Short Course in Geology: Volume 8

Fundamentals of Geostatistics in Five Lessons

Andre G. Journel

American Geophysical Union

Short Course in Geology: Volume 8

Fundamentals of Geostatistics in Five Lessons

Andre G. Journel

Short Course Presented at the
28th International Geological Congress
Washington, D.C.

American Geophysical Union, Washington, D.C.

Maria Luisa Crawford and Elaine Padovani
Short Course Series Editors

Library of Congress Cataloging-in-Publication Data

Journel, A. G.
Fundamentals of geostatistics in five lessons.
Bibliography: p.
1. Geology–Statistical methods. I. Title.
QE33.2.S82J68 1989 551'.072 89-14911
ISBN 0-87590-708-3

Printed in the United States of America.

CONTENTS

PREFACE

From its inception as a separate discipline, geostatistics sought recognition from practitioners, not from mathematicians or physicists, and rightfully so. Indeed, the theory was essentially established by the 1950's by Kolmogorov and Wiener and exposed by Matern (1960), Whittle (1963), and Matheron (1965), among others. But there is a long, hard way between a concept expressed by matrix notations in a Hilbert space and its implementation and routine application. It is my opinion that the main contribution of geostatistics has been and still is *implementation*, an essential follow-up step much too often forsaken by theoreticians.

Implementation requires a prior effort of simplification. A concept or algorithm will take root only if understood by the user, who can customize it to the ever changing needs of his various projects. Practice over time is a merciless judge that will strip all concepts of their fancy dressing, whether wording or computer coding, and let the sole essence stand for itself. Geostatisticians would accelerate the diffusion process if they could go about extracting the essence of their favorite tools - a very humbling task - and deliver it in simple terms.

Stressing the essence of one's finding may not make it into a publication list but would help the understanding and correct application of the corresponding algorithm. Behind most sophisticated concepts, there is a simple idea sometimes so simple that we feel like dressing it up. Practitioners face real data with their extraordinary complexity that defies any pre-conceived model, and they are the best positioned to customize the algorithm to make it work. Thus, it is of great importance that they (the practitioners) understand what we (the academics) are proposing.

Andre G. Journel
Stanford University

Fundamentals of Geostatistics in Five Lessons

Andre G. Journel

Stanford Center for Reservoir Forecasting
Department of Applied Earth Sciences
Stanford University, Stanford, California 94035

Introduction

These lessons, except for the fourth, were "speed"-written as support for a "Geostatistics for Reservoir Characterization" course given in Dallas, December of 1987. There is definitely a need for new books in Geostatistics that would acknowledge the contribution of new application fields and sort the wheat from the tares, the theory that did yield from that which remained but elegant. I know of at least two such books in the mill. In the meantime, course supports were needed and I attempted this quick draw. I ask for the reader indulgence and patience until the availability of official books.

Textbooks are polished logical constructions which do not lend themselves to the spot-painting and diversions that could sometimes enlighten a class. Lessons allow digressions, returns to fundamentals, parallels, that could reveal a hidden side of the theory being developed and, in the best case, the essence of an algorithm that which makes it work. Building from rigorous random function theory, how could one tell that the essence of ordinary kriging is:

1. the usage of a structural distance, specific to the variable being considered, which need not be a variogram
2. the possibility of accounting for data redundancy, as figured by the data covariance matrix?

If kriging sometimes works it is not because of its probabilistic pedigree - in fact the algorithm could be established without a single reference to random variables - but because it extends well-proven and intuitive distance-weighting criteria. Associating kriging to distance-weighting algorithms, or in a dual fashion to surface-fitting algorithms, makes it more "ordinary" now that it is severed from that hazy random function source. Geostatistics may end up looking less prestigious (or mysterious?), but being better understood will be better applied.

After much hard selling, the time for a fresh and more temperate look at geostatistics has come. Geostatistics is foremost Data Analysis and Spatial Continuity Modeling. Such analysis and modeling cannot be done without a clear understanding of the origin of the data, including geological interpretation. The main reason for modeling spatial continuity is to assess spatial uncertainty. As for using probabilistic models, it is naive to think that any statistical tool provides objectivity, it should provide though consistency once a prior model has been chosen. Geostatistics is a set of numerical tools to be added to the large tool chest of the geologist; it allows transporting quantitatively a geological model all the way to process design and engineering. That geological model should not stem from a blackbox package, particularly if that package is cryptic. Good geology based on well understood data is still the only recipe for good reservoir/site/deposit characterization.

These five lessons address the fundamentals of geostatistical theory relevant to spatial interpolation, image reconstitution and uncertainty modeling. The practice of geostatistics, although of paramount importance, is not covered here for lack of space. The book from Srivastava and Isaaks, which should be in the shelves by mid-1989, will fulfill that need beautifully. The remarkably user-friendly and yet complete software "Geostat Toolbox", made public-domain by its author Roland Froidevaux, provides the tools for anyone to get started.

Lesson I proposes a brief review of statistics and no-

tations needed for developing the further lessons. The reader is supposed to have a prior familiarity with statistics, integral and differential calculus at an introductory college level.

Lesson II presents the classical linear regression theory with the particular geostatistical twist, in the sense that the data used (the so-called "independent" variables of classical regression) are themselves dependent one upon each other and thus needed to be made independent in a first step. The simple kriging system (SK) is shown to be but a variant of the normal system of equations. A dual interpretation of the simple kriging algorithm shows it amounts to fit covariance-type interpolation functions to the data values at their locations.

Simple kriging requires that the mean of the variable over the field being estimated be constant and known. Ordinary kriging does not require knowledge of that mean, as long as it remains constant. Kriging with a trend model allows considering a variable mean, function of the coordinates values. That function is everywhere unknown but is of known functional form and could represent a local trend component being added to residual values. In Lesson III, it is shown that ordinary kriging and kriging with a trend model are achieved by adding specific constraints to the normal system of equations.

Lesson IV is possibly the most important of this set of five lessons, for it addresses the critical problem of uncertainty modelling. No estimator is devoid of potential error, thus qualified decision-making requires an assessment (model) of the uncertainty underlying each estimate. Too often, a Gaussian error distribution is casually taken as model for uncertainty although evidence for the inadequacy of such symmetric distribution model do exist. Alternative models based on the actual distribution of neighboring data are proposed building on an indicator data kriging paradigm. Construction of models of uncertainty precedes the derivation of an estimate for the unknown, which allows retaining non-least squares, i.e. non-kriging-type estimates possibly better suited to the project at hand.

All moving average-type estimates, including all kriging estimates, provide a smooth image of the underlying reality: the variogram of these estimates would not reproduce the data variogram. The concept of conditional simulation allows generating alternative, equiprobable, images which honor data values at their locations and reflect a series of spatial continuity functions. The novel technique of Indicator conditional simulations, presented in Lesson V allows generation of images that do not suffer from the maximum entropy (maximum disorganization for a given variogram model) limitation of Gaussian-related random function models. Also, indicator simulations allows honoring, in addition to hard data, soft information whether local (e.g. constraint intervals at specific locations) or global (e.g. interpretative structural geology). A set of simulated images provide an assessment of joint spatial uncertainty rather than a series of, say, local estimation variances.

References are proposed at the end of each lesson, and an index of subjects is given.

Lesson I: Statistics Review and Notations

In this lesson, we will review only those basic notions of statistics that are useful to geostatistics and spatial interpolation problems. Engineering-type definitions are preferred over more rigorous but less intuitive axiomatic definitions.

