Mathematical Geology, Vol. 18, No. 1, 1986

Geostatistics: Models and Tools for the Earth Sciences[1]

A. G. Journel[2]

The probability construct underlying geostatistical methodology is recalled, stressing that stationary is a property of the model rather than of the phenomenon being represented. Geostatistics is more than interpolation and kriging(s) is more than linear interpolation through ordinary kriging. A few common misconceptions are addressed.

KEY WORDS: random function model, stationarity, interpolation, local accuracy, conditional probability.

FOREWORD

This presentation of geostatistics has been triggered by two communications recently published in Mathematical Geology, cf Shurtz (1985) and Philip and Watson (1986). These communications attempt to demolish the whole of geostatistics, finding little or no value in any of its contributions.

In the face of such a systematic disparagement based on a few genuine concerns and a lot of convenient misconceptions, one's first reaction is to ignore the flash in the pan, or just appreciate the fine polemical construction, counting such "confettis" as "grotesque—limeade—idiot—fool—sham—" and the number of self-quotes yet to be published.

But not to answer may appear like hedging and could leave the way to another discussant of the polemical kind, resulting in even more confusion. A positive answer may consist of writing yet another presentation of geostatistics insisting particularly on the probabilistic construct and clarifying some common misconceptions which, possibly, could have generated such controversy.

[1]Manuscript received 22 July 1985; accepted 22 July 1985.
[2]Department of Applied Earth Sciences, Stanford University, Stanford, California 94305.

0882-8121/86/0100-0119$05.00/0

INTRODUCTION

Geostatistics has been defined commonly as the application of the "Theory of Regionalized Variables" to the study of spatially distributed data. From a theoretical standpoint, little is new in this theory, which borrows most of its models and tools from the concept of stationary random functions and techniques of analysis of variance and generalized least-squares prediction. This fact is recognized to a certain degree in all geostatistical textbooks.

If stochastic processes, moving averages, and generalized least-squares predictors actually were used under various names in time series analyses of pseudo-periodic phenomena, with emphasis on spectral analysis, the multidimensional random function model was virtually unknown to practitioners in the earth sciences as late as the 1960s.

We wish to emphasize here that major differences exist between a one-dimensional time series, usually sampled on a regular grid and a two- or three-dimensional realization usually sampled at irregular, if not preferential, locations. First, no unique definition of past, present, and future is found; second, the problem of anisotropy appears with paramount importance; third, multidimensional earth sciences phenomena are not all seen as combinations of periodic pulses or wavelengths. Last, in a multidimensional space, emphasis is usually shifted from extrapolation to interpolation with data surrounding the point or region to be estimated.

The contribution of geostatistics was to use the random function (RF) model, thus capitalizing on existing probabilistic tools, and customizing them when necessary. The novelty of geostatistics resides not in the model and tools being used, but in the analysis of the specifics of some earth sciences problems and their expression in terms allowing usage of these tools.

Did Kolmogorov, Feller, Wiener, and other luminaries know about the specifics of ore reserve assessment and that it calls for much more than interpolation? Were they aware of the fundamental importance of notions such as the change of support and spatial smoothing? The unfortunate tendency of some scholars in geostatistics to put new names on old hats should not mask their real contribution in the diffusion of probabilistic thinking among practitioners.

Geostatistics is a methodology, a tool box and a machine tool, based on essentially one model, the random function. To think that it is some kind of descriptive natural science such as crystallography or hydrothermal geochemistry would lead to severe disappointments. A random function, characterized only by its bivariate distribution or a few moments thereof, cannot claim to have any genetic significance. However, the RF model provides the most complete set of tools yet available to apprehend various facets of a spatially distributed data set. If a more flexible and user-friendly set of tools were developed, I have little doubt that engineers, including this author, would switch to it.

A model is no scientific theory, it needs no a priori justification, and it can only be refuted a posteriori if proven to be inadequate for the goal at hand.

We may now redefine geostatistics as a branch of statistics dealing with spatial phenomena.

THE PROBABILITY CONSTRUCT

Consider the spread over an area A of an attribute z. This attribute z is measured over a support of constant geometry and volume, say a core grade. Perfect knowledge of the deterministic field A would consist in an exhaustive sampling which is always possible, if not practical, because both A and the data support are of finite size.

By a slight abuse of language, and to allow utilization of spatial integrals instead of cumbersome discrete summations, we shall speak of an attribute value $z(x)$ measured at a "point" of coordinates $x \in A$.

Information available over A is seen usually as a set of n data values

$$S_n = \{z(x_\alpha),\ \alpha = 1, \ldots, n\}.$$

Any probability construct requires the definition of a population. The population considered here is the finite (although extremely large) set of all possible measurements $z(x)$ that can be made over area A, a sample of which is S_n. Note that area A may be a subzone of a larger field D, say a deposit, i.e., other data may be available outside A. The decision to limit the population, hence all further statistics, to area A is subjective, based on such criteria as geological homogeneity, goal of the study, and availability of sufficient data.

Any actual value $z(x)$, $x \in A$, is interpolated as a realization (outcome) of a random variable (RV) $Z_A(x)$, $x \in A$. The finite set of all such random variables $\{Z_A(x),\ x \in A\}$ is called a random function (RF) and denoted simply by $Z_A(x)$.

The RF $Z_A(x)$ has no providential, nor genetic connotation; it is not some kind of physical law that has generated mineralization over area A, among many others. The RF $Z_A(x)$ is simply a model reproducing some aspects (statistics) of the information available, to be used for understanding the very particular and *unique* area A.

The question is which aspects of the data set are deemed important for the goal at hand and should be reproduced by the model? The more such aspects are retained, the better the model, provided that:

(i) these aspects are considered reliable enough; the model is supposed to represent the entire population, not any futile detail of the data set

(ii) the model does not become so complex as to hamper its usage.

Once a model is constructed, its usage draws from all aspects of it and particularly from those that were not identified to any data information. When drawing a conclusion from a model, one always should be extremely careful in evaluating how much of that conclusion is coming from the initial data and how much from the model construct.

In geostatistics, spatial dependence between two attribute values $z(x)$, $z(x')$, located at different locations x, $x' \in A$, is considered a most important aspect. All geostatistical models strive to model that spatial dependence to various degrees of completeness.

