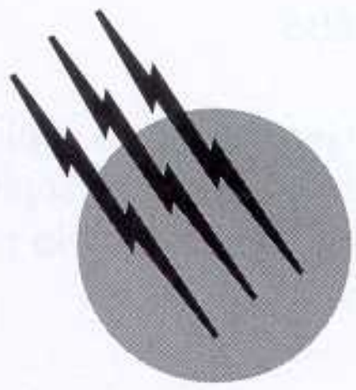


# Geostatistics

**Clayton V. Deutsch**
*University of Alberta, Edmonton*



## GLOSSARY

**Declustering** Technique to assign relative weights to different data values based on their redundancy with nearby data. Closely spaced data get less weight.

**Kriging** After the name of D. G. Krige, this term refers to the procedure of constructing the best linear unbiased estimate of a value at a point or of an average over a volume.

**Realization** Nonunique grid of simulated values. A set of realizations is used as a measure of uncertainty in the variable being studied.

**Simulation** Procedure of adding correlated error by Monte Carlo to create a value that reflects the full variability.

**Variogram** Basic tool of the theory used to characterize the spatial continuity of the variable.

**GEOSTATICS** commonly refers to the theory of regionalized variables and the related techniques that are used to predict variables such as rock properties at unsampled locations. Matheron formalized this theory in the early 1960s (Matheron, 1971). Geostatistics was not developed as a theory in search of practical problems. On the contrary, development was driven by engineers and geologists faced with real problems. They were searching for a consistent set of numerical tools that would help them address real problems such as ore reserve estimation, reservoir performance forecasting, and environmental site characterization. Reasons for seeking such comprehensive technology included (1) an increasing number of data to deal with, (2) a greater diversity of available data at different scales and levels of precision, (3) a need to address problems with consistent and reproducible methods, (4) a belief that improved numerical models should be possible by exploiting computational and mathematical developments in related scientific disciplines, and (5) a belief that more responsible decisions would be made with improved numerical models. These reasons explain the continued expansion of the theory and practice of geostatistics. Problems in mining, such as unbiased estimation of recoverable reserves, initially drove the development of geostatistics. Problems in petroleum, such as realistic heterogeneity models for unbiased flow predictions, were dominant from the mid-1980s through the late 1990s.

*Encyclopedia of Physical Science and Technology, Third Edition, Volume 6*
Copyright © 2002 by Academic Press. All rights of reproduction in any form reserved.

Geostatistics is applied extensively in these two areas and is increasingly applied to problems of spatial modeling and uncertainty in environmental studies, hydrogeology, and agriculture.

## I. ESSENTIAL CONCEPTS

Geostatistics is concerned with constructing high-resolution three-dimensional models of categorical variables such as rock type or facies and continuous variables such as mineral grade, porosity, or contaminant concentration. It is necessary to have *hard* truth measurements at some volumetric scale. All other data types, including remotely sensed data, are called *soft* data and must be calibrated to the hard data. It is neither possible nor optimal to construct models at the resolution of the hard data. Models are generated at some intermediate geologic modeling scale, and then scaled to an even coarser resolution for process performance. A common goal of geostatistics is the creation of detailed numerical three-dimensional geologic models that simultaneously account for a wide range of relevant data of varying degrees of resolution, quality, and certainty. Much of geostatistics relates to data calibration and reconciling data types at different scales.

At any instance in geologic time, there is a single true distribution of variables over each study area. This true distribution is the result of a complex succession of physical, chemical, and biological processes. Although some of these processes may be understood quite well, we do not completely understand all of the processes and their interactions, and could never have access to the boundary conditions in sufficient detail to provide the unique true distribution of properties. We can only hope to create numerical models that mimic the physically significant features. Uncertainty exists because of our lack of knowledge. Geostatistical techniques allow alternative realizations to be generated. These realizations are often combined in a histogram as a model of uncertainty.

Conventional mapping algorithms were devised to create smooth maps to reveal large-scale geologic trends; they are low-pass filters that remove high-frequency property variations. The goal of such conventional mapping algorithms, including splines and inverse distance estimation, is *not* to show the full variability of the variable being mapped. For many practical problems, however, this variability has a large affect on the predicted response. Geostatistical simulation techniques, conversely, are devised with the goal of introducing the full variability, that is, creating maps or realizations that are neither unique nor smooth. Although the small-scale variability of these realizations may mask large-scale trends, geostatistical simulation is more appropriate for most engineering applications.

There are often insufficient data to provide reliable statistics. For this reason, data from analogous, more densely sampled study areas are used to help infer spatial statistics that are impossible to calculate from the available data. There are general features of certain geologic settings that can be transported to other study areas of similar geologic setting. Although the use of analogous data is often essential in geostatistics, it should be critically evaluated and adapted to fit any hard data from the study area.

A sequential approach is often followed for geostatistical modeling. The overall geometry and major layering or zones are defined first, perhaps deterministically. The rock types are modeled within each major layer or zone. Continuous variables are modeled within homogeneous rock types. Repeating the entire process creates multiple equally probable realizations.

### A. Random Variables

The uncertainty about an unsampled value $z$ is modeled through the probability distribution of a random variable (RV) $Z$. The probability distribution of $Z$ after data conditioning is usually location-dependent; hence the notation $Z(\mathbf{u})$, with $\mathbf{u}$ being the coordinate location vector. A random function (RF) is a set of RVs defined over some field of interest, e.g., $Z(\mathbf{u})$, $\mathbf{u} \in$ study area $A$. Geostatistics is concerned with inference of statistics related to a random function (RF).

Inference of any statistic requires some repetitive sampling. For example, repetitive sampling of the variable $z(\mathbf{u})$ is needed to evaluate the cumulative distribution function: $F(\mathbf{u}; z) = \text{Prob}\{Z(\mathbf{u}) \leq z\}$ from experimental proportions. However, at most, one sample is available at any single location $\mathbf{u}$; therefore, the paradigm underlying statistical inference processes is to trade the unavailable replication at location $\mathbf{u}$ for replication over the sampling distribution of $z$ samples collected at other locations within the same field.

This trade of replication corresponds to the decision of stationarity. Stationarity is a property of the RF model, not of the underlying physical spatial distribution. Thus, it cannot be checked from data. The decision to pool data into statistics across rock types is not refutable *a priori* from data; however, it can be shown inappropriate *a posteriori* if differentiation per rock type is critical to the undergoing study.

## II. QUANTIFICATION OF SPATIAL VARIABILITY

### A. Declustering

Data are rarely collected with the goal of statistical representivity. Wells are often drilled in areas with a greater

probability of good reservoir quality. Core measurements are taken preferentially from good-quality reservoir rock. These data-collection practices should not be changed; they lead to the best economics and the greatest number of data in portions of the reservoir that contribute the greatest flow. There is a need, however, to adjust the histograms and summary statistics to be representative of the entire volume of interest.