Random variable (RV)

A random variable can be seen as a variable, say, Z, in capital letters, that can take a series of outcomes or realizations (z_i , $i = 1, \ldots, N$) with a given set of probability of occurrence (p_i , $i = 1, \ldots, N$)

When the number N of occurrences is finite, one speaks of a "discrete" RV. The N probabilities of occurrences must verify the conditions

$$p_i \geq 0 \quad , \text{ for all } i = 1, \cdots, N \tag{1}$$

$$\sum_{i=1}^{N} p_i = 1 \tag{2}$$

If the number of possible occurrences is infinite, say, a porosity value within the interval $[0, 100\%]$, the RV Z is said to be continuous, and its probability distribution is characterized by the cumulative distribution function (cdf) defined as:

$$F(z) = \text{Prob}\{Z \leq z\} \in [0, 1] \tag{3}$$

The cdf corresponds to the notion of cumulative histogram in the discrete case:

$$F(z_i) = \sum_{(j)} p_j \quad , \quad \begin{array}{l}\text{with the set } (j) \text{ corresponding to} \\ \text{all realizations } z_j \leq z_i \, .\end{array}$$

The cdf fully characterizes the RV Z. Probability intervals can be derived:

$$\text{Prob}\{Z \in]\, a, b\,]\,\} = F(b) - F(a)$$

Similarly, the probability of exceedence of any threshold value can be derived:

$$\text{Prob}\{Z > a\} = 1 - F(a)$$

Quantiles

The p-quantile of the distribution $F(z)$ is the value z_p such that: $F(z_p) = p \in [0,1]$, i.e., the value z_p which has a probability p not to be exceeded by the RV Z. Defining the inverse $F^{-1}(p)$ of the cdf:

$$p - \text{quantile } z_p = F^{-1}(p)\ , \text{ with } p \in [0,1] \qquad (4)$$

Expressing p in percent, the p-percentile is defined:

$$p - \text{percentile } z_p = F^{-1}(100\,p)\ , \text{ with } p \in [0,100]$$

Quantiles of interest are:

- the .5 quantile (50th percentile) or median:

$$M = F^{-1}(.5) \qquad (5)$$

- the lower and upper quartile:

$$z_{.25} = F^{-1}(.25)\ ,\ z_{.75} = F^{-1}(.75)$$

The interquartile range (IR) is defined as the interval bounded by the upper and lower quartile: IR $= [z_{.25}\ ,\ z_{.75}]$

Expected value

The expected value is the probability-weighted sum of all possible occurrences of the RV.

The expected value of Z, also called mean of Z, is defined as:

- in the discrete case, corresponding to relations (1) and (2):

$$E\{Z\} = m = \sum_{i=1}^{N} p_i z_i \qquad (6)$$

- in the continuous case with cdf $F(z)$, under conditions of existence of the integrals:

$$E\{Z\} = m = \int_{-\infty}^{+\infty} z\, dF(z) = \int_{-\infty}^{+\infty} z\, f(z)\, dz \qquad (7)$$

$$= \lim_{K \to \infty} \sum_{k=1}^{K} z'_k \left[F(z_{k+1}) - F(z_k)\right]\ , \text{ with } z'_k \in \left]\, z_k, z_{k+1}\,\right]$$

where: $f(z) = F'(z)$ is the probability density function (pdf) defined as the derivative, when it exists, of the cdf $F(z)$. The integral $\int_{-\infty}^{+\infty} z\, dF(z)$ is approximated by K classes of respective frequencies $[F(z_{k+1}) - F(z_k)]$, and z'_k is a value within the kth class, say, the center of that class:

$$z'_k = (z_{k+1} + z_k)/2$$

The expected value of any well-behaved function of Z, say, $\varphi(Z)$, can also be defined, under conditions of existence of the integral:

- in the discrete case:

$$E\{\varphi(Z)\} = \sum_{i=1}^{N} p_i\, \varphi(z_i) \qquad (8)$$

- in the continuous case:

$$E\{\varphi(Z)\} = \int_{-\infty}^{+\infty} \varphi(z)\, dF(z) = \int_{-\infty}^{+\infty} \varphi(z) \cdot f(z)\, dz$$

$m = E\{Z\}$ being the mean of Z, the variance of the RV Z is defined as the expected squared deviation of Z about its mean:

$$\text{Var}\{Z\} = \sigma^2 = E\left\{[Z - m]^2\right\} \geq 0 \qquad (9)$$

$$= \sum_{i=1}^{N} p_i (z_i - m)^2\ , \text{ in the discrete case}$$

$$= \int_{-\infty}^{+\infty} (z-m)^2 dF(z) = \int_{-\infty}^{+\infty} (z-m)^2 f(z) dz\ , \qquad (10)$$

in the continuous case.

The mean is a central location characteristic of the pdf $f(z)$, see Figure 1. The variance is a characteristic of spread of that distribution around the mean.

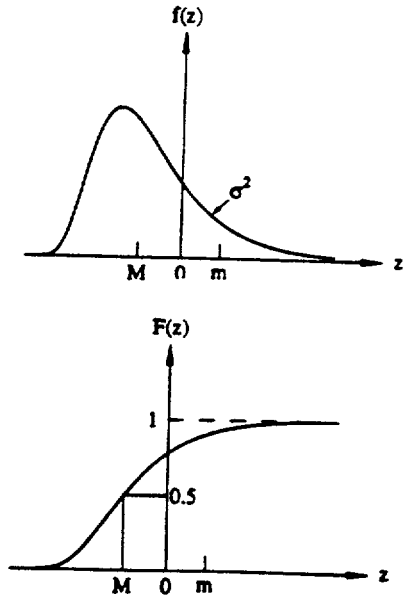


Fig. 1. Probability density function (pdf), and Cumulative density function (cdf)

Another central location characteristic is the median M. A corresponding characteristic of spread would be the mean absolute deviation about the median (mAD):

$$\text{mAD} = E\{|\, Z - M \,|\}$$

The skewness sign is defined as:

$$\text{Sign of}(m - M)$$

A positive skewness usually indicates a long tail of the pdf towards the high outcome values.

Order relations

The cdf $F(z)$ and its derivative, the pdf $f(z)$, being probability-related functions, must verify the following conditions:

$$F(z) = \int_{-\infty}^{z} f(u)du \in [0,1] \qquad (11)$$

$F(z) \geq F(z')$, for all $z \geq z'$, i.e., the cdf is a non-decreasing function of z.

Correspondingly:

$$f(z) = F'(z) \geq 0$$

$\int_{-\infty}^{+\infty} f(z)dz = F(+\infty) = 1$, i.e., the integral of the pdf is equal to 1.

Linear property of the expected value

The expected value is a "linear operator," in the sense that the expected value of a linear combination of RV's is the linear combination of the expected values of these RV's:

$$E\left\{\sum_k a_k Z_k\right\} = \sum_k a_k E\{Z_k\} \qquad (12)$$

whatever the RV's Z_k whether dependent or independent from each other, and whatever the constant a_k's.

In particular, for any transform functions φ_1 and φ_2:

$$E\{\varphi_1(Z) + \varphi_2(Z)\} = E\{\varphi_1(Z)\} + E\{\varphi_2(Z)\}$$

provided the corresponding expected values exist.

Application:

$$\begin{aligned} \text{Var}Z &= E\{[Z-m]^2\} = E\{Z^2 - 2mZ + m^2\} \\ &= E\{Z^2\} - 2mE\{Z\} + m^2 = E\{Z^2\} - m^2 \end{aligned}$$

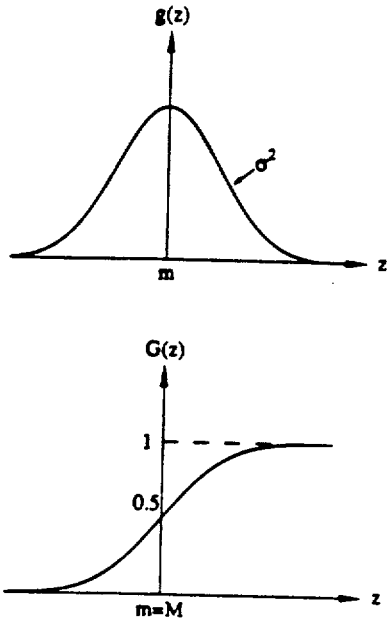


Fig. 2. Normal (Gaussian) pdf and cdf

Gaussian (normal) model

It is a distribution model fully characterized by its two parameters, mean m and variance σ^2. The pdf $g(z)$ is, see Figure 2:

$$g(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z-m}{\sigma}\right)^2\right] \qquad (13)$$

The standard normal pdf corresponds to $m = 0$, $\sigma^2 = 1$

$$g_0(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

The corresponding cdf's have no close-form analytical expression, but the standard normal cdf $G_0(z)$ is well tabulated:

$$G_0(z) = \int_{-\infty}^{z} g_0(u)du \qquad (14)$$

$$G(z) = \int_{-\infty}^{z} g(u)du = G_0\left(\frac{z-m}{\sigma}\right)$$

Symmetry. The Gaussian distribution is symmetric about its mean, thus:

$$\begin{aligned} m &= M \,,\; g(z+m) = g(m-z) \,, \\ G(m-z) &= 1 - G(m+z) \,, \text{ for all } z \end{aligned} \qquad (15)$$

$z_{1-p} = 2m - z_p$, for all $p \in [0, .5]$, z_p being the p-quantile.