By construction of the model, the component RVs $Z_A(x)$, $x \in A$, are all attached to a common marginal cumulative probability distribution function (cdf) defined as the exhaustive proportion of actual values $z(x) \leq z$ within A

$$F_A(z) = \text{Prob}\{Z_A(x) \leq z\} = \frac{1}{\{A\}} \int_A i(u; z)\, du \qquad \text{for all } x \in A \tag{1}$$

with: $\{A\}$ being the measure of A, and the indicator function of z defined as

$$i(x; z) = \begin{cases} 0 & \text{for } z < z(x) \\ 1 & \text{otherwise} \end{cases} \tag{1a}$$

Similarly, by construction of the model, any two component RVs $Z_A(x)$, $Z_A(x + h)$, are linked through a bivariate cdf identified to the exhaustive proportion, within A, of actual values $z(x) \leq z$ and $z(x + h) \leq z'$

$$\begin{aligned} F_A(h; z, z') &= \text{Prob}\{Z_A(x) \leq z;\, Z_A(x + h) \leq z'\} \\ &= \frac{1}{2\{A(h)\}} \left[\int_{A \cap A_{-\mathbf{h}}} i(u; z) \cdot i(u + h; z')\, du \right. \\ &\quad \left. + \int_{A \cap A_{+h}} i(u; z) \cdot i(u - h; z')\, du \right] \end{aligned} \tag{2}$$

$$\text{with } A(h) = A \cap A_{-\mathbf{h}} \quad \text{and} \quad \{A(h)\} = \{A(-h)\}$$

When the modulus of vector h is small with regard to the dimension of A in the direction of h, the representativity area of $F_A(h; z, z')$, i.e., $(A \cap A_{-\mathbf{h}}) \cup (A \cap A_{+\mathbf{h}})$ (cf hatched area on Fig. 1) is equal to A.

Stationarity, a Model Property

The two integral constructs (1) and (2) amount to building the component RVs $Z_A(x)$, $x \in A$ stationary up to their bivariate distributions. This stationary property is a constitutive property of the RF model $Z_A(x)$, and as such need not (nor can it) be a priori checked or refuted. Nor is the decision to average statistics over a given area A irreversible if more data becomes available or if the goal of the study changes.

The construct $F_A(z)$ defined by the integral (or summation) (1) is in no way original. In virtually any statistical endeavor, a population is defined a priori—here A—and the distribution is identified to exhaustive proportions over that

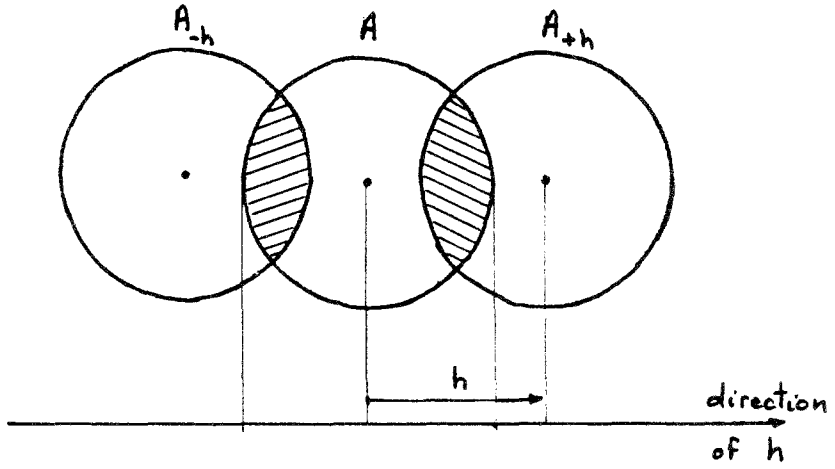

Fig. 1. Representativity area of bivariate statistics A_{-h} and A_{+h} are translations of the area A by the vector $-h$ and $+h$, respectively. $A \cap A_{-h}$ is the intersection of A with its translated A_{-h}. For h small, typically less than half the dimension of A in the direction of h

$$(A \cap A_{-h}) \cup (A \cap A_{+h}) = A$$

population. The construct $F_A(h; z, z')$ is merely an extension of that idea, where bivariate probabilities are identified to exhaustive proportions of bivariate events from the population. This additional construct $F_A(h; z, z')$ allows modeling of the capital notion of spatial dependence.

The prior decision about the "stationarity area" A is in all points similar to the prior decision of a demographer to define the "stationarity population" of American citizens of Caucasian origin. Should that prior decision be tested? Does that decision imply that any two citizens of Caucasian origin are either identical or unrelated? That prior decision is not cast forever either, and would change if the goal of the study changes, or if more data becomes available allowing consideration of Caucasian subpopulations. Last, that demographer need not think that citizens of Caucasian origin were procreated by some Aryan random function.

Global Statistics and Local Applications

Any statistics, be it a mean, a histogram, a variogram, or a bivariate distribution of type (2), is necessarily global in the sense that it is representative of a whole population, not specific to any individual or pair of individuals. Still these global statistics can be used for prediction or estimation of a specific individual or location.

Take the example of the bivariate probability $F_A(h; z, z')$ defined in expres-

sion (2); it is by construction, a statistic averaged over the entire area A. Nevertheless, it does provide the probability that an unknown value $z(x)$, at any specific location $x \in A$, be less than any threshold value z given that a datum value $z(x + h) = z_\alpha$ has been observed at a nearby location $x + h \in A$. Let Prob $\{Z_A(x) \leq z | Z_A(x + h) = z_\alpha\}$ be that probability. Again this probability has no magic, nor genetic origin; it should be placed in the context of the probability construct (2) and understood as:

"Define within A the subpopulation of attribute values $z(x)$ such that their h- neighbor $z(x + h) = z_\alpha$, the probability Prob $\{Z_A(x + h) = z_\alpha\}$ is none other than the proportion of such attributes $z(x)$ with value $\leq z$."

This probability is a function of the threshold value z, vector h (direction and modulus) characterizing the data configuration, *and* the datum (or data) value z_α.

If a roulette gambler has observed over a long series (area A) a large probability for one 0 following another, wouldn't he or she bet on a second 0 after observing the first?

STATISTICAL INFERENCE

Consider the RF model $Z_A(x)$, or more precisely the class of RF models, characterized by the bivariate cdf $F_A(h; z, z')$ defined by expression (2).

The problem is not the inference of an elusive bivariate distribution of a no less elusive "generator" RF, but the estimation of well-defined spatial integrals of type (1) or (2) from discrete sums using the finite data set $S_n = \{z(x_\alpha), x_\alpha \in A, \alpha = 1, \ldots, n\}$.