Most contouring or mapping algorithms automatically correct this preferential clustering. Closely spaced data inform fewer grid nodes and, hence, receive lesser weight. Widely spaced data inform more grid nodes and, hence, receive greater weight. Geostatistical mapping algorithms depend on a global distribution that must be equally representative of the entire area being studied.

*Declustering* techniques assign each datum a weight, $w_i, i = 1, \ldots, n$, based on its closeness to surrounding data. Then the histogram and summary statistics are calculated with the declustering weights. The weights $w_i, i = 1, \ldots, n$, are between 0 and 1 and add up to 1.0. The height of each histogram bar is proportional to the cumulative weight in the interval, and summary statistics such as the mean and variance are calculated as weighted averages. The simplest approach to declustering is to base the weights on the volume of influence of each sample. Determining a global representative histogram is the first step of a geostatistical study. The next step is to quantify the spatial correlation structure.

## B. Measures of Spatial Dependence

The covariance, correlation, and variogram are related measures of spatial correlation. The decision of stationarity allows inference of the stationary covariance (also called auto covariance):

$$C(\mathbf{h}) = E[Z(\mathbf{u}+\mathbf{h}) \cdot Z(\mathbf{u})] - m^2,$$

where $m$ is the stationary mean. This is estimated from all pairs of $z$-data values approximately separated by vector $\mathbf{h}$. At $\mathbf{h} = 0$ the stationary covariance $C(0)$ equals the stationary variance $\sigma^2$. The standardized stationary correlogram (also called auto correlation) is defined as

$$\rho(\mathbf{h}) = C(\mathbf{h})/\sigma^2.$$

Geostatisticians have preferred another two-point measure of spatial correlation called the variogram:

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]^2\}$$

The variogram does not call for the mean $m$ or the variance $\sigma^2$; however, under the decision of stationarity the covariance, correlogram, and variogram are equivalent tools for characterizing two-point correlation:

$$C(\mathbf{h}) = \sigma^2 \cdot \rho(\mathbf{h}) = \sigma^2 - \gamma(\mathbf{h})$$

This relation depends on the model decision that the mean and variance are constant and independent of location. These relations are the foundation of variogram interpretation. That is, (1) the "sill" of the variogram is the variance, which is the variogram value that corresponds to zero correlation; (2) the correlation between $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ is positive when the variogram value is less than the still; and (3) the correlation between $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ is negative when the variogram exceeds the sill.

## C. Anisotropy

Spatial continuity depends on direction. Anisotropy in geostatistical calculations is *geometric*, that is, defined by a triaxial Cartesian system of coordinates. Three angles define orthogonal $x$, $y$, and $z$ coordinates and then the components of the distance vectors are scaled by three range parameters to determine the scalar distance, that is,

$$h = \sqrt{\left(\frac{h_x}{a_x}\right)^2 + \left(\frac{h_y}{a_y}\right)^2 + \left(\frac{h_z}{a_z}\right)^2},$$

where $h_x$, $h_y$, and $h_z$ are the components of a vector $\mathbf{h}$ in three-dimensional coordinate space and $a_x$, $a_y$, and $a_z$ are scaling parameters in the principal directions. Contour lines of equal "distance" follow ellipsoids. The use of $z$ for the random variable and a coordinate axis is made clear by context. The three $x$, $y$, and $z$ coordinates must be aligned with the principal directions of continuity. A coordinate rotation may be required.

The directions of continuity are often known through geologic understanding. In case of ambiguity, the variogram may be calculated in a number of directions. A *variogram map* could be created by calculating the variogram for a large number of directions and distances; then, the variogram values are posted on a map where the center of the map is the lag distance of zero.

## D. Variogram Modeling

The variogram is calculated and displayed in the principal directions. These experimental directional variogram points are not used directly in subsequent geostatistical steps such as kriging and simulation; a parametric variogram model is fitted to the experimental points. There are two reasons why experimental variograms must be modeled: (1) there is a need to interpolate the variogram function for $\mathbf{h}$ values where too few or no experimental data pairs are available, and (2) the variogram measure $\gamma(\mathbf{h})$ must have the mathematical property of "positive definiteness" for the corresponding covariance model—that