Some Gaussian values

$$G(m+\sigma) = .84\ ,\ G(m-\sigma) = .16$$

$$G(m+2\sigma) = .977\ ,\ G(m-2\sigma) = .023$$

Thus:

$$\text{Prob}\{Z \in [m \pm \sigma]\} = .84 - .16 = .68$$

$$\text{Prob}\{Z \in [m \pm 2\sigma]\} = .977 - 0.23 = .954 \simeq .95$$

The "success" of the Gaussian model stems from a series of Central Limit theorems which state that:

The sum of a not too large number of independent, equally distributed (although not necessarily Gaussian), standardized RV's tend to be normally distributed, i.e., if the n RV's Z_i's have same cdf and zero means, the RV $Y = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$ tend towards a normal cdf, as $n \to \infty$.

The most restrictive constraint to application of Central Limit theorems is the condition of independence: the n RV's Z_i must be independent.

In the highly controlled environment of a laboratory experiment where measurement devices are carefully chosen and monitored, one may expect normal distributions for measurement errors. In fact, the experiment(s) may have been designed specifically to generate such normal distribution. However, in the nature-controlled environment of a spatial distribution, say, that of porosity within a layer or permeability within a sandstone formation, there is no reason to believe a priori that the conditions of a Central Limit theorem would apply. Nor should these conditions apply to the various sources of error involved in a spatial interpolation exercise, that of estimating, say, a porosity value from neighboring well data.

The lognormal model

The normal (Gaussian) model is very convenient because it is fully characterized by only two parameters, its mean and variance. However, it is symmetric and allows for the occurrence of negative outcomes. Many experimental distributions tend to be skewed (asymmetric) with mean different from median, and most earth sciences variables are non-negative valued.

Various transforms of the normal model have been defined to accommodate the usual features of earth sciences data distributions.

A positive RV, $Y > 0$, is said to be lognormally distributed if its logarithm $X = \ln Y$ is normally distributed:

$$Y > 0 \rightsquigarrow \log N(m, \sigma^2)\ \text{, if } X = \ln Y \rightsquigarrow N(\alpha, \beta^2) \quad (16)$$

The lognormal distribution is a two-parameter distribution fully characterized by:

- either its mean and variance (m, σ^2) also called arithmetic parameters,
- or the mean and variance (α, β^2) of the log transform $X = \ln Y$, also called logarithmic parameters.

The lognormal cdf is more easily expressed as a function of its logarithmic parameters (α, β^2).

$$\text{Prob}\{Y \leq y\} = F_Y(y) = G_0\left(\frac{\ln y - \alpha}{\beta}\right) \quad \text{for all } y > 0 \quad (17)$$

where $G_0(\cdot)$ is the standard normal cdf defined in (15).

The corresponding pdf is:

$$f_Y(y) = F'_Y(y) = \frac{1}{\beta y} g_0\left(\frac{\ln y - \alpha}{\beta}\right)$$

where $g_0(\cdot)$ is the standard normal pdf defined in (14).

The relations between arithmetic parameters and logarithmic parameters are:

$$\begin{cases} m = e^{\alpha+\beta^2/2} \\ \sigma^2 = m^2\left[e^{\beta^2} - 1\right] \end{cases} \quad \begin{cases} \alpha = \ln m - \beta^2/2 \\ \beta^2 = \ln\left(1 + \frac{\sigma^2}{m^2}\right) \end{cases} \quad (18)$$

The lognormal pdf $f_Y(y)$ plots as a positively skewed distribution with a long tail allowing for very large y-outcomes, although with increasingly low probability of occurrence, see Figure 3.

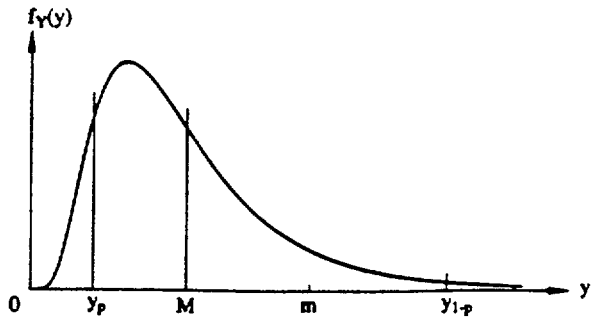


Fig. 3. Lognormal density function

If ν_p is the p-quantile of a standard normal distribution, i.e.,

$$\nu_p = G_0^{-1}(p)\ ,\ p \in [0,1]\ ,$$

the p-quantile of the lognormal distribution (α, β^2) is:

$$y_p = e^{\alpha+\beta\nu_p} \tag{19}$$

entailing the following relation which can be used to determine the logarithmic mean α from estimates of opposite quantiles y_p, y_{1-p}:

$$y_p \cdot y_{1-p} = e^{2\alpha} = M^2 \Longrightarrow 2\alpha = \ln y_p + \ln y_{1-p} \tag{20}$$

where: $M = y_{.5}$: median of the lognormal distribution.

As a corollary of the Central Limit theorem, the product of a great number of independent, identically distributed RV's tend to be lognormally distributed. Indeed,

$$Y = \Pi_{i=1}^{n} Y_i \Longrightarrow \ln Y = \sum_{i=1}^{n} \ln Y_i \rightarrow \text{Normal, as } n \rightarrow \infty$$

thus: $Y \rightarrow$ Lognormal, as $n \rightarrow \infty$.

However, there is no a priori reason to believe that the various sources of spatial variability of permeability/transmissivity are all independent, of roughly equal variance, and are multiplicative. The traditional modeling of permeability distributions by a lognormal model is more a question of convenience than a hard fact supported either by data or theory. There exist many other possible models for positively-skewed distributions of non-negative variables.

Bivariate distribution

Up to now we have considered only one random variable at a time, whether Z_i, or the sum $Y = \sum_{i=1}^{n} Z_i$ of n RV's.

In earth sciences, what is often most important is the pattern of dependence relating one variable X to another Y, relating the outcome of one RV X to the outcome of another Y. This pattern of dependence is particularly critical if one wishes to estimate, say, the outcome of X (core porosity) from the known outcome of Y (well log datum).

Just like the distribution of outcomes of a single RV X is characterized by a cdf $F_X(x) = \text{Prob}\{X \leq x\}$, see relation (3), the joint distribution of outcomes from a pair of RV's X and Y is characterized by a joint cdf:

$$F_{XY}(x,y) = \text{Prob}\{X \leq x\ ,\ \text{and}\ Y \leq y\}\ , \tag{21}$$

in practice estimated by the proportion of pairs of data values jointly below the respective threshold values x, y.

The bivariate (X, Y) equivalent of a histogram is a scattergram, where each data pair (x_i, y_i) is plotted as a point, see Figure 4.

The degree of dependence between the two variables X and Y can be characterized by the spread of the previous scattergram around any regression line, with perfect linear dependence corresponding to all experimental pairs $(x_i, y_i)\ ,\ i = 1, \cdots, N$ plotting on that line.

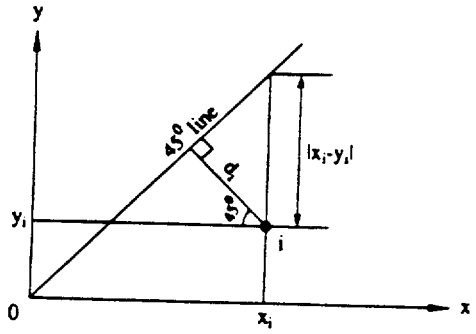


Fig. 4. Pair (x_i, y_i) on a scattergram

Thus, the moment of inertia of the scattergram around e.g. the 45° line would be a characteristic of lack of dependence, see Figure 4:

$$\gamma_{XY} = \frac{1}{N}\sum_{i=1}^{N} d_i^2 = \frac{1}{2N}\sum_{i=1}^{N} [x_i - y_i]^2 \tag{22}$$

This moment of inertia is called the "semi-variogram" of the set of data pairs (x_i, y_i) ,
$i = 1, \cdots, N$. The variogram, $2\gamma_{XY}$, is none other than the average squared difference between the two components of each pair. The greater the variogram value, the greater the scatter and the less related are the two variables X and Y.