If n data locations are spaced regularly over area A, an estimate of the spatial integral $F_A(h; z, z')$ is

$$F_A^*(h; z, z') = \frac{1}{n(h)} \sum_{\alpha=1}^{n(h)} i(x_\alpha; z) \cdot i(x_\alpha + h; z') \tag{3}$$

where $\{i(x_\alpha; z), i(x_\alpha + h; z'), \alpha = 1, \ldots, n(h) < n\}$ are indicator data—see definition (1a)—attached to the $n(h)$ pairs of z data approximately distant of vector $\pm h$.

Correspondingly, bivariate moments such as the variogram

$$2\gamma_A(h) = E\{[Z_A(x + h) - Z_A(x)]^2\}$$

would be estimated by the discrete sum

$$2\gamma_A^*(h) = \frac{1}{n(h)} \sum_{\alpha=1}^{n(\mathbf{h})} [z(x_\alpha) - z(x_\alpha + h)]^2 \tag{4}$$

Remarks

If the $n(h)$ data pairs are not spread evenly over area A, more sophisticated estimates than the equal-weighted averages (3) or (4) should be considered (see Omre, 1984).

At this stage, the only relevant error is the representativity error $[\gamma_A(\mathbf{h}) - \gamma_A^*(h)]$ or, more generally $[F_A(h; z, z') - F_A^*(h; z, z')]$. Clearly both the estimate of $\gamma_A^*(h)$ and the magnitude of the potential error $\gamma_A(h) - \gamma_A^*(h)$ cannot be inferred simultaneously using (only) the same data set. However, if one uses only part of the data available to construct the $\gamma_A^*(h)$s, the remaining data can be used to evaluate the potential error. In such an exercise, one should keep in mind that even exhaustive spatial averages $\gamma_A(h)$ or $F_A(h; z, z')$ are still models delivering but a limited image of the phenomemon under study.

The experimental semivariogram graph, or cloud $[\gamma_A^*(h), h]$ is fitted by a usually continuous model $\gamma_A(h)$, except at the origin ($h = 0$). As long as it meets some evident conditions of positive definiteness, the analytical expression of the interpolation function is irrelevant. The spherical variogram model is a convenient, polynomial-type, positive definite (in its covariance form), interpolation function with *no* a priori genetic meaning. One could as well use a series of exponentials.

The reason why geostatisticians prefer the variogram tool to the covariance tool is simple and practical. The covariance tool would require prior estimation of the mean attribute over area A,

$$z_A = \frac{1}{\{A\}} \int_A z(x)\, dx$$

which may be a goal to the study.

The variogram model $2\hat{\gamma}_A(h)$ is more than estimation of the correlation range, if only for the single reason that the experimental cloud $[\gamma_A^*(h), h]$ may not present any plateau for any distance h within A, and would thus be modeled by a steadily increasing function such as: $\gamma_A(h) = |h|$. The most important features of a variogram model are its behavior at the origin (differentiability or discontinuity) and its pattern of anisotropy. Indeed, in three-dimensional space, the model is a three-dimensional function $\hat{\gamma}_A(h_1, h_2, h_3)$ usually anisotropic, e.g.

$$\hat{\gamma}_A(1, 0, 0) \neq \hat{\gamma}_A(0, 1, 0) \neq \hat{\gamma}_A(0, 0, 1)$$

Whether the semivariogram, $\gamma_A(h)$ or $\hat{\gamma}_A(h)$, has a range or not is irrelevant to most geostatistical calculations, including kriging. To limit the search for data to the range of the variogram model, when doing ordinary kriging, may be a convenience to reduce computation but it has no theoretical justification.

Although models $\hat{\gamma}_A(h)$ and/or $\hat{F}_A(h; z, z')$ do not carry any a priori genetic significance, they may signal features that may have been overlooked, such as

anisotropies *and* spatial heterogeneities, which put us back to the discussion on stationarity. Consideration of the experimental semivariogram cloud $[\gamma_A^*(h, h)]$ or, better, the experimental h scattergram $[z(x_\alpha), z(x_\alpha + h), \alpha = 1, \ldots, n(h)]$, often points out the necessity of further splitting the stationarity area chosen initially.

Any serious practitioner of geostatistics would expect to spend a good half of his or her time looking at all faces of a data set, relating them to various geological interpretations, prior to any kriging; he or she may even decide not to do any kriging!

NONSTATIONARY MODELS

The nonstationary models used in geostatistics are, in fact, the association of a deterministic component accounting for the nonstationarity and a stationary residual RF model.

The model underlying the techniques of trend surface analysis and "universal" kriging provides a good example

$$Z(x) = m(x) + R(x), \tag{5}$$

$$\text{where } m(x) = \sum_{l=1}^{L} a_l f_l(x)$$

is a deterministic trend, an unknown combination (the coefficients a_ls are unknown) of known functions, the $f_l(x)$s. The deterministic component $m(x)$ is usually identified as the nonstationary expected value of the RF model.

$$m(x) = E\{Z(x)\}$$

$R(x)$ is a stationary RF modeling the residual. Its characteristics are either assumed (e.g., $R(x)$ is Gaussian with no correlation) or inferred from the data.

An alternative model could be to replace the deterministic component $m(x)$ by a strongly continuous RF model (e.g., with a Gaussian variogram and a large range).

The decision to consider the two-fold model (5), again cannot be refuted a priori, but should be documented and will be judged on its efficiency at solving the problem at hand.

The choice of the deterministic functions $f_l(x)$ may be based on some knowledge about the physics of the problem—e.g., sonic signals are sine waves with known frequency range—but most often they are chosen on simplicity grounds—e.g., monomials to a given order. In the latter case, model (5) is nonunique, not a priori refutable, but again this can be said about any model.

Inference Problems

A nonstationary RF mdoel of type (5) poses serious problems for statistical inference. Indeed within the predefined R-stationarity area A, only z data are available usually, not the R data required to infer characteristics of the stationary RF model $R(x)$. Various solutions out of this problems are:

(i) A cheap solution consists of considering a RF $R(x)$ with no spatial autocorrelation; this is the model underlying traditional least-squares trend fitting and is adequate if residual variance (not explained by the trend) is small.

(ii) Find subareas and/or directions within area A where the trend contribution can be considered as approximately constant: $m(x) \simeq a_1$, for $x \in A_1 \subset A$. Within such subareas or along such directions, experimental z increments identify R increments; hence, the R variogram model can be inferred from the experimental z variogram. This is the inference procedure underlying the "universal" kriging algorithm.

(iii) Infer directly "generalized covariance" of the nonstationary RF model $z(x)$, using the more esoteric formalism of "intrinsic RFs of order k" (IRF k) if the trend $m(x)$ is a polynomial of order k. A generalized covariance is a covariance model that allows calculation of the variance of any linear combination of z data that filters the trend $m(x)$. Inference of the generalized covariance model requires defining such linear combinations; this essentially is not different from the previous inference technique which requires finding areas and/or directions filtering out the trend.