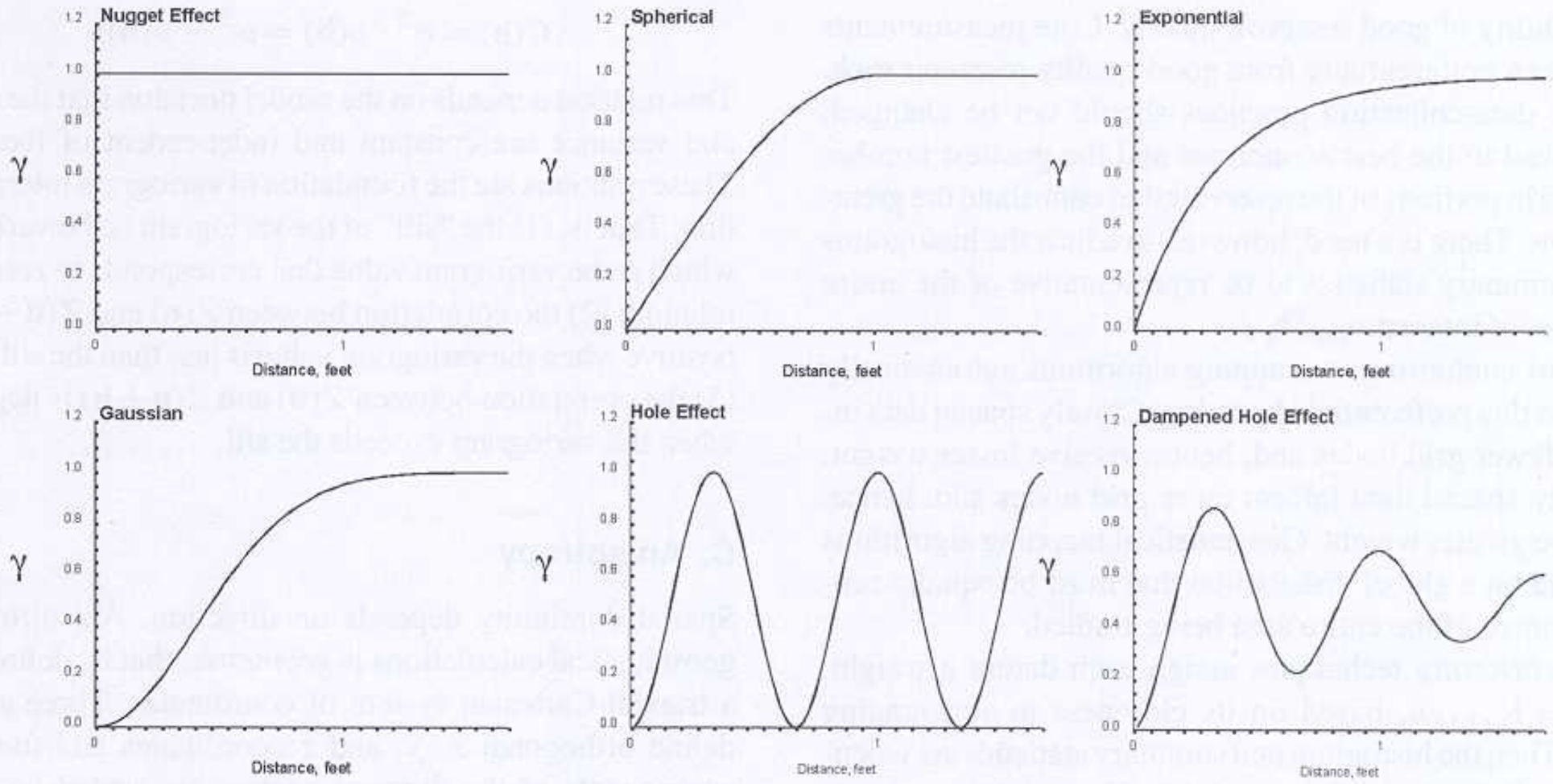



**FIGURE 1** Typical variogram structures that are combined together in nested structures to fit experimental variograms. Anisotropy, that is, different directional variograms, are brought to the same distance units by geometric scaling.

is, we must be able to use the variogram and its covariance counterpart in kriging and stochastic simulation. For these reasons, geostatisticians have fitted sample variograms with specific known positive definite functions such as the spherical, exponential, Gaussian, and hole-effect variogram models (see Fig. 1).

A variogram model can be constructed as a sum of known positive-definite licit variogram functions called nested structures. Each nested structure explains a fraction of the variability. All nested structures together describe the total variability, $\sigma^2$. Interactive software is typically used to fit a variogram model to experimental points in different directions.

## III. SPATIAL REGRESSION OR KRIGING

### A. Point Estimation

An important application of geostatistics is to calculate estimates at unsampled locations. The basic idea is to propose a liner estimate of the residual from the mean:

$$z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha \cdot [z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)],$$

where $z^*(\mathbf{u})$ is an estimate made with $n$ data, $m(\mathbf{u})$ is the mean value known at all locations, and $\lambda_\alpha$, $\alpha = 1, \ldots, n$, are weights that account for how close the $n$ data are to the location being estimated and how redundant the data are with each other. The weights could be assigned inversely proportional to the distance between the data $\mathbf{u}_\alpha$ and location being estimated, $\mathbf{u}$; however, a better procedure is to use the variogram and minimize the error variance.

### B. Simple Kriging

Least-squares optimization has been used for many years. The idea, proposed by early workers in geostatistics, was to calculate the weights to be optimum in a minimum squared error sense, that is, minimize the squared difference between the true value $z(\mathbf{u})$ and the estimator $z^*(\mathbf{u})$. Of course, the true values are known only at the data locations, *not* at the locations being estimated. Therefore, as is classical in statistics, the squared error is minimized in the expected value.

The geostatistical technique known as simple kriging is a least-squares regression procedure to calculate the weights that minimize the squared error. A set of $n$ equations must be solved to calculate the $n$ weights:

$$\sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\beta - \mathbf{u}_\alpha) = C(\mathbf{u} - \mathbf{u}_\alpha), \quad \alpha = 1, \ldots, n.$$

Recall that $C(\mathbf{h}) = \sigma^2 - \gamma(\mathbf{h})$; therefore, knowledge of the variogram model permits calculation of all needed covariance terms. The left-hand side contains all of the information related to redundancy in the data, and the right-hand side contains all of the information related to closeness of the data to the location being estimated. Kriging is the best estimator in terms of minimum error variance.

Kriging is an exact estimator; that is, the kriging estimator at a data location will be the data value. The minimized error variance or *kriging variance* can be calculated for all estimated locations:

$$\sigma_K^2(\mathbf{u}) = \sigma^2 - \sum_{\alpha=1} \lambda_\alpha \cdot C(\mathbf{u} - \mathbf{u}_\alpha),$$

where the kriging variance is the global variance, $\sigma^2$, in the presence of no local data and 0 at a data location. The kriging estimates and kriging variance can be calculated at each location and posted on maps.

### C. Constrained Kriging

The basic estimator written in Section III.A requires the mean $m(\mathbf{u})$ at all locations. A number of techniques have been developed in geostatistics to relax this requirement. Ordinary kriging, for example, assumes that the mean $m$ is constant and unknown. A constraint is added to the kriging equations to enforce the sum of the weights to equal 1, which amounts to estimating the mean at each location. Universal kriging assumes the mean follows a particular parametric shape; the parameters are estimated at each location. These constrained versions of kriging make a different decision regarding stationarity.

### D. Multiple Variables

The term *kriging* is traditionally reserved for linear regression using data with the same variable as that being estimated. The term *cokriging* is reserved for linear regression that also uses data defined on different attributes. For example, the porosity value $z(\mathbf{u})$ may be estimated from a combination of porosity samples and related acoustic impedance values, $y(\mathbf{u})$. Kriging requires a model for the $Z$ variogram. Cokriging requires a *joint* model for the matrix of variogram functions including the $Z$ variogram, $\gamma_Z(\mathbf{h})$, the $Y$ variogram, $\gamma_Y(\mathbf{h})$, and the cross $Z$–$Y$ variogram $\gamma_{Z-Y}(\mathbf{h})$. When $K$ different variables are considered, the covariance matrix requires $K^2$ covariance functions. The inference becomes demanding in terms of data and the subsequent joint variogram modeling; however, cokriging provides the minimum error-variance estimator of the variable at an unsampled location using multiple data variables.

### E. Smoothing

Kriging estimates are smooth. The kriging variance is a quantitative measure of the smoothness of the kriging estimates. There is no smoothing when kriging at a data location, $\sigma_K^2 = 0$. There is complete smoothness when kriging with data far from the location being estimated; the estimate is equal to the mean and the kriging variance is the full variance, $\sigma_K^2 = \sigma^2$. This nonuniform smoothing of kriging is the largest shortcoming of kriging for map making. A map of kriging estimates gives an incorrect picture of variability, and calculated results such as recoverable reserves and flow properties are wrong. Simulation corrects for the smoothing of kriging.

## IV. SIMULATION

### A. Sequential Gaussian Simulation

The idea of simulation is to draw multiple, equally probable realizations from the random function model. These realizatios provide a *joint* measure of uncertainty. Each realization should reproduce (1) the local data at the correct scale and measured precision, (2) the global stationary histogram within statistical fluctuation, and (3) the global stationary variogram or covariance within statistical fluctuation. There is much discussion in the geostatistical literature about different random function models. The most commonly used, however, is the multivariate Gaussian model. The data are first transformed so that the global stationary histogram is Gaussian or normal. Then, all multivariate distributions of $n$ points taken at a time are assumed to follow the mathematically congenial Gaussian distribution. There are many techniques to draw simulations from a multivariate Gaussian random function. The sequential approach gained wide popularity in the 1990s because of its simplicity and flexibility. The sequential Gaussian simulation (SGS) algorithm is as follows.