Just like one has defined the expected value of, say, X, as the frequency-weighted average of the X-outcomes, estimated here by:

$$m_X = \frac{1}{N}\sum_{i=1}^{N} x_i$$

one could define the expected value of the product XY as the bivariate probability-weighted average of the joint outcomes $XY = xy$; more precisely and similarly to relation (7):

$$E\{XY\} = \iint_{-\infty}^{+\infty} xy\, d^2F_{XY}(x,y) = \iint_{-\infty}^{+\infty} xy\, f_{XY}(x,y)dxdy \quad (23)$$

in practice, estimated by:

$$\frac{1}{N}\sum_{i=1}^{N} x_i y_i$$

where: $f_{XY}(x,y) = \frac{d^2F_{XY}(x,y)}{dxdy}$ is the bivariate probability density function (pdf).

$d^2F_{XY}(x,y) = f_{XY}(x,y)dxdy$ is the probability of occurrence of the joint outcome $\{X = x \pm dx\ ,\ y = y \pm dy\}$. Similarly, the probability of outcome of the pair (x_i, y_i) is $\frac{1}{N}$ if N pairs were sampled.

The bivariate moment (23) is called the non-centered covariance of the two RV's X, Y. The centered covariance, or simply "covariance," is defined as:

$$\begin{aligned}\text{Cov}\{X,Y\} &= \sigma_{XY} = E\{[X - m_X]\cdot[Y - m_Y]\} \\ &= E\{XY\} - m_X \cdot m_Y\end{aligned} \quad (24)$$

in practice estimated by:

$$\frac{1}{N}\sum_{i=1}^{N}(x_i - m_X)(y_i - m_Y) = \frac{1}{N}\sum_{i} x_i y_i - m_X m_Y$$

with: $m_X = \frac{1}{N}\sum_{i=1}^{N} x_i\ ,\ m_Y = \frac{1}{N}\sum_{i=1}^{N} y_i$

The variance, say, of X, as defined in relation (9) is but the autocovariance of X with itself:

$$\sigma_X^2 = \text{Var}\{X\} = \text{Cov}\{X,X\} = E\{[X - m_X]^2\} \geq 0$$

Note that although the variances σ_X^2 and σ_Y^2 are necessarily non-negative, the covariance σ_{XY} may be negative if a positive deviation $[x - m_X]$ is related, in average, with a negative deviation $[y - m_Y]$.

The coefficient of correlation ρ_{XY} between the two RV's X and Y is but a covariance standardized to be unit-free:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Cov}\{X,Y\}}{\sqrt{\text{Var}\{X\}\cdot\text{Var}\{Y\}}} \in [-1,+1] \quad (25)$$

It can be shown (Schwarz inequality) that the coefficient of correlation is necessarily within $[-1,+1]$.

Relation between covariance and variogram

Consider the bivariate data set $(x_i, y_i), i = 1,\cdots,N$ and its moments:

$$2\gamma_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2\ ,\ \sigma_{XY} = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - m_X m_Y$$

$$2\gamma_{XY} = \left[\frac{1}{N}\sum_{i=1}^{N} x_i^2 - m_X^2\right] + \left[\frac{1}{N}\sum_{i=1}^{N} y_i^2 - m_Y^2\right] -$$

$$- \left[\frac{2}{N}\sum_{i=1}^{N} x_i y_i - 2m_X \cdot m_Y\right] + \left(m_X^2 + m_Y^2 - 2m_X m_Y\right)$$

i.e.,

$$2\gamma_{XY} = \sigma_X^2 + \sigma_Y^2 + (m_X - m_Y)^2 - 2\sigma_{XY} \geq 0 \quad (26)$$

Thus, whenever the variogram $2\gamma_{XY}$ increases, the covariance σ_{XY} decreases: the greater the spread of the scattergram $(x_i, y_i), i = 1,\cdots,N$ around the 45° line, see Figure 4, the larger γ_{XY} and the smaller the covariance σ_{XY} and the correlation coefficient ρ_{XY}.

γ_{XY} is a measure of variability, while σ_{XY} and ρ_{XY} are measures of similarity.

Both measures γ_{XY} and σ_{XY} are dependent on a linear transform of the variables X, Y, either a translation $X + b$, or a unit rescaling aX. A more intrinsic, unit-free, measure of variability/similarity should therefore be defined on the standardized variables:

$$X' = (X - m_X)/\sigma_X \text{ and } Y' = (Y - m_Y)/\sigma_Y$$

The corresponding statistics are:

$$m_{X'} = m_{Y'} = 0\ ,\ \sigma_{X'}^2 = \sigma_{Y'}^2 = 1 \quad (27)$$

$$\sigma_{X'Y'} = E\left\{\frac{X - m_X}{\sigma_X}\cdot\frac{Y - m_Y}{\sigma_Y}\right\} = \rho_{XY} \in [-1,+1]$$

$$\gamma_{X'Y'} = 1 - \rho_{XY} \in [0,2]$$

The coefficient of correlation ρ_{XY} is unchanged by any linear transform applied on either RV X or Y.

Note that, when $\rho_{XY} = 1 \implies \gamma_{X'Y'} = 0$, thus all standardized pairs $\left(\frac{x_i - m_X}{\sigma_X}\ ,\ \frac{y_i - m_Y}{\sigma_Y}\right)$, $i = 1,\cdots,n$ are aligned around the 45° line.
Therefore:

$$x_i = m_x + \frac{\sigma_X}{\sigma_Y}(y - m_Y) = ay + b \quad (28)$$

A unit correlation, and more generally $\rho_{XY}^2 = 1$, characterizes two RV's which are linearly related. The coefficient ρ_{XY} is a measure of linear (cor)relation between the two RV's X, Y. However, zero correlation, i.e., $\rho_{XY} = 0$, does not entail necessarily that the two variables are in-

dependent; it indicates simply that their pattern of dependence cannot be modeled by a linear relation of type (28).

Random function

In any of the previous discussions, the two RV's X, Y can represent:

(i) either two different attributes measured at the same location, say, porosity and permeability measured from the same core plug;

(ii) the same attribute measured at two different locations in space, say, porosity at locations x and $x+h$ distant of a vector h: $X = Z(x)$, $Y = Z(x+h)$; or

(iii) two different attributes measured at two different locations, say, porosity at location x and permeability at vector h apart: $X = \phi(x)$, $Y = K(x+h)$.

In all cases, the semi-variogram γ_{XY} or the correlation ρ_{XY} would measure the degree of variability/similarity between the two variables X, Y.

The second case (ii) is of particular interest for spatial interpolation problems, where a whole field of a given attribute, $Z(x), x \in$ Site A, has to be inferred (mapped) from a limited number of locations sampled for that same attribute.

Pooling all $n(h)$ pairs of data found on attribute Z over the same site/zone/layer A, these pairs being approximately distant of the same vector h (in length and direction), one could estimate the variogram characteristic of spatial variability over A:

$$\gamma_{Z(x),Z(x+h)} = \hat{\gamma}_A(h) = \frac{1}{n(h)} \sum_{\alpha=1}^{n(h)} [z(x_\alpha) - z(x_\alpha + h)]^2 \tag{29}$$

By varying the vector h (in length and direction), that characteristic is made into a vectorial function $\gamma(h)$.

The experimental semi-variogram $\hat{\gamma}_A(h)$ is not the estimate of some elusive expected value of squared increments of a no less elusive set of random variables $Z(x), Z(x+h)$. It is the discrete estimate of a well-defined spatial integral defining an average over A:

$$\gamma_A(h) = \frac{1}{A(h)} \int_{A \cap A_{-h}} [z(x) - z(x+h)]^2 \, dx \tag{30}$$

with: $A \cap A_{-h}$ being the intersection of A and its translate A_{-h} by vector $-h$. If $x \in A \cap A_{-h}$, then both x and $x+h$ are in A. $A(h)$ is the measure (area/volume) of that intersection.