Some Incidental Remarks

The fact that "universal" kriging and formalism of IRF k result in exactly the same system of linear equations is little publicized, allowing the latter to appear more up-to-date. The formalism of IRF k allows, supposedly, the automatic inference of both order k of the polynomial trend and the corresponding generalized covariance, and is unquestionably at the origin of a substantial software business. However, this author is less than impressed by the performance of IRF k when used in the automatic mode.

Also, the experience of this author is that "universal" kriging, or for that matter IRF k, is needed only in cases of extrapolation when the point being estimated is beyond the correlation range of any datum.

GEOSTATISTICS IS MUCH MORE THAN INTERPOLATION

If all attribute values $z(x)$ for all locations $x \in A$ were known, all problems would be solved. Building on this triviality, all estimation problems are often

naively thought to reduce to the interpolation of any unknown value $z(x)$ from neighboring data. For example, a good estimate of any known function $\varphi[z(x)]$ of the unknown $z(x)$ is thought to be $\varphi[z^*(x)]$. Unfortunately, the fact that a local estimate $z^*(x)$ is "good" in some sense—not always clearly defined—does not entail that the estimates $\varphi[z^*(x)]$, or more generally $\varphi[z^*(x), x \in A]$, are good in the same sense, or any good at all.

In fact, one can hardly imagine a real problem for which the point estimate $z^*(x)$ is a real goal instead of an intermediary step toward a tougher estimation problem. Moreover, that step often turns out to be in the wrong direction!

The following is but a short list of some typical problems, calling for much more than interpolation, for which geostatistics has brought a solution.

Estimation of Recoverable Reserves

More generally, estimation of reserves involves characterization of excursions above a threshold value z_0. Consider the paving of area A, or of any panel $P \subset A$, by N blocks $v(x_k)$, $k = 1, \ldots, K$ of same size $\{v\}$

$$P = \bigcup_{k=1}^{K} v(x_k) \subset A$$

The block mean grades are

$$z_v(x_k) = \frac{1}{\{v\}} \int_{v(x_k)} z(u)\, du, \qquad k = 1, \ldots, K \tag{6}$$

The problem at hand is to select those blocks with mean grade greater than a given threshold z_0, for processing (in the mining industry) or for cleaning (in pollution control).

Short of exhaustive sampling, the selection has to be made on the estimated values $z_v^*(x_k)$. Once a block has been selected, on the basis that $z_v^*(x_k) > z_0$, its actual mean grade $z_v(x_k)$ is processed. Total yield (or recovery) from panel P is proportional to the following sum

$$\varphi[z_v(x_k), k = 1, \ldots, K] = \sum_{k=1}^{K} j_v^*(x_k; z_0) \cdot z_v(x_k) \qquad \text{with} \tag{7}$$

$$j_v^*(x_k; z_0) = \begin{cases} 1 & \text{for } z_v^*(x_k) > z_0 \\ 0 & \text{otherwise} \end{cases}$$

Note that the indicator function $j_v^*(x_k; z_0)$ is defined not on the actual grade value $z_v(x_k)$ but on the estimated value $z_v^*(x_k)$ upon which selection is performed.

The problem is to estimate that yield, i.e., the nonlinear function φ of the

K, yet unknown, recovered block grades. The information is, say, the set S_n of n core data $z(x_\alpha)$, $x_\alpha = 1, \ldots, n$. Short of being perfect, what makes a block estimate $z_v^*(x_k)$ "good" for this problem? In other words, what are the relevant goodness criteria?

Efficiency Criterion

Estimates $z_v^*(x_k)$ should be such that to yield a recovery function φ the closest possible to the ideal recovery φ_0 defined as

$$\varphi_0[z_v(x_k),\ k = 1, \ldots, k] = \sum_{k=1}^{K} j_v(x_k;\ z_0) \cdot z_v(x_k) \tag{8}$$

where indicators, i.e., the selection, are defined now on actual block grades not their estimates

$$j_v(x_k;\ z_0) = \begin{cases} 1 & \text{for } z_v(x_k) > z_0 \\ 0 & \text{otherwise} \end{cases}$$

An efficiency loss function can thus be defined as the difference $|\varphi - \varphi_0|$, and should be minimized.

Accuracy Criterion

The actual recovery function is φ, not φ_0, for selection is based always on estimates $z_v^*(x_k)$ not on actual values $z_v(x)$. The next question is to predict that actual recovery. For such prediction, again, only estimated block grades are available, and the estimated recovery function may be written

$$\varphi^*(z_v^*(x_k),\ k = 1, \ldots, K) = \sum_{k=1}^{K} j_v^*(x_k;\ z_0) \cdot z_v^*(x_k) \tag{9}$$

The estimates $z_v^*(x_k)$ should be such that they yield a recovery estimate the closest possible to the actual recovery φ. An accuracy loss function to be minimized should also be defined; e.g., the difference $|\varphi - \varphi^*|$.

For the purpose of reserve estimation, a block estimate should be both efficient and accurate. Efficiency alone is not enough if an excellent actual recovery is badly overestimated and leads to overexpenditures. Accuracy alone is not enough, if one departs too much from the ideal recovery.

The preceding is but a limited presentation of the thorough analysis that geostatistics has given to the problem of ore reserves assessment, and more generally to the problem of locating and estimating extremes of a spatial distribution. It should suffice, though, to prove that much more is involved in ore reserves assessment (and geostatistics) than interpolation.

In fact, the almost sure way to meet the well-known curse of "vanishing high-grade tons" is to map naively the area A or panel P, planimeter to obtain K estimates $z_v^*(x_k)$, which are then plugged into expression (9) of the recovery estimate φ^*. The linear kriging interpolator, although an imperfect approach to this problem, does however outperform other more traditional linear interpolation techniques (see Helwick, 1983).

Geostatistical approaches to this problem are to be found in the relevant literature; see in particular the proceedings of Geostatistics Tahoe (1984) and the case study performed by Verly and Sullivan (1985) on a major U.S. gold deposit.

Evaluation of Spatial Variability

For homogenization and blending problems, spatial variability of mining unit grades within any particular panel must be determined. Consider such a panel containing K blocks of equal size $\{v\} : P_i = \Sigma_{k=1}^{K} v(x_k) \subset A$.

We wish to provide a characteristic of variability of the K block grades $z_v(x_k)$ within the panel P_i, i.e., some sort of roughness index function of the block geometry and size and the panel geometry and size, and independent of the density of data available over the panel P_i. Indeed, the fact that a panel P_i is more densely drilled than another P_j should not make its in situ variability greater.