1. Transform the original $Z$ data to a standard normal distribution (all work will be done in "normal" space). There are different techniques for this transformation. The normal score transformation whereby the normal transform $y$ is calculated from the original variable $z$ as $y = G^{-1}[F(z)]$, where $G(\cdot)$ is the standard normal cumulative distribution function (cdf) and $F(\cdot)$ is the cdf of the original data.
2. Go to a location $\mathbf{u}$ (chosen randomly from the set of locations that have not been simulated yet) and perform kriging to obtain a kriged estimate and the corresponding kriging variance.
3. Draw a random residual $R(\mathbf{u})$ that follows a normal distribution with mean of 0.0 and a variance of $\sigma_K^2(\mathbf{u})$. Add the kriging estimate and residual to get a simulated value. The independent residual $R(\mathbf{u})$ is drawn with classical Monte Carlo simulation.
4. The simulated value is added to the data set and used in future kriging and simulation to ensure that the

variogram between all of the simulated values is correct. A key idea of sequential simulation is to add previously simulated values to the data set.

5. Visit all locations in a random order (return to step 2). There is no theoretical requirement for a random order or path; however, practice has shown that a regular path can induce artifacts. When every grid node has been assigned, the data values and simulated values are back-transformed to real units.

Repeating the entire procedure with a different random number seeds creates multiple realizations. The procedure is straightforward; however, there are a number of implementation issues, including (1) a reasonable three-dimensional model for the mean $m(\mathbf{u})$ must be established, (2) the input statistics must be reliable, and (3) reproduction of all input statistics must be validated.

### B. Alternatives to Sequential Approach

Many algorithms can be devised using the properties of the multi-Gaussian distribution to create stochastic simulations: (1) matrix approaches (LU decomposition), which are not used extensively because of size restrictions (an $N \times N$ matrix must be solved, where $N$ could be in the millions for reservoir applications); (2) turning bands methods, where the variable is simulated on one-dimensional lines and then combined into a three-dimensional model, which is not commonly used because of artifacts; (3) spectral methods using fast Fourier transforms can be CPU-fast, but the grid size $N$ must be a power of 2 and honoring conditioning data requires an expensive kriging step; (4) fractals, which are not used extensively because of the restrictive assumption of self-similarity, and (5) moving-average methods, which are used infrequently due to CPU requirements.

### C. Indicator Simulation

The aim of the indicator formalism for categorical variables is to simulate the distribution of a categorical variable such as rock type, soil type, or facies. A sequential simulation procedure is followed, but the distribution at each step consists of estimated probabilities for each category: $p^*(k)$, $k = 1, \ldots, K$, where $K$ is the number of categories. The probability values are estimated by first coding the data as indicator or probability values—that is, an indicator is 1 if the category is present, and 0 otherwise. The Monte Carlo simulation at each step is a discrete category. Requirements for indicator simulation include $K$ variograms of the indicator transforms and $K$ global proportions.

## V. SPECIAL TOPICS

### A. Object-Based Modeling

Object-based models are becoming popular for creating facies models in petroleum reservoirs. The three key issues to be addressed in setting up an object-based model are (1) the geologic shapes, (2) an algorithm for object placement, and (3) relevant data to constrain the resulting realizations. There is no inherent limitation to the shapes that can be modeled with object-based techniques. Equations, a raster template, or a combination of the two can specify the shapes. The geologic shapes can be modeled hierarchically—that is, one object shape can be used at large scale and then different shapes can be used for internal small-scale geologic shapes. It should be noted that object-based modeling has nothing to do with object-oriented programming in a computer sense.

The typical application of object-based modeling is the placement of abandoned sand-filled fluvial channels within a matrix of floodplain shales and fine-grained sediments. The sinuous channel shapes are modeled by a one-dimensional centerline and a variable cross section along the centerline. Levee and crevasse objects can be attached to the channels. Shale plugs, cemented concretions, shale clasts, and other non-net facies can be positioned within the channels. Clustering of the channels into channel complexes or belts can be handled by large-scale objects or as part of the object-placement algorithm.

Object-based facies modeling is applicable to many different depositional settings. The main limitation is coming up with a suitable parameterization for the geologic objects. Deltaic or deep-water lobes are one object that could be defined. Eolean sand dunes, remnant shales, and different carbonate facies could also be used.

### B. Indicator Methods

The indicator approach to categorical variable simulation was mentioned earlier. The idea of indicators has also been applied to continuous variables. The key idea behind the indicator formalism is to code all of the data in a common format, that is, as *probability* values. The two main advantages of this approach are (1) simplified data integration because of the common probability coding, and (2) greater flexibility to account for different continuity of extreme values. The indicator approach for continuous data variables requires significant additional effort versus Gaussian techniques.

The aim of the indicator formalism for continuous variables is to estimate directly the distribution of uncertainty $F^*(z)$ at unsampled location $\mathbf{u}$. The cumulative distribution function is estimated at a series of threshold values:

$z_k$, $k=1,\ldots,K$. The indicator coding at location $\mathbf{u}_\alpha$ for a particular threshold $z_k$ is

$$i(\mathbf{u}_\alpha; z_k) = \text{Prob}[Z(\mathbf{u}_\alpha) \leq z_k]$$
$$= \begin{cases} 1, \text{ if } z(\mathbf{u}_\alpha) \leq z_k, \\ 0, \text{ otherwise.} \end{cases}$$

All hard data $z(\mathbf{u}_\alpha)$ are coded as discrete zeros and ones. Soft data can take values between zero and one. The indicator transform for a threshold less than the data value is zero, since there is no probability that the data value is less than the threshold; the indicator transform for a very high threshold is one, since the data value is certainly less than the threshold.

The cumulative distribution function at an unsampled location at threshold $z_k$ can be estimated by kriging. This "indicator kriging" or IK requires a variogram measure of correlation corresponding to each threshold $z_k$, $k=1,\ldots,K$. The IK process is repeated for all $K$ threshold values that discretize the interval of variability of the continuous attribute $Z$. The distribution of uncertainty, built from assembling the $K$ indicator kriging estimates, can be used for uncertainty assessment or simulation.

### C. Simulated Annealing

The method of simulated annealing is an optimization technique that has attracted significant attention. The task of creating a three-dimensional numerical model that reproduces some data is posed as an optimization problem. An objective function measures the mismatch between the data and the numerical model. An initial random model is successively perturbed until the objective function is lowered to zero. The essential contribution of simulated annealing is a prescription for when to accept or reject a given perturbation. This acceptance probability distribution is taken from an analogy with the physical process of annealing, where a material is heated and then slowly cooled to obtain low energy.

Simulated annealing is a powerful optimization algorithm that can be used for numerical modeling; however, it is more difficult to apply than kriging-based methods because of difficulties in setting up the objective function and choosing many interrelated parameters such as the annealing schedule. Therefore, the place of simulated annealing is not for conventional problems where kriging-based simulation is adequate. Simulated annealing is applicable to difficult problems that involve (1) dynamic data, (2) large-scale soft data, (3) multiple-point statistics, (4) object placement, or (5) special continuity of extremes.