Just as the random variable $Z(x)$ and its distribution characterize the uncertainty about the attribute value at location x, the random function $Z(x), x \in A$, defined as a set of dependent RV's, will characterize the joint spatial uncertainty over A. The variogram $2\gamma(h)$ of that random function (RF) characterizes the degree of spatial variability between any two RV's $Z(x), Z(x+h)$ distant of vector h. The variogram $2\gamma(h)$ is modeled after the spatial average $2\gamma_A(h)$ which can be estimated by the discrete sum $2\hat{\gamma}_A(h)$ defined in (29).

The modeling of $\gamma(h)$ after the spatial average $\gamma_A(h)$ amount to decide to average statistics over a given area/site/zone/layer A. That decision is necessarily subjective, usually based on geological criteria, and cannot be proven (or refuted) by statistical tests. Once the decision of modeling $\gamma(h)$ after $\gamma_A(h)$ is taken, $\gamma(h)$ just like $\gamma_A(h)$ is not any more location (x) dependent within A:

$$2\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} \text{ is independent of } x \in A. \tag{31}$$

The moment $\gamma(h)$ is said to be stationary. Stationarity is a model decision, not some intrinsic property of the actual distribution of z-values over A.

Once a model $\gamma(h)$ is obtained, the corresponding covariance model can be deduced by the classical relation:

$$C(h) = \text{Constant} - \gamma(h) \tag{32}$$

Indeed:

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

$$\left[E\left\{Z^2(x)\right\} - m^2\right] + \left[E\left\{Z^2(x+h)\right\} - m^2\right] -$$

$$2\left[E\left\{Z(x)Z(x+h)\right\} - m^2\right] =$$

$$\text{Var}\{Z(x)\} + \text{Var}\{Z(x+h)\} - 2\,\text{Cov}\{Z(x), Z(x+h)\} =$$

$$2[C(o) - C(h)], \text{ thus: } C(h) = C(o) - \gamma(h)$$

The constant of formula (32) can either be set equal to the sill $\gamma(\infty)$ of the semi-variogram *model* if it exists, or can be set to any arbitrary large value. That arbitrary constant should then be filtered out from all further algorithms, as will be the case in ordinary kriging, see Lesson III, System (4).

The very idea of geostatistics is remarkably simple. It consists in the following sequence of steps:

(i) Define an area/site/population/$\cdots A$, deemed homogeneous enough to warrant statistical averaging within it.

(ii) Scan all data available within A to calculate the experimental h-characteristics of spatial variability, i.e., calculate the experimental variogram values (30).

(iii) Smooth and complete the partial experimental image thus obtained into a model $\gamma(h)$, available for all interdistance vectors h.

(iv) Use this model $\gamma(h)$, or the corresponding covariance model $C(h)$ for traditional, well-proven regression techniques, also called "kriging", see Lesson II and III.

The most important steps, and by far the most consequential of any geostatistical study, are: step (i) - decision of averaging or stationarity, and, step (iii) - the modeling. Step (iv) is but well-known calculus.

The decision of stationarity is implicit in all statistics, it is not particular to the geostatistical approach. That decision allows defining a pool of data (area A) over which experimental averages and proportions will be calculated and assumed representative of the population as a whole, not of any particular individual location ($x \in A$).

In the practice of reservoir characterization, that decision corresponds to the traditional split of the reservoirs into zones/layers deemed homogeneous enough by the geologist. Of course, that split should not be so fine as to retain one zone per datum, for then no averaging (no statistics) is possible.

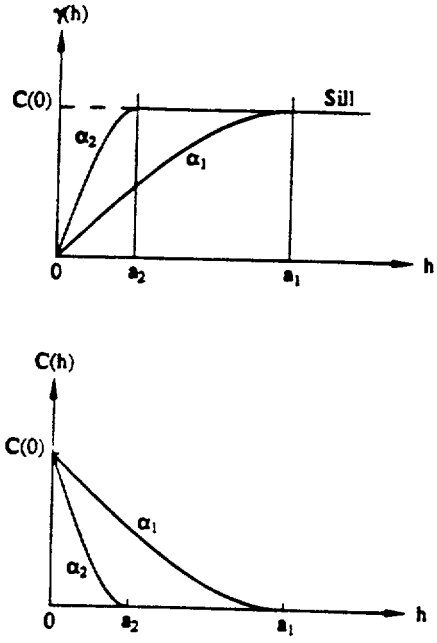


Fig. 5. Anisotropic variogram and covariance
α_1: direction of continuity
α_2: direction across continuity
a_1: distance at which spatial correlation vanishes in direction α_1:
$C(|h|, \alpha_1) = 0$, $\gamma(|H|, \alpha_1) =$ sill value $= C(0)$,
for all $|h| \geq a_1$

Typical variograms

Variogram and covariance as defined in relations (23) to (26) are functions of the vector h (length and directions). When that function depends only on the length of the vector h, the model is said to be isotropic. When it depends also on the direction of the vector h, the model is said to be anisotropic. The variograms in a direction (α_1) along continuity, say, of a sedimentation channel, will present less variability than in a direction (α_2) across continuity; correspondingly, the covariance or correlogram will indicate greater correlation in the direction of sedimentation, see Figure 5.

The variance operator

Knowledge of the covariance function $C(h)$ for a random function $\{Z(x), x \in A\}$, or the covariance matrix $[C_{\alpha\beta}, \alpha, \beta = 1, \cdots, n]$ for a set of n RV's $Z_\alpha, \alpha = 1, \cdots, n$, not necessarily related to the same attribute, allows calculation of the variance of any linear combination of the component RV's.

Let:

$$Y = \sum_{\alpha=1}^{n} a_\alpha Z(x_\alpha) \text{ , or } Y = \sum_{\alpha=1}^{n} a_\alpha Z_\alpha$$

then:

$$\begin{aligned} \text{Var}\{Y\} &= \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} a_\alpha a_\beta C(x_\alpha - x_\beta) \geq 0 \\ &= \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} a_\alpha a_\beta C_{\alpha\beta} \geq 0 \end{aligned} \tag{33}$$

A variance is necessarily non-negative, thus the previous expression must be non-negative whatever the choice of the n weights a_α, possibly negative. The variance is nought only if all weights a_α are zero, assuming that none of the component RV's are exactly constant. Thus, the covariance function $C(h)$ or the covariance matrix $[C_{\alpha\beta}]$ must be such to ensure the positivity of the variance operator (33) whatever the weights a_α chosen: This is the condition of "positive definiteness" of either the function $C(h)$ or the matrix $[C_{\alpha\beta}]$

Not all functions nor all matrix of numbers can be taken as a covariance/variogram model, respectively, as a covariance/variogram matrix. When modeling an experimental covariance/variogram, a valid function $C(h)$ or $\gamma(h)$ must be chosen.

Selected Bibliography

Gardner, W. A., 1986, Introduction to Random Processes, MacMillan, 434 p.

Isaaks, E. and Srivastava, R., 1988, "Spatial continuity measures for probabilistic and deterministic geostatis tics,' Math. Geol., v. 20, no. 4.

Journel, A. and Huijbregts, C., 1978, Mining Geostatistics, Chap. II-III, Academic Press, 600 p.

Lindley, D. V., 1965, Introduction to Probability and Statistics: Part 1, Cambridge University Press, 259 p.

Thomas, J. B., 1986, Introduction to Probability, Springer-Verlag, 247 p.

Lesson II: Linear Regression Theory or Simple Kriging (SK)

The basic idea of linear regression is to estimate an unknown value Z_0 by a linear combination of n known values $Z_\alpha, \alpha = 1, \cdots, n$. These n known values may correspond to the same attribute as Z_0, in which case one speaks of "kriging," or to attributes different from Z_0, in which case it is "cokriging." From an algorithmic point of view, there is no difference between kriging and cokriging; the difference appears only at the stage of inference of the covariance/variogram models required for solving for the linear regression weights.

Consider an unknown value z_0, interpreted as an outcome of a random variable (RV) Z_0.

The n data values $z_\alpha, \alpha = 1, \cdots, n$ are themselves interpreted as n outcomes of n RV's $Z_\alpha, \alpha = 1, \cdots, n$.