If the K block grades $z_v(x_k)$ were known (see definition 6), their experimental variance within panel P_i could be calculated. One can debate, using a variance as an index of roughness; it is but a convenient measure accessible from the variogram tools. Again, the experimental variance appears as a nonlinear function of the K block grade:

$$s^2(v/P_i) = \frac{1}{K} \sum_{k=1}^{K} z_v^2(x_k) - \left[\frac{1}{K} \sum_{k=1}^{K} z_v(x_k)\right]^2 \tag{10}$$

If one were to replace the unknown block grades $z_v(x_k)$ in expression (10) by some interpolated estimates $z_v^*(x_k)$, the resulting variance $s^2*(v/P_i)$ would be biased, underestimated (usually) if a linear kriging interpolator is used, overestimated if some triangulation interpolator is used.

The intermediary step consisting in interpolating the block grades from neighboring data (usually defined on much smaller volumes) is neither necessary nor adequate, unless much data is available.

In geostatistics, spatial variance is estimated by a composite variance, a function of size and geometry of both v and P_i, and conditional to the data values

$$s^2*(v/P_i) = \frac{1}{K} \sum_{k=1}^{K} E\{[Z_v(x_k) - Z_{P_i}]^2 | Z_A(x_\alpha) = z(x_\alpha), \alpha = 1, \cdots, n\} \tag{11}$$

where $Z_v(x_k)$ is the RV modeling the particular block grade $z_v(x_k)$; $Z_{P_i} = 1/K \sum_{k=1}^{K} Z_v(x_k)$ is the panel grade RV model, and $Z_A(x_\alpha) = z(x_\alpha)$, $\alpha = 1, \cdots, n$, are the n data RVs conditioned to the n actual data values.

Expression (11) changes from one panel P_i to another, because the data environment and geometry (x_k, x_α) change. Also it does not increase necessarily if density of data (number n) increases. The statistics and models used to calculate expression (11) are global, in the sense that they are derived from an area larger than the particular panel P_i under consideration. However, the spatial variance estimate (11) is local in the sense that it is conditional to the particular data environment available within and around P_i.

Calculation of expression (11) does not require any interpolation, but requires much more than a mere z variogram $2\hat{\gamma}_A(h)$ such as the one fitted to graph (4). In all rigor, one needs the multivariate distribution of the RF model $Z_A(x)$. Approximations for such multivariate distributions are discussed later in the section on nonlinear geostatistics.

An Incidental Remark

The "dispersion variance" of geostatisticians is the nonconditional or deconditioned version of expression (11)

$$D^2(v/P) = \frac{1}{K} \sum_{k=1}^{K} E\{[Z_v(x_k) - Z_{P_i}]^2\} \tag{11a}$$

whatever the panel $P_i \subset A$.

This dispersion variance, together with the kriging variance (see later discussion), are possibly the most misused geostatistical tools. Being nonconditional to the data values, they cannot be considered as local measures of spatial variability, in the case of $D^2(v/P)$, or accuracy, in the case of kriging variances.

The local measure of spatial variability $s^2(v/P_i)$ may vary so much from one panel P_i to another P_j, as to render meaningless the estimation of any one particular $s^2(v/P_i)$ by the average value $D^2(v/P)$ (see Sullivan, 1984).

In addition, direct calculation of the dispersion variance $D^2(v/P)$ from the semivariogram model $\hat{\gamma}_A(h)$ depends essentially on non-data-based features such as the behavior of that model for short distances $|h| \leq \{v\}$.

A careful analysis of the limitations of the dispersion variance and other non-data-values-conditioned geostatistical tools is, however, no justification for the introduction of "roughness" indexes and other substitutes not yet fully analyzed.

Constrained Interpolation and Soft Information

The information available is seen usually as a set of n data values

$$z(x_\alpha), \qquad \alpha = 1, \cdots, n \tag{12a}$$

taken at n different locations $x_\alpha \in A$.

In practice, many other types of information are available. They are:

i. hard inequality-type data

$$z(x_\alpha) \in [a_\alpha, b_\alpha] \tag{12b}$$

where the datum value is known to lie within an interval of known bounds

ii. data defined on different supports and/or filters such as

$$z_v(x) = \frac{1}{\{v\}} \int_{v(x)} z(u)\, du, \qquad \text{or} \qquad \int_{v(x)} z(u) \cdot w(u)\, du \tag{12c}$$

where $w(u)$ is a known weighting function defined over the volume $v(x) \subset A$.

iii. soft information, of local nature such as geological interpretation about fault locations, or of global nature such as definition of zones of "homogeneous" mineralizations, directions of deposition, trends, etc.

In many earth sciences applications, hard data of type (12a) are scarce. Practitioners have no choice but to use other types of data, including soft information; unfortunately most often they do so in an ad hoc fashion, for lack of a clear methodology. Development of methodologies allowing a controlled usage of soft information should be a most urgent priority in earth sciences. The consequence of a "clean hand" attitude of considering only hard data is to accept, later, arbitrary decisions taken by personnel possibly less informed or careful.

At the least, hard data of type (12c) defined on supports other than cores should be used. Let us take this opportunity to recall a useful property of linear kriging, often overlooked by practitioners. As long as a "point," or core-based variogram model is available, linear kriging systems can accommodate *simultaneously* all linearly averaged data of type (12c). In other words, the kriged estimate can be a linear combination of data defined on different supports. The compositing process on cores of equal length is necessary for variogram inference, not for the kriging process. Everything else being equal, the kriging process would give a greater weight to a datum of larger support. Influence of data supports (i.e., volumes of definition) is a factor totally ignored by most traditional interpolation algorithms, including the most fancy triangulation techniques.

Inequality data of type (12b) are usually hard information and, as such, should be used. When a well stops short of meeting an horizon, hard information does exist at that location: the horizon depth $z(x_\alpha)$ is greater than the well depth a_α. Anywhere else, if that horizon is known to never outcrop, other hard information exists; the horizon depth is positive.

Recent developments in geostatistics, including but not limited to the indicator formalism, provide a common coding of hard, inequality, and soft data

which allows their simultaneous utilization; see Kulkarni (1984) and Kostov and Journel (1985). If the z attribute interval of variation is discretized by K thresholds: z_k, $k = , \cdots, K$, a hard datum value $z(x_\alpha)$ is seen as generating a completely informed vector of K indicator data defined as

$$i(x_\alpha; z_k) = \begin{cases} 0 & \text{if } z_k < z(x_\alpha) \\ 1 & \text{if } z_k \geq z(x_\alpha) \end{cases}$$

whereas an inequality datum of type (12b) would generate a vector where only those values $i(x_\alpha; z_k)$ such that $z_k \notin [a_\alpha, b_\alpha]$ are informed. Interpolation of the hypersurface $i(x; z_k)$ is then proceeded upon, using all indicator data available, both in the x and the z_k dimensions.

