most common mapping programs allow the user to specify only one radius; thus, the search ellipse is circular (isotropic). However, during geostatistical analysis, we often find that the spatial model is anisotropic. Thus, we should design the search ellipse based on the spatial model correlation scales, aligning the search ellipse azimuth with the major axis of anisotropy (Figure 3).

Model crossvalidation. Crossvalidation tests the "goodness" of the spatial model and the search ellipse design. The procedure compares estimated values with measured values, just as one computes residuals between predicted and observed values in regression or analysis of variance. The procedure is:

> For each sample in the data set, compute a kriged estimate at the same location, using the spatial model and search ellipse parameters but ignoring that sample value during reestimation. Thus, each sample value of a data set has a reestimated value and a measure of the kriging variance. From this information, various displays are created which are often humbling. One common display is a scatterplot of measured versus reestimated values. If the model were perfect, the scatterplot would be a straight line, but this never occurs. However, if the kriged estimates are unbiased, averages of the estimated and measured values should be equal.

Another usual display is the histogram of the standardized estimation error, which is the reestimated minus the observed values, divided by the kriging variance. If the histogram is symmetrical about a mean of 0, the estimates are unbiased. This ensures that anywhere in the mapped area, interpolated values have an equal chance of being over- or underestimates of the true value.

Kriging. Contouring data by hand or by computer uses some type of interpolation procedure. Many algorithms are used in computer mapping, and all require some criterion to be satisfied. Quite often a computer-generated map is unsatisfactory because it doesn't look "real"—that is, it doesn't depict the geology as we envision it. Thus, the computer map often requires much editing. The geoscientist working by hand interpolates between data points, draws connecting contours, smoothes the map to make it look real, and biases the contours based on a geologic model.

Inverse weighted distance is a commonly used mapping algorithm, and its formulation is easily understood. The weights used in the interpolation are based on how far each control point (observed value) is from the target (grid node). Thus, control points closest to the target receive higher weights. However, if the data exhibit strong anisotropy, it does not hold that the closest control point should receive the greatest weight; rather, more distant control points need to have greater influence on the interpolated value.

Kriging is a geostatistical interpolation technique. It is a linear weighted-averaging method, similar to inverse weighted distance. However, kriging weights depend on a model of spatial correlation. Therefore, it is possible to create a map exhibiting strong anisotropy, resulting in a map that "looks" more geologically plausible (Figure 4).

Conclusions. Geostatistics is not magic or a panacea. It is not a replacement for good data and is not a replacement for thorough understanding and analysis of the data. The results must be interpreted and validated in light of reservoir geology, rock physics, and reservoir engineering information/principles. Geostatistics is a tool for helping to incorporate geologic concepts into a quantitative 2-D or 3-D representation.

Kriging is a deterministic method that has a unique solution offering the best estimate. It does not pretend to represent the actual variability of the studied attribute. It can be used in the traditional way that other mathematical interpolation methods have been used. It has the added value of incorporating the spatial model and thus more reliably depicting the shapes of geologic features.

Next month's paper will develop the concept of stochastic modeling and how it is used to model reservoir heterogeneity more realistically by creating many plausible representations of an attribute of interest. This method not only captures the spatial continuity like kriging but, unlike kriging, it also preserves the variability.

Suggestions for further reading. *Geostatistics in Petroleum Geology* by Dubrule (AAPG Continuing Education Course Note Series 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). *An Introduction to Applied Geostatistics* by Isaaks and Srivastava (Oxford, 1989). E

Corresponding author: geodoc547@aol.com