The $(n+1)$ RV's $Z_0, Z_1, \cdots, Z_n$ are characterized by

- their means, for now assumed known:

$$E\{Z_\alpha\} = m_\alpha \ , \ \alpha = 0, \cdots, n \tag{1}$$

- their covariances:

$$\begin{aligned} \text{Cov}\{Z_\alpha, Z_\beta\} &= E\{Z_\alpha, Z_\beta\} - m_\alpha m_\beta = C_{\alpha\beta} \\ &= \text{for all } \alpha, \beta = 0, \cdots, n. \end{aligned}$$

The $(n+1)$ variances are none other than the autocovariances values:

$$\text{Var}\{Z_\alpha\} = E\{Z_\alpha^2\} - m_\alpha^2 = C_{\alpha\alpha} \ , \ \alpha = 0, \cdots, n \tag{2}$$

One can also define the standardized covariances or coefficient of correlation between any two variables Z_α, Z_β:

$$\rho_{\alpha\beta} = \frac{C_{\alpha\beta}}{\sqrt{C_{\alpha\alpha} \cdot C_{\beta\beta}}} \in [-1, +1] \tag{3}$$

We will distinguish the data covariance matrix $\underline{K}$ (n x n):

$$\underline{K} = [C_{\alpha\beta} \ , \ \alpha, \beta = 1, \cdots, n] \tag{4}$$

from the data-to-unknown covariance column $\underline{k}$ (n x 1):

$$\underline{k}^T = [C_{\alpha 0} \ , \ \alpha = 1, \cdots, n] \tag{5}$$

The critical step of covariance/variogram inference and modeling allows determination of all the required covariance values.

Linear estimator:

The unknown value z_0 is estimated by a linear combination of the n data plus a shift parameter λ_0:

$$z_0^* = \lambda_0 + \sum_{\alpha=1}^{n} \lambda_\alpha z_\alpha \tag{6}$$

z_0^* is said to be the "estimate." Correspondingly, the linear combination of the n RV's Z_α is itself an RV called the "estimator":

$$Z_0^* = \lambda_0 + \sum_{\alpha=1}^{n} \lambda_\alpha Z_\alpha \tag{7}$$

The actual error, $z_0 - z_0^*$, is usually unknown and little can be done about it. However, some moments of the RV error, $Z_0 - Z_0^*$, can be calculated thus acted upon.

The expected value of the RV error is:

$$E\{Z_0 - Z_0^*\} = E\{Z_0\} - E\{Z_0^*\}$$

$$= E\{Z_0\} - \lambda_0 - \sum_{\alpha=1}^{n} \lambda_\alpha E\{Z_\alpha\} = m_0 - \lambda_0 - \sum_{\alpha=1}^{n} \lambda_\alpha m_\alpha$$

We would wish that expected error to be zero ensuring "unbiasedness," thus the shift parameter λ_0 is taken as:

$$\lambda_0 = m_0 - \sum_{\alpha=1}^{n} \lambda_\alpha m_\alpha$$

Thus:

$$Z_0^* = m_0 + \sum_{\alpha=1}^{n} \lambda_\alpha [Z_\alpha - m_\alpha] \tag{8}$$

The unbiased estimator Z_0^* appears as the result of a linear estimation of the residual $[Z_0 - m_0]$ from the residual RV data $[Z_\alpha - m_\alpha]$:

$$[Z_0^* - m_0] = [Z_0 - m_0]^* = \sum_{\alpha=1}^{n} \lambda_\alpha [Z_\alpha - m_\alpha] \qquad (9)$$

There remains to determine the n weights $\lambda_\alpha, \alpha = 1, \cdots, n$. For this, we will act upon (minimize) the error variance. Viewing the RV error $(Z_0 - Z_0^*)$ as a linear combination Y of $(n+1)$ RV's:

$$Y = Z_0 - Z_0^* = [Z_0 - m_0] - [Z_0 - m_0]^* = \sum_{\alpha=0}^{n} a_\alpha [Z_\alpha - m_\alpha] \qquad (10)$$

with: $\begin{cases} a_0 = 1 \\ a_\alpha = -\lambda_\alpha \ , \ \alpha = 1, \cdots, n \end{cases}$

The error variance is expressed as a double sum of the covariances $C_{\alpha\beta}$, see Lesson I, relation (33):

$$\text{Var } Y = \sum_{\alpha=0}^{n} \sum_{\beta=0}^{n} a_\alpha a_\beta C_{\alpha\beta} \qquad (11)$$

We would wish the error variance to be the smallest possible. Thus, the n weights $a_\alpha = -\lambda_\alpha, \alpha = 1, \cdots, n$ are chosen to minimize Var Y. This is done by setting to zero the n partial derivatives of Var Y with regard to each of the n parameters a_α , $\alpha = 1, \cdots, n$:

$$\frac{1}{2} \frac{\partial \text{Var } Y}{\partial a_\alpha} = \sum_{\beta=0}^{n} a_\beta C_{\alpha\beta} = a_0 C_{\alpha 0} + \sum_{\beta=1}^{n} a_\beta C_{\alpha\beta}$$

$$= C_{\alpha 0} - \sum_{\beta=1}^{n} \lambda_\beta C_{\alpha\beta} = 0 \ , \ \text{for all } \alpha = 1, \cdots, n$$

We have just established the well known "normal system" of equations, also known as "linear regression" equations, and belatedly renamed "simple kriging" (SK) system:

$$\sum_{\beta=1}^{n} \lambda_\beta C_{\alpha\beta} = C_{\alpha 0} \ , \ \alpha = 1, \cdots, n \qquad (12)$$

The corresponding minimized error variance, also called "simple kriging" variance, is then written:

$$\sigma_{SK}^2 = \text{Var } \{Z_0 - Z_0^*\} = \sum_{\alpha=0}^{n} \sum_{\beta=0}^{n} a_\alpha a_\beta C_{\alpha\beta}$$

$$= \underbrace{C_{00} - \sum_{\beta=1}^{n} \lambda_\beta C_{0\beta}}_{\alpha=0} - \underbrace{\sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha 0}}_{\alpha\neq 0 \, , \, \beta = 0} + \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C_{\alpha\beta}$$

$$= C_{00} - 2 \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha 0} + \sum_{\alpha=1}^{n} \lambda_\alpha \left[\sum_{\beta=1}^{n} \lambda_\beta C_{\alpha\beta} = C_{\alpha 0} \right]$$

Finally:

$$\sigma_{SK}^2 = \text{Var } \{Z_0 - Z_0^*\} = C_{00} - \sum_{\alpha=1}^{n} \lambda_\alpha C_{\alpha 0} \geq 0 \qquad (13)$$

In matrix notations, the SK system (12) is written:

$$\underline{K} \, \underline{\lambda} = \underline{k} \iff \underline{\lambda} = \underline{K}^{-1} \cdot \underline{k} \qquad (14)$$

with:

$$\underline{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \ , \quad \underline{K} \text{ and } \underline{k} \text{ being the covariance matrices defined in (4),(5)}$$

Defining the residual data matrix

$$\underline{R}_\alpha^T = [Z_1 - m_1, \cdots, Z_n - m_n] \ ,$$

the estimator (9) is written:

$$[Z_0 - m_0]^* = \underline{\lambda}^T \cdot \underline{R}_\alpha = \underline{k}^T \cdot \underline{K}^{-1} \cdot \underline{R}_\alpha \qquad (15)$$

since: $\underline{K} = \underline{K}^T$ (the data covariance matrix is symmetric), $\underline{K}^{-1} = (\underline{K}^{-1})^T$.

The minimized estimation variance is:

$$\sigma_{SK}^2 = C_{00} - \underline{\lambda}^T \cdot \underline{k} = C_{00} - \underline{k}^T \cdot \underline{K}^{-1} \cdot \underline{k}$$

The system, (12) or (14), yields one and only solution as soon as the data covariance matrix $\underline{K}$ is positive definite, in practice as soon as no two columns (or rows) of that matrix are identical, i.e., there is no two data totally redundant:

$$C_{\alpha\beta} = C_{\alpha'\beta} \ , \ \text{for all } \beta \ , \ \text{if and only if: } \alpha = \alpha'.$$

Of course, $\underline{K}$ should be a covariance matrix, i.e., the n^2 values cannot be picked arbitrarily, see Lesson I (33).