### D. Change of Support

Reconciling data from different scales is a long-standing problem in geostatistics. Data from different sources, including remotely sensed data, must all be accounted for in the construction of a geostatistical reservoir model. These data are at vastly different scales, and it is wrong to ignore the scale difference when constructing a geostatistical model. Geostatistical scaling laws were devised in the 1960s and 1970s primarily in the mining industry, where the concern was mineral grades of selective mining unit (SMU) blocks of different sizes. These techniques can be extended to address problems in other areas, subject to implicit assumptions of stationarity and linear averaging.

The first important notion in volume-variance relations is the spatial or dispersion variance. The dispersion variance $D^2(a, b)$ is the variance of values of volume $a$ in a larger volume $b$. In a geostatistical context, all variances are dispersion variances. A critical relationships in geostatistics is the link between the dispersion variance and the average variogram value:

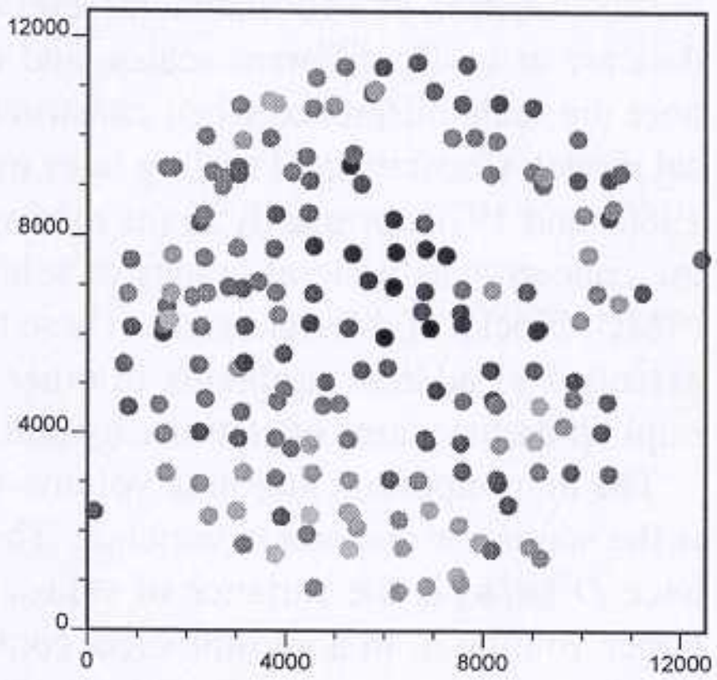
$$D^2(a, b) = \bar{\gamma}(b, b) - \bar{\gamma}(a, a).$$

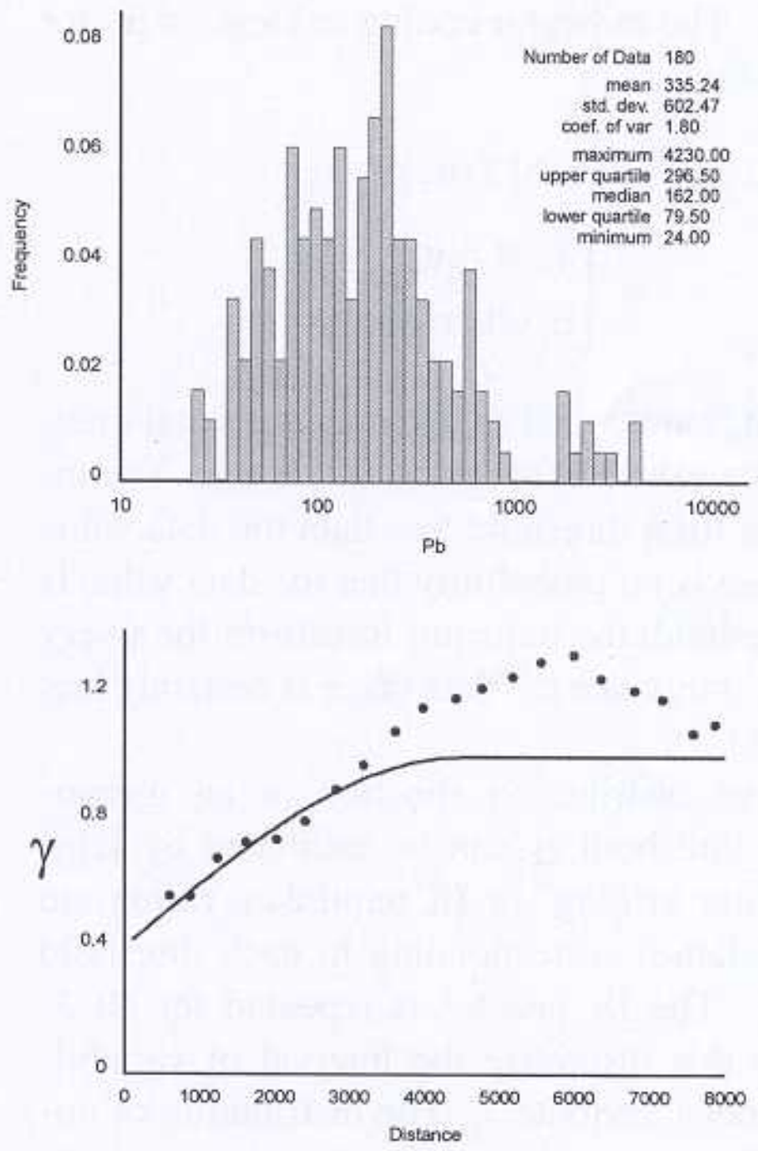
This tells us how the variability of a variable changes with the volume scale and variogram. The variability of a variable with high short-scale variability decreases quickly, since high and low values average out.

## VI. APPLICATIONS AND EXAMPLES

### A. Environmental

Figure 2 illustrates some of the geostatistical operations applied to characterize the spatial distribution of lead contamination over a 12,500-ft$^2$ area. There are five parts to Fig. 2: (1) the upper left shows the location map of the 180 samples—there is no evident clustering that would require declustering; (2) the equal-weighted histogram, at the upper right, shows the basic statistics related to the measurements—note the logarithmic scale; (3) the variogram, shown below the histogram, is of the normal scores transform of the lead data—about 40% of the variability is at very short distances and the remaining 60% of the variability is explained over 4500 ft—the black dots are the experimentally calculated points and the solid line is the fitted model; (4) a map of kriging estimates on a 100-ft$^2$ grid is shown at the lower left—note the smoothness of the kriging estimates; and (5) a sequential Gaussian simulation (SGS) realization is shown at the lower right—this realization reproduces the 180 sample data, the input histogram, and the variogram model. A set of realizations could be used to assess the probability that each location exceeds some critical threshold of lead concentration.