Knowledge of an estimate of the indicator function (of z_k) at a point x allows derivation of not only one, but several, estimates of the attribute $z(x)$, all different through their goodness criterion; see, hereafter, the section on nonlinear geostatistics.

Another Incidental Remark

All kriging algorithms, including all linear versions, may yield negative kriging (interpolation) weights, a fact well known to all beginners. Negative weights are no evil; on the contrary, they allow estimates to take on values outside the range of data.

Negative kriged weights are expected whenever a continuous spatial model is considered, such as a Gaussian variogram with no nugget effect. Such a continuous variogram would be appropriate to model a gentle undulating surface, which would result in a kriged map generating domes and troughs between data locations.

However, negative weights applied to outlier, rich grade data may result in awkward negative estimated grades. The problem lies not in the kriging algorithm(s) nor does the solution lie in constraining all kriging weights to be nonnegative. The solution is to inform the system that at any point x within the field an inequality datum exists: $z(x) > 0$, or, more generally for a grade: $z(x) \in [0, 100 \text{ percent}]$, and to use that information. The indicator formalism, previously mentioned, offers a full solution to the old problem of negative kriged grades.

KRIGING IS MUCH MORE THAN INTERPOLATION

Kriging, sensu lato, is a generic name for a variety of generalized least-squares (LS) estimation algorithms. However, to limit the scope of that set of tools to ordinary linear kriging applied to the interpolation of a surface $z(x)$, $x \in A$, would be a severe oversight.

Besides interpolation, and under various names, the least-squares principle is responsible for a great many applications in engineering and geophysics. In nonlinear geostatistics, kriging is used as a machine tool to build other tools, essentially probability distributions, from which estimates of $z(x)$ and functionals of $z(x)$ are derived.

Ordinary Kriging

Consider the predefined stationarity area A, informed by the data set S_n: $\{z(x_\alpha),\ \alpha = 1, \cdots, n\}$, and modeled by a RF $Z_A(x)$ with semivariogram $\hat{\gamma}_A(h)$ (see the section on the probability construct).

Any particular unknown attribute $z(x)$, $x \in A$, is estimated by a linear combination of the $n(x)$ neighboring data

$$z^*(x) = \sum_{\alpha=1}^{n(x)} \lambda_\alpha(x) \cdot z(x_\alpha) \tag{13}$$

with $z(x_\alpha) \in S_n$, and usually $n(x) < n$.

The weights $\lambda_\alpha(x)$ are determined to minimize a quadratic norm (a LS criterion) under the constraint

$$\sum_{\alpha=1}^{n(x)} \lambda_\alpha(x) = 1 \tag{14a}$$

The criterion to be minimized is a data configuration index, also called "estimation variance," and defined from the two RVs $Z_A(x)$ and $Z_A^*(x_\alpha) = \Sigma_\alpha \lambda_\alpha(x)\, Z_A(x_\alpha)$ representative of $z(x)$ and $z^*(x)$, respectively

$$E\{[Z_A(x) - Z_A^*(x)]^2\} = 2 \sum_{\alpha=1}^{n(x)} \lambda_\alpha(x) \cdot \hat{\gamma}_A(x - x_\alpha) - \sum_{\alpha=1}^{n(x)} \sum_{\beta=1}^{n(x)} \lambda_\alpha(x)\, \lambda_\beta(x)\, \hat{\gamma}_A(x_\alpha - x_\beta) \tag{14b}$$

Minimization of the quadratic form (14b) under constraint (14a) leads to a normal system of linear equations, also called the ordinary kriging (OK) system. This system provides the OK weights and the corresponding minimized estimation variance called "kriging variance."

From expression (14b) of the minimization criterion, several remarks can be made:

(i) Because the model $\hat{\gamma}_A(h)$ is data-based only for small distances h, typically smaller than one-half the dimensions of A in the direction of h (Fig. 1), the $n(x)$ data retained should be such that all semivariogram values $\hat{\gamma}_A(x - x_\alpha)$, $\hat{\gamma}_A(x_\alpha - x_\beta)$, required by expression (14b) are available and reliable. The range of the model $\hat{\gamma}_A(h)$, which incidentally may not exist, should play no role in the selection of the $n(x)$ data.

(ii) The constraint (14a) entails "unbiasedness" of the estimator $Z_A^*(x)$ and criterion (14b) entails that the error $[Z_A(x) - Z_A^*(x)]$ has a minimum variance. According to our probability construct, this means no more (and no less) that the weights $\lambda_\alpha(x)$s are determined such that, if all values $z(x)$, $x \in A$, were estimated using the same data configuration and that set of weights, the resulting distribution of errors $[z(x) - z^*(x)]$ would have zero mean and variance less than with any other linear weighting scheme. Hence, the name "best linear unbiased estimate" (BLUE), which is admittedly misleading because the term "best" is not qualified. "Minimum variance estimator" or "LS estimator" would be more appropriate.

(iii) Neither the estimation variance (14b), nor its minimized version the kriging variance are data-values dependent. They depend solely on the data configuration and the variogram model and are thus area A-specific indexes of data configuration, *not* measures of local accuracy. Only an arbitrary Gaussian construct could allow deriving confidence intervals from these variances. Data values-dependent statistics exist that can be used for measures of local accuracy, but they call for much more than a z-variogram model (see hereafter).

(iv) The kriging weights account for the spatial continuity specific to area A through the variogram model $2\hat{\gamma}_A(h)$, for the structural (note merely Euclidean) distances between location x and data locations through the terms $\hat{\gamma}_A(x - x_\alpha)$, *and* for the redundancy between the data as characterized by the terms $\hat{\gamma}_A(x_\alpha - x_\beta)$ in expression (14b).

Any substitute to OK should have these features plus a measure of *local* accuracy. Traditional interpolation algorithms, including inverse distance weighting and various triangulations, do not have all these features, nor do they provide any local measure of accuracy, which should not be mixed with local roughness of the surface $z(x)$. Last, in which sense these interpolators are "good" is yet unclear, which does not mean that they are "bad."

Nonlinear Geostatistics

The key to providing a local estimate qualified with a measure of local accuracy is to use estimations criteria and tools that are not only data-configuration dependent such as (14b) but also data-value $[z(x_\alpha)$s$]$ dependent.