Independence case

Assume that the n RV data Z_α , $\alpha = 1, \cdots, n$, are independent one from another; they are then called "independent variables," the unknown RV Z_0 being the "dependent variable."

The data covariance matrix $\underline{K}$ thus reduces to a diagonal matrix with for diagonal elements the data variances:

$$\underline{K} = [C_{\alpha\beta}] = \begin{bmatrix} C_{11} & & \underline{0} \\ & \ddots & \\ \underline{0} & & C_{nn} \end{bmatrix}$$

$$\text{,i.e. } C_{\alpha\beta} = \begin{cases} 0 \ , \ \text{for } \alpha \neq \beta \\ C_{\alpha\alpha} \ , \ \text{for } \alpha = \beta \end{cases}$$

The SK system (12) is then written:

$$\lambda_\alpha C_{\alpha\alpha} = C_{\alpha 0} \implies \lambda_\alpha = C_{\alpha 0} / C_{\alpha\alpha} \ , \ \alpha = 1, \cdots, n$$

Introducing the coefficient of correlation:

$$\rho_{\alpha 0} = C_{\alpha 0}/\sqrt{C_{\alpha\alpha} C_{00}} \implies \lambda_\alpha = \rho_{\alpha 0} \frac{\sqrt{C_{00}}}{\sqrt{C_{\alpha\alpha}}}\ ,\ \alpha = 1, \cdots, n$$

and the SK estimator (9) becomes:

$$[Z_0 - m_0]^* = \sum_{\alpha=1}^{n} \rho_{\alpha 0} \frac{\sqrt{C_{00}}}{\sqrt{C_{\alpha\alpha}}} \cdot [Z_\alpha - m_\alpha]$$

i.e., with the standard deviation notation: $\sigma_\alpha = \sqrt{C_{\alpha\alpha}}$

$$\left[\frac{Z_0 - m_0}{\sigma_0}\right]^* = \sum_{\alpha=1}^{n} \rho_{\alpha 0} \left[\frac{Z_\alpha - m_\alpha}{\sigma_\alpha}\right] \qquad (16)$$

In presence of independent data, the SK (regression) weights λ_α , $\alpha = 1, \cdots, n$, are none other than the correlation coefficients $\rho_{0\alpha}$ between each datum and the unknown (dependent) variable Z_0.

The greater the correlation $\rho_{0\alpha}$ of a particular datum with Z_0, the greater its weight.

Note that:

$\sum_{\alpha=1}^{n} \rho_{\alpha 0} \neq 1$, usually.

$$\sigma^2_{SK} = C_{00}\left[1 - \sum_\alpha \lambda_\alpha \frac{C_{\alpha 0}}{C_{00}}\right] = C_{00}[1 - \sum_\alpha \rho^2_{\alpha 0}]$$

Moreover, if the n data are independent of the unknown Z_0, i.e., if $\rho_{\alpha 0} = 0$, the SK estimator reduces to the mean: $Z_0^* = m_0$.

Example. $n = 1$ (1 single datum Z_1).

Expressions (12) or (16) reduces to:

$$\left[\frac{Z_0 - m_0}{\sigma_0}\right]^* = \rho_{10}\left[\frac{Z_1 - m_1}{\sigma_1}\right] \ ,\ \text{i.e.}$$

$$Z_0^* = m_0 + \rho_{10}\frac{\sigma_0}{\sigma_1}(Z_1 - m_1)\ ,$$

which is the traditional expression of the regression line related to a scattergram $\left(z_0^{(i)}, z_1^{(i)}, i = 1, \cdots, L\right)$, see Figure 1:

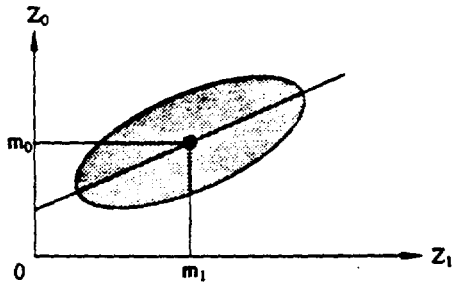


Fig. 1. Scattergram (Z_0, Z_1)

$$m_0 = \frac{1}{L}\sum_{i=1}^{L} z_0^{(i)}\ ,\ m_1 = \frac{1}{L}\sum_{i=1}^{L} z_1^{(i)}$$
$$\sigma_0^2 = \frac{1}{L}\sum_{i=1}^{L}\left[z_0^{(i)} - m_0\right]^2\ ,\ \sigma_1^2 = \frac{1}{L}\sum_{i=1}^{L}\left[z_1^{(i)} - m_1\right]^2$$
$$\sigma_{10} = \frac{1}{L}\sum_{i=1}^{L}\left[z_0^{(i)} - m_0\right]\cdot\left[z_1^{(i)} - m_1\right],$$
$$\rho_{10} = \sigma_{10}/\sigma_1\sigma_0 \in [-1, +1]$$

$m_0, m_1, \sigma_0^2, \sigma_1^2, \rho_{10}$ are the experimental statistics of the L pairs $\left(z_0^{(i)}, z_1^{(i)}\right)$, $i = 1, \cdots, L$, available.

The exactitude property

Consider the case when one of the data, say, $Z_{\alpha'}$, is considered the unknown:
$Z_{\alpha'} \equiv Z_0$. We would then wish the SK estimate to identify that single datum.
In such a case: $C_{\alpha\alpha'} = C_{\alpha 0}$, for all $\alpha = 1, \cdots, n$, and the SK system (12) becomes:

$$\sum_{\substack{\beta=1 \\ \beta\neq\alpha'}}^{n} \lambda_\beta C_{\alpha\beta} + \lambda_{\alpha'} C_{\alpha 0} = C_{\alpha 0}\ ,\ \alpha = 1, \cdots, n$$

The unique solution is: $\lambda_{\alpha'} = 1$, $\lambda_\beta = 0$, for all $\beta \neq \alpha'$.

Thus, the SK estimate is:

$$Z_0^* = Z_{\alpha'}\ ,\ \text{and}\ \sigma^2_{SK} = 0 \qquad (17)$$

Spatial interpolation

Spatial interpolation corresponds to the case where the n data Z_α relate to the same attribute Z but at different locations $x_\alpha \neq x_0$.

For example, the porosity $Z(x_0)$ at unsampled location x_0 is to be estimated from n porosity data $Z(x_\alpha)$ taken at n surrounding locations x_α:

$$Z^*(x_0) - m = \sum_{\alpha=1}^{n} \lambda_\alpha [Z(x_\alpha) - m] \qquad (18)$$

with m being the common expected value of the $(n+1)$ RV's $Z(x_0)$, $Z(x_\alpha)$, $\alpha = 1, \cdots, n$. In practice, m would be the mean of all porosity data available within the same layer or homogeneous formation.

Defining the covariance function $C(h)$ as the covariance between any two porosity RV's $Z(x)$, $Z(x+h)$, distant of vector h but still within the same formation:

$$C(h) = \text{Cov}\{Z(x)\ ,\ Z(x+h)\} \qquad (19)$$

The simple kriging system (12) is:

$$\sum_{\beta=1}^{n} \lambda_\beta C(x_\alpha - x_\beta) = C(x_\alpha - x_0)\ ,\ \alpha = 1, \cdots, n \qquad (20)$$

Remark. The SK estimator (18) requires prior knowledge or estimation of the common mean m. That prior step can be pooled with the error variance minimization step into a single algorithm, called "ordinary kriging," see Lesson III, system (4).

Simple cokriging

When the n data Z_α relate to one or more (up to n) attributes different from the attribute corresponding to the unknown Z_0, the previous SK process is called "simple cokriging."

For example, Z_0 may be the sought after core porosity at any particular given location. The Z_α's may be n porosity-related well log data at that same location. The corresponding simple cokriging estimate is none other than the traditional linear regression of core porosity from the log data:

$$[Z_0 - m_0]^* = \sum_{\alpha=1}^{n} \lambda_\alpha [Z_\alpha - m_\alpha] \text{ , according to (9)}$$

Remark. If one of the data, say, $Z_{\alpha'}$, happens to be either the sought after core porosity, or a log datum perfectly correlated to the core porosity Z_0, the previous SK estimate would identify that datum:

$$Z_0^* \equiv Z_{\alpha'} \text{ , see the exactitude property (17).}$$

The previous linear regression considers for estimation of the core porosity Z_0 only those well log data at the same location. In fact, there also exists core data at other locations. If these data at other locations are correlated to Z_0, they should be included to improve the regression defining a full cokriging pattern.

Consider for the estimation of an unknown porosity $Z_0(x_0)$ at a location x_0 the information constituted by:

- n_0 core porosity data $Z_0(x_{\alpha_0})$, $\alpha_0 = 1, \cdots, n_0$ obtained from neighboring cored well(s)
- K sets of log data $Z_k(x_{\alpha_k})$, $\alpha_k = 1, \cdots, n_k$, $k = 1, \cdots, K$ coming from various locations x_{α_k} corresponding to either cored or simply logged wells.

The generalized linear regression, or "cokriging" estimator, is written:

$$\begin{aligned}[Z_0(x_0) - m_0]^* &= \sum_{\alpha_0=1}^{n_0} \lambda_{\alpha_0} [Z_0(x_{\alpha_0}) - m_0] \\ &+ \sum_{k=1}^{K} \sum_{\alpha_k=1}^{n_k} \lambda_{\alpha_k} [Z_k(x_{\alpha_k}) - m_k]\end{aligned} \tag{21}$$

This expression is no different from the general SK expression (9). The total number of data being:

$$n = n_0 + \sum_{k=1}^{K} n_k$$

The corresponding cokriging system to determine the n weights, λ_{α_0} , λ_{α_k}, is no different from the general SK system (12) except for the notations differentiating each type of log data, see hereafter System (22).

The traditional regression using only the data available at the location to be informed fails to use existing spatial correlations existing between further away data and the attribute $Z_0(x_{\alpha_0})$ to be estimated.

Simple cokriging system

$$\begin{aligned}&\sum_{\beta_0=1}^{n_0} \lambda_{\beta_0} C_{\alpha_0\beta_0} + \sum_{k=1}^{K} \sum_{\beta=1}^{n_k} \lambda_{\beta_k} C_{\alpha_0\beta_k} = C_{\alpha_0 0} \\ &\text{for } \alpha_0 = 1, \cdots, n_0 \\ &\sum_{\beta_0=1}^{n_0} \lambda_{\beta_0} C_{\alpha_{k'}\beta_0} + \sum_{k=1}^{K} \sum_{\beta=1}^{n_k} \lambda_{\alpha_{k'}} C_{\beta_{k'}\beta_k} = C_{\alpha_{k'}0} \\ &\text{for } \alpha_{k'} = 1, \cdots, n_{k'} \text{ and } k' = 1, \cdots, K\end{aligned} \tag{22}$$

with:

$$\begin{aligned}C_{\alpha_0\beta_0} &= \text{Cov}\{Z_0(x_{\alpha_0}) , Z_0(x_{\beta_0})\} \\ C_{\alpha_0\beta_k} &= \text{Cov}\{Z_0(x_{\alpha_0}) , Z_k(x_{\beta_k})\} \\ C_{\alpha_0 0} &= \text{Cov}\{Z_0(x_{\alpha_0}) , Z_0(x_0)\} \\ C_{\alpha_{k'}\beta_0} &= \text{Cov}\{Z_{k'}(x_{\alpha_{k'}}) , Z_0(x_{\beta_0})\} \\ C_{\alpha_{k'}\beta_k} &= \text{Cov}\{Z_{k'}(x_{\alpha_{k'}}) , Z_k(x_{\beta_k})\} \\ C_{\alpha_{k'}0} &= \text{Cov}\{Z_{k'}(x_{\alpha_{k'}}) , Z_0(x_0)\}\end{aligned}$$

In matrix terms, the general system (14) holds true with simply the various matrices being partitioned, e.g.,

$$\underline{K} = \begin{bmatrix} [C_{\alpha_0\beta_0}] & [C_{\alpha_0\beta_1}] & \cdots & [C_{\alpha_0\beta_K}] \\ & [C_{\alpha_1\beta_1}] & \cdots & [C_{\alpha_1\beta_K}] \\ & & \ddots & \vdots \\ & & & [C_{\alpha_K\beta_K}] \end{bmatrix} \tag{23}$$

Data orthogonalization

In most practical situations, the n data $Z_\alpha, \alpha = 1, \cdots, n$ are dependent on each other. The idea is to define from the initial set of dependent data $Z_\alpha, \alpha = 1, \cdots, n$, a new set of independent data $Y_\alpha, \alpha = 1, \cdots, n$, on which the simpler regression (16) could be applied.

Given any initial data vector: $\underline{Z}_\alpha^T = [Z_1, \cdots, Z_n]$ with non-diagonal covariance matrix $\underline{K}$, there exists a linear transform $\underline{Y}_\alpha = \underline{A} \cdot \underline{Z}_\alpha$ which will transform $\underline{Z}_\alpha$ into a vector $\underline{Y}_\alpha$ with an identity diagonal covariance matrix, i.e., the n transformed RV's $\underline{Y}_\alpha$ are uncorrelated.

The initial matrix $\underline{K}$ being a covariance matrix always admits the following LU decomposition:

$$\underline{\underline{K}} = \underline{\underline{L}} \cdot \underline{\underline{U}} \qquad (24)$$

with: $\underline{\underline{L}}$ being a lower triangular matrix [\º]
and: $\underline{\underline{U}} = \underline{\underline{L}}^T$ being an upper triangular matrix [º\],
[Golub and van Loan 1983, p. 54].

The inverse of $\underline{\underline{K}}$ always exists and is written:

$\underline{\underline{K}}^{-1} = \underline{\underline{U}}^{-1} \cdot \underline{\underline{L}}^{-1} = \underline{\underline{L}}^{-1} \cdot \underline{\underline{U}}^{-1}$, since $\underline{\underline{K}}$ and $\underline{\underline{K}}^{-1}$ are symmetric

Consider the linear transform $\underline{\underline{A}} = \underline{\underline{L}}^{-1}$:

$$\underline{Y}_\alpha = \underline{\underline{L}}^{-1} \cdot \underline{Z}_\alpha \;\Rightarrow\; \underline{Y}_\alpha^T = \underline{Z}_\alpha^T \cdot \underline{\underline{U}}^{-1} \qquad (25)$$

The covariance matrix of the Y_α's is written:

$$\begin{aligned} \underline{\underline{K}}_Y = E\left\{\underline{Y}_\alpha \cdot \underline{Y}_\alpha^T\right\} &= E\left\{\underline{\underline{L}}^{-1} \underline{Z}_\alpha \underline{Z}_\alpha^T \underline{\underline{U}}^{-1}\right\} \\ &= \underline{\underline{L}}^{-1} \cdot E\left\{\underline{Z}_\alpha \underline{Z}_\alpha^T\right\} \cdot \underline{\underline{U}}^{-1} \\ &= \underline{\underline{L}}^{-1} \underline{\underline{K}}\, \underline{\underline{U}}^{-1} = \underline{\underline{L}}^{-1} (\underline{\underline{L}}\,\underline{\underline{U}}) \underline{\underline{U}}^{-1} = \underline{\underline{I}} \end{aligned}$$

Thus, the transform $\underline{Y}_\alpha = \underline{\underline{L}}^{-1} \cdot \underline{Z}_\alpha$ "uncorrelates" the data. The new data Y_α, however, remains correlated to the unknown RV Z_0, with for covariance matrix:

$$\underline{k}_Y = E\{Z_0 \underline{Y}_\alpha\} = \underline{\underline{L}}^{-1} \cdot E\{Z_0 \underline{Z}_\alpha\} = \underline{\underline{L}}^{-1} \cdot \underline{k}$$

Consider the SK estimate based on the Y_α data:

$$[Z_0 - m_0]^{**} = \sum_{\alpha=1}^{n} \nu_\alpha [Y_\alpha - m_{Y_\alpha}] = \underline{\nu}^T \cdot \underline{\underline{L}}^{-1} \cdot \underline{R}_\alpha$$

The corresponding SK system (14) is written, with $\underline{\underline{K}}_Y = \underline{\underline{I}}$

$$\underline{\underline{I}} \cdot \underline{\nu} = \underline{k}_Y \;\Rightarrow\; \underline{\nu} = \underline{k}_Y$$

Thus: $