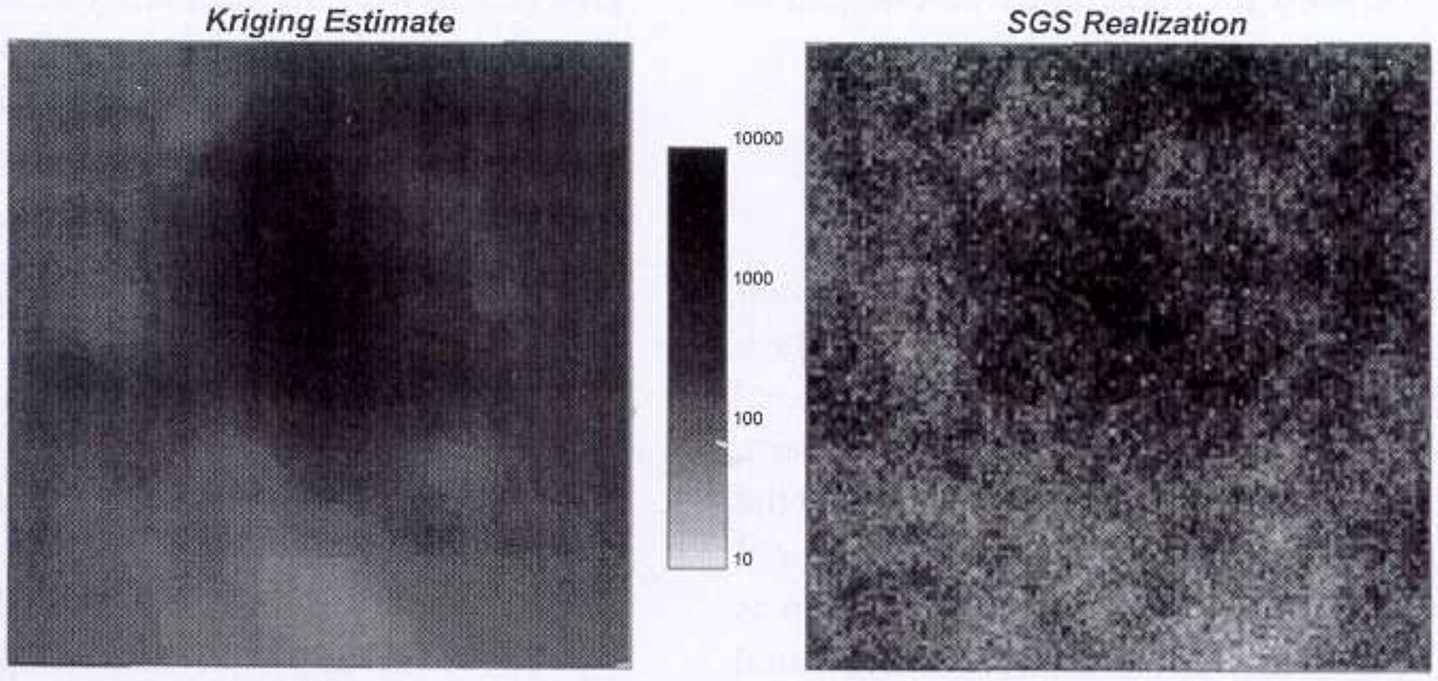



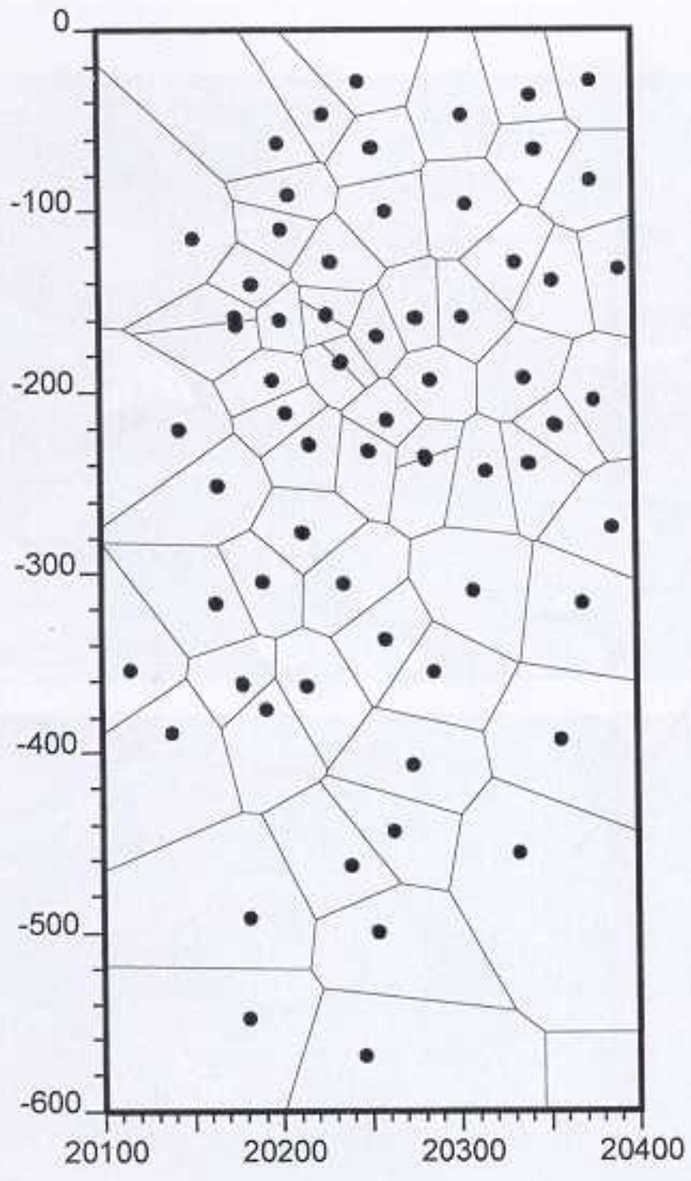

**FIGURE 2** Location map (distance units in feet) of 180 samples, histogram of lead concentration, variogram of the normal scores transform (hence the *sill* value of 1.0), a map of kriging estimates on a 100-ft$^2$ grid and an SGS realization over the same domain.

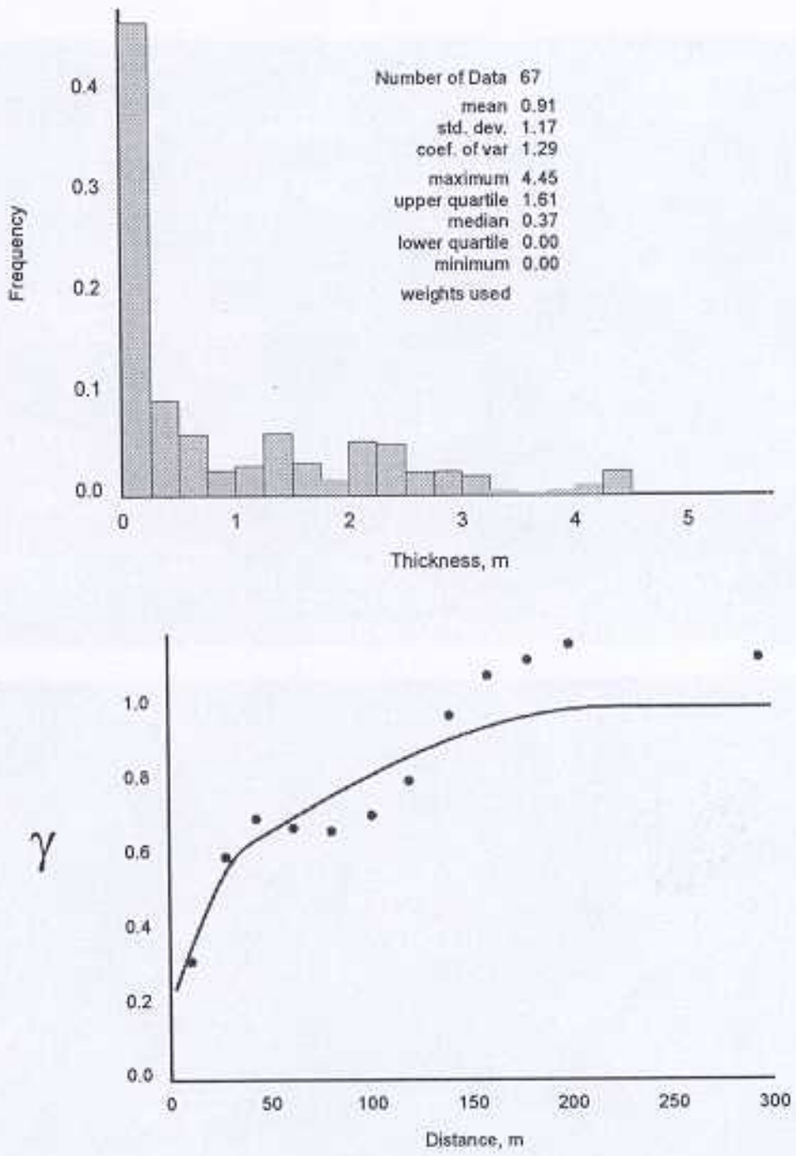
## B. Mining

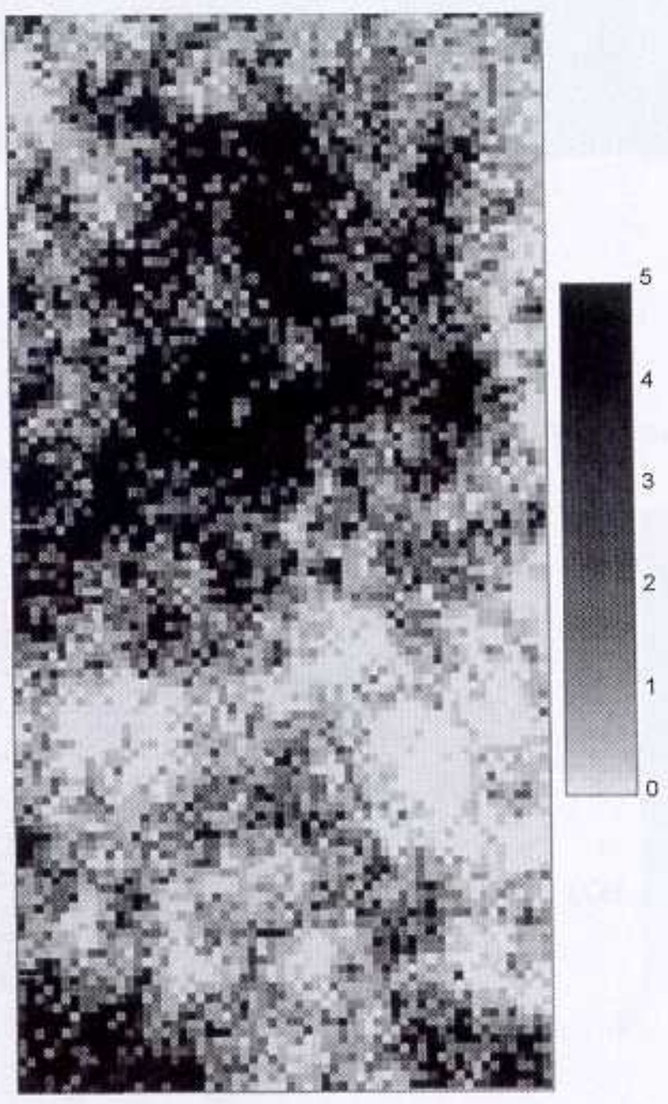
Figure 3 illustrates an example application to a vein-type mineral deposit. The cross-sectional view at the upper left is a vertical cross section facing west; the vertical coordinate is meters below the surface. The drillhole intersections are clustered in the thickest part of the vein. The polygonal areas of influence plotted on the location map are used for declustering weights. The histogram at the upper right of the figure considers the declustering weights. The variogram is shown below the histogram. Two nested structures were used to fit this variogram. One sequential Gaussian realization is shown at the lower left; 150 realizations were generated. The probability of exceeding 1-m thickness is plotted at the lower right. The black locations are where the vein is measured to be greater than 1 m in thickness (probability of 1), and the white locations are where the vein is measured to be less than 1 m (probability of 0).

## C. Petroleum

The profile of porosity and permeability from two wells from an offshore petroleum reservoir are shown at the



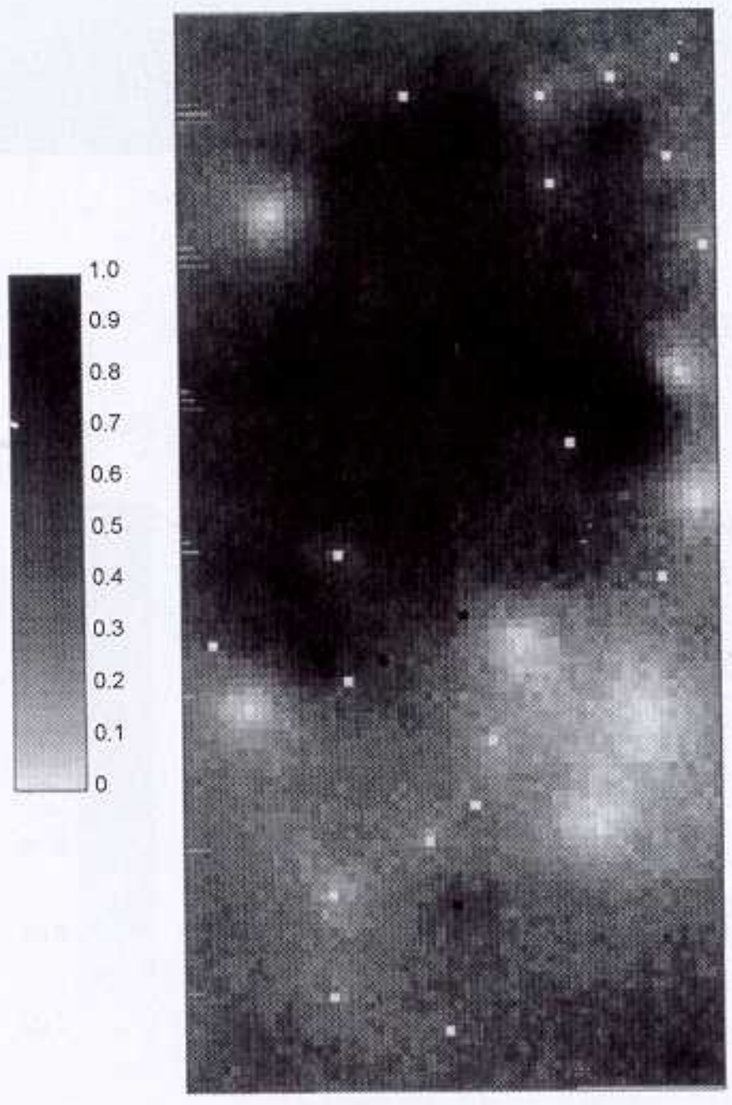




**FIGURE 3** Location map (distance units in meters) of 67 drillholes with polygonal areas of influence for declustering weights, histogram of vein thickness, variogram of the normal scores transform, an SGS realization over the same domain, and the probability to exceed 1.0-m thickness calculated from 100 realizations.

bottom of Fig. 4. A porosity and permeability realization are shown at the top. Simulation of porosity and permeability were done simultaneously to reproduce the correlation between these two variables. The vertical variograms were calculated and modeled easily; however, the horizontal variograms are impossible to discern from two wells. A 50:1 horizontal-to-vertical anisotropy was considered from analog data.

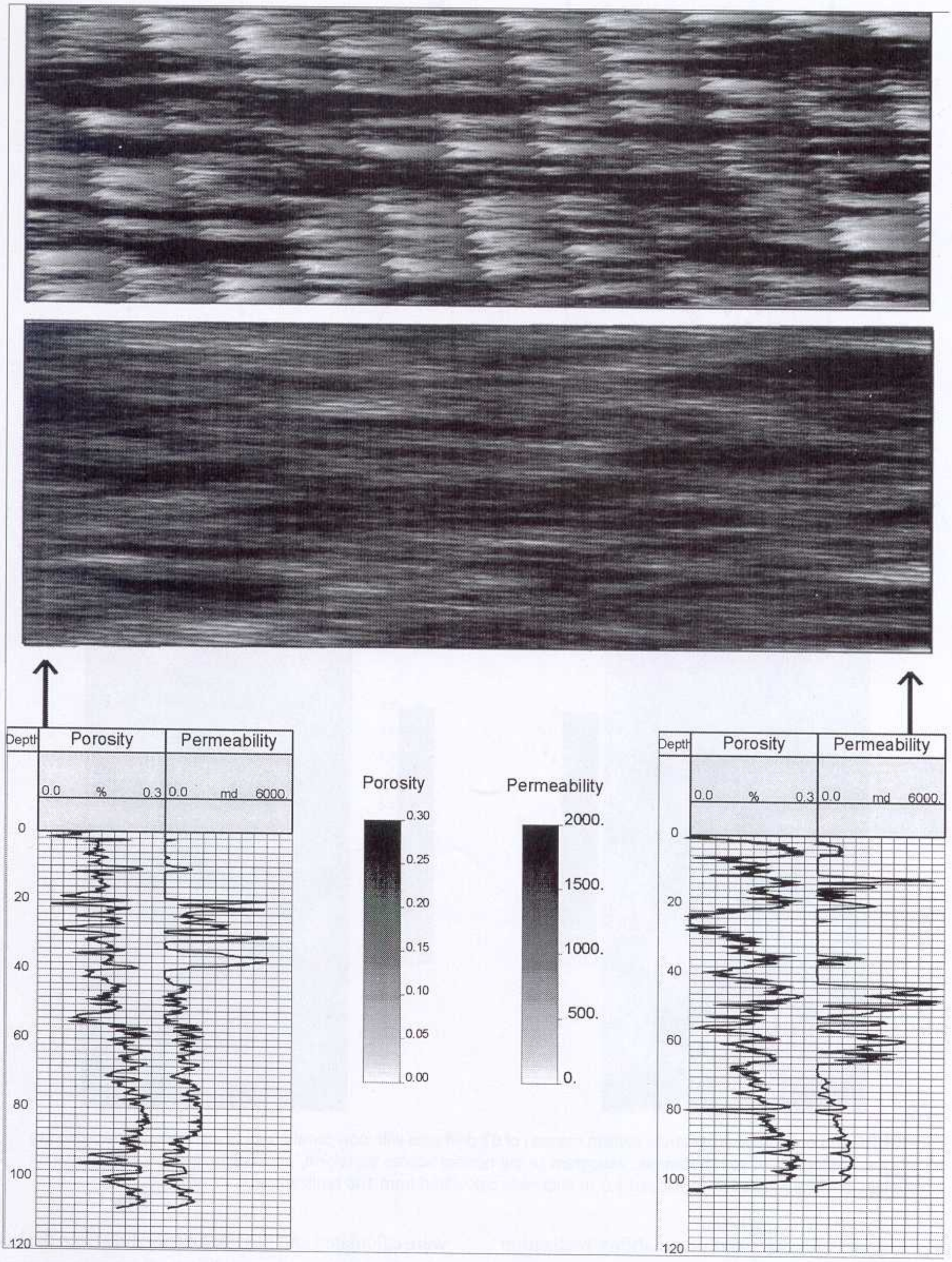



**FIGURE 4** Permeability realization (top) and porosity realization (middle) constrained to two wells 600 m apart (shown at the bottom).

## SEE ALSO THE FOLLOWING ARTICLES

MINING ENGINEERING • ORE PETROLOGY • PETROLEUM GEOLOGY • STATISTICS, FOUNDATIONS

## BIBLIOGRAPHY


Chiles, J. P., and Delfiner, P. (1999). "Geostatistics: Modeling Spatial Uncertainty" (Wiley Series in Probability and Statistics, Applied Probability and Statistics), Wiley, New York.

Cressie, N.(1991). "Statistics for Spatial Data," Wiley, New York.

David, M. (1977). "Geostatistical Ore Reserve Estimation," Elsevier, Amsterdam.

Deutsch, C. V., and Journel, A. G. (1997). "GSLIB: Geostatistical Software Library," 2nd ed., Oxford University Press, New York.

Goovaerts, P. (1997). "Geostatistics for Natural Resources Evaluation," Oxford University Press, New York.

Isaaks, E. H., and Srivastava, R. M. (1989). "An Introduction to Applied Geostatistics," Oxford University Press, New York.

Journel, A. G., and Huijbregts, C. (1978). "Mining Geostatistics," Academic Press, London.

Matheron, G. (1971). "The theory of regionalized variables and its applications," *Cahiers du CMM* **5,** Ecole Nationale Superieure des Mines de Paris, Paris.

Ripley, B. D. (1981). "Spatial Statistics," Wiley, New York.