In nonlinear geostatistics, the basic tool is not the variogram any more but a conditional probability distribution function defined as

$$F_x[z\,|\,(n)] = \text{Prob}\{Z_A(x) \leq z\,|\,Z_A(x_\alpha) = z(x_\alpha),\ \alpha = 1, \cdots, n\} \quad (15a)$$

or more generally

$$F_\varphi[b\,|\,(n)] = \text{Prob}\{\varphi[Z_A(x),\ x \in A] \leq b\,|\,Z_A(x_\alpha) = z(x_\alpha),\ \alpha = 1, \cdots, n\} \quad (15b)$$

where z and b are parameters of the distribution functions, $Z_A(x)$ is the RV modeling uncertainty about the unknown $z(x)$, for any $x \in A$, $\varphi[z(x), x \in A]$ is any known functional, not necessarily linear, of any number of unknown attributes $z(x)$, $x \in A$. For example, φ may be the functional (8) related to the problem of ore reserves assessment, or the functional (10) related to the problem of spatial variability assessment: $Z_A(x_\alpha)$s are the RVs conditioned to the data values $z(x_\alpha)$.

Before showing how such conditional probabilities can be inferred, it is pertinent to see for what they can be used.

Knowledge of the conditional probability function (15a) would provide a full solution to the problem of estimation of the unknown value $z(x)$

(i) confidence intervals are straightforwardly derived

$$\text{Prob}\{Z_A(x) \in [a, b] \,|\, (n)\} = F_x[b \,|\, (n)] - F_x[a \,|\, (n)] \tag{16}$$

(ii) various "optimal" estimates, related to different goodness criteria or loss functions, can be obtained. The estimate $[z(x)]^*_L$ that minimizes any predefined loss function, or error impact function $L(z - z^*)$, is the value z^* that minimizes the expected (average) loss value

$$[z(x)]^*_L = z^* \quad \text{such that} \tag{17}$$

$$E\{L[Z_A(x) - z^*] \,|\, (n)\} = \int_{-\infty}^{+\infty} L(z - z^*) \cdot f_x[z \,|\, (n)]\, dz \text{ is minimum}$$

where $f_x[z \,|\, (n)] = F'_x[z \,|\, (n)]$ is the density function associated with the cumulative probability function $F_x[z \,|\, (n)]$.

For example, if the loss function is the square function $L(z - z^*) = (z - z^*)^2$, the optimal estimate is the LS estimate equal to the conditional expectation of $Z_A(x)$, i.e., the expected value of the distribution $F_x[z \,|\, (n)]$.

If the loss function is linear and symmetric, $L(z - z^*) = |z - z^*|$, the optimal estimate is the conditional median, or median associated with the distribution $F_x[z \,|\, (n)]$.

More generally, if the goal of the study is not to interpolate a value $z(x)$ but to estimate a known functional $\varphi[z(x), x \in A]$ defined over A, the conditional probability distribution (15b) provides various estimates (besides the LS one) with confidence intervals.

Remarks

As opposed to the ordinary kriging variance, the widths of the confidence intervals (16) are both data configuration *and* data value dependent.

These confidence intervals do not depend on the particular loss function and estimate retained. In particular, they need not be centered on the estimate

retained. For example, if underestimation of a plutonium pollution is unacceptable (infinite loss), the "optimal" estimate for the unknown concentration is $+\infty$, a value outside any confidence interval calling for immediate cleaning.

Inferring Conditional Probability Distributions

Conditional probability functions (15a, b) are either inferred from a model for the multivariate distribution of the RF $Z_A(x)$, or estimated directly from the data. The multi-Gaussian formalism relates to the first approach whereas the indicator formalism relates to the second (see Verly, 1983, and Journel, 1983, respectively; Verly and Sullivan, 1985, and Geostatistics Tahoe, 1984, for applications).

The Multi-Gaussian Formalism

If the RF model $Z_A(x)$ was Gaussian, it would be fully characterized by the sole semivariogram model $\hat{\gamma}_A(h)$. Unfortunately, nature is never as congenial as to provide us with standard normal histogram $F_A^*(z)$ (see definition 1). The idea, common in statistics, is to "normal scores transform" the original data $z(x_\alpha)$, $\alpha = 1, \cdots, n$ into standard normal data $y(x_\alpha)$, $\alpha = 1, \cdots, n$, with zero mean and unit variance. Let T be that transform; $y(x) = T[z(x)]$, $x \in A$.

From the transformed data $\{ y(x_\alpha),\ \alpha = 1, \cdots, n,\ x_\alpha \in A \}$, a Gaussian RF model $Y_A(x)$ is built, fully characterized by the y-variogram model $2\hat{\Gamma}_A(h)$.

The data-based features of the RF model $Y_A(x)$ are its marginal distribution, standard normal by construction of the data $y(x_\alpha)$, and its variogram $2\hat{\Gamma}_A(h)$. Any other features, such as normality of all conditional distributions, come from the model, not necessarily from the data. But, as mentioned previously, the convenience of a model is to allow utilization of all its aspects, including those that were not identified with prior data information. In this regard, the multi-Gaussian model $Y_A(x)$ is remarkably powerful, but also potentially hazardous if one mistakes model for reality.

The original attribute $z(x)$ is but a particular known functional of $y(x)$, more precisely $z(x) = T^{-1}[y(x)]$. Thus the conditional distribution (15a) can be derived from the known multi-Gaussian distribution of $Y_A(x)$. Similarly, the functional $\varphi[Z_A(x),\ x \in A]$ can be expressed as a known functional $\varphi\{T^{-1}[Y_A(x)],\ x \in A\}$ of the Gaussian RF $Y_A(x)$, and its conditional distribution (15b) calculated. Note that, unless the transform T is linear, the conditional distribution (15a) is not normal, resulting in non-Gaussian confidence intervals (16).

Remarks

The multivariate distribution of the RF model $Z_A(x)$ is said to be multi-T-normal, in the sense that its T transform $Y_A(x)$ is multi-Gaussian.

Derivation of the conditional probability functions (15a,b) does require some specific krigings on the transformed variable $y(x)$, but also much more (see the references to Verly). These conditional distributions can also be sampled from simulated realizations of the RF model $Z_A(x) = T^{-1}[Y_A(x)]$ conditioned by the data $Z_A(x_\alpha) = z(x_\alpha)$, $\alpha = 1, \cdots, n$.

Log-normal kriging, i.e., linear kriging on the logarithmic transform $y(x) = \ln[z(x]$, is but a particular case and a particular aspect of the multi-Gaussian formalism (Parker, 1984).

The normal scores transform T requires knowledge of the initial z-histogram $F_A^*(z)$, but not its modeling.

The Indicator Formalism

Instead of being inferred from a prior Gaussian model, the conditional probability distribution (15a), not (15b), can be estimated directly from the z data.

Indeed, the probability (15a) for a given value of the parameter z appears as the conditional expectation of an indicator transform (i.e., nonlinear) of $Z_A(x)$, more precisely

$$\begin{aligned} E\{I(x; z)\,|\,(n)\} &= 1 \cdot F_x[z\,|\,(n)] + 0 \cdot \{1 - F_x[z\,|\,(n]\} \\ &= F_x(z\,|\,(n)) \end{aligned} \tag{18}$$

with $$I(x; z) = \begin{cases} 1, & \text{if } Z_A(x) \leq z \\ 0 & \text{otherwise} \end{cases}$$

Now, the previous indicator transform can also be applied to the z data, defining indicator data functions of the parameter z

$$i(x_\alpha; z) = \begin{cases} 1 & \text{if } z(x_\alpha) \leq z \\ 0 & \text{otherwise} \end{cases}$$

$$\text{for } \alpha = 1, \cdots, n$$

These indicator data can be used to estimate the indicator expected values $E\{I(x; z)|(n)\}$, i.e., the conditional probability values $F_x[z\,|\,(n)]$. Because the problem is estimating an expected value, LS estimates are relevant; the expected values (18) are estimated by kriging the indicator data

$$\begin{aligned} \{F_x[z_k\,|\,(n)]\}^* &= [i(x; z_k)]^* \\ &= \sum_{k'=1}^{K} \sum_{\alpha=1}^{n(x)} \lambda_{\alpha, k'}(x) \cdot i(x_\alpha; z_{k'}) \end{aligned} \tag{19}$$

where the z_k, $k = 1, \cdots, K$ are K threshold values discretizing the z variability

interval and $n(x)$ are the number of data locations retained for estimation of location x.

Such indicator (co)kriging requires inference of all indicator variograms and cross-variograms over the area A, i.e., of a discretized version of the bivariate probability distribution $F_A(h; z_k, z_{k'})$ as defined in (2) and (3). The data-based features of the indicator RF models $I(x; z_k)$, $k = 1, \cdots, K$ are thus much richer than those of the multi-Gaussian model $Y_A(x)$—a bivariate distribution instead of a mere bivariate moment $\Gamma_A(h)$. However, a distribution such as (15b) can be provided only by more complete models such as the multi-Gaussian RF $Y_A(x)$.

Thus in nonlinear geostatistics, kriging(s) is but a machine tool allowing construction of probabilistic tools—functions (15a, b)—used to approach the problems at hand; these problems being different, in nature, from interpolation.

CONCLUSIONS

Geostatistics is a branch of Applied Statistics specializing in the analysis and modeling of spatial variability in earth sciences. Its contribution to earth sciences has been essentially the formulation of complex problems in terms allowing usage of probabilistic tools.

The random function model should not be seen as some kind of generator of mineralizations, pollution, or metal prices. A model captures those aspects of the available information deemed important and reliable and allows deductive exercises (e.g., prediction) through appropriate tools. A model is never unique and can be judged only on its adequacy to solve the particular problem at hand. A model is not cast forever and should be updated as more information becomes available or retooled if the goal of the study changes.

Stationarity is a constitutive property of the random function model, not an intrinsic property of the phenomenon under study. It is no more than an a priori decision to average statistics over a given population or area. Linkage between the well-defined stationarity property (of a model) and the subjective notion of homogeneity (of a population) is tenuous and possibly misleading.

The key to defining local statistics and estimators within a larger stationarity area is to condition these statistics and estimators to the local data configuration *and* values. Such local geostatistics do exist and provide estimates informed with a measure of local accuracy.

Although not carrying any measure of local accuracy (except under a Gaussian construct), linear kriging features:

(i) utilization of a measure of spatial continuity, the variogram, specific to the stationarity area or population under study

(ii) consideration of data redundancy

(iii) a clear statement of the goodness criterion, and

(iv) a reasonable record of successful applications under its various names, including "generalized least-squares regression"

Geostatistics as a methodology addresses more challenging problems than interpolation. Kriging tools, although least-squares criterion-based, can be used to build models from which nonlinear, non-least-squares estimates can be derived.

As with any other methodology and/or set of tools, geostatistics should be questioned. Alternative models and tools should be developed and used if proven of superior performance. Such proofs should come from actual practice, not from aggressive polemics.

REFERENCES

Geostatistics-Tahoe (1984), refers to the two-volume proceedings of the 1983 NATO ASI "Geostatistics for Natural Resources Characterization," Verly et al., Eds. published by Reidel, Dordrecht.

Helwick, S. J., 1983, Geostatistics vs. conventional methods: Economical impact of misclassification of ore and waste," Proceedings of the 112th AIME Annual Meeting, Atlanta: AIME, New York.

Journel, A. G., 1983, Nonparametric estimation of spatial distributions: Math Geol. v. 15, no. 3, p. 445–468.

Kostov, C. and Journel, A. G., 1985, Coding and extrapolating expert information for reservoir description, Proceedings of the Reservoir Characterization Technical Conference, Dallas, April 1985: National Institute for Petroleum and Energy Research, Bartlesville, Oklahoma.

Kulkarni, R., 1984, Bayesian kriging in geotechnical problems, *in* Geostat.-Tahoe, part 2, p. 261–270.

Omre, 1984, Alternative variogram estimators in geostatistics. Ph.D. thesis, Stanford University.

Parker, H., 1984, Trends in geostatistics in the mining industry, *in* Geostat.-Tahoe, Part 2, p. 915–934.

Philip, G. M. and Watson, D. F., 1986, Matheronian geostatistics—Quo vadis?: Math. Geol., v. 18, no. 1, p. 93–117.

Shurtz, R. F., 1985, A critique of A. Journel's 'The deterministic side of geostatistics': Math. Geol. v. 17, no. 8, p. 861–868.

Sullivan, J., 1984, Non-parametric estimation of spatial distributions, Ph.D. thesis, Stanford University.

Verly, G., 1983, The multigaussian approach and its applications to the estimation of local reserves: Math Geol., v. 15, no. 2, p. 263–290.

Verly, G. and Sullivan, J., 1985, Multigaussian and probability krigings—Application to the Jerritt Canyon Deposit: Min. Eng., v. 37, no. 6, p. 568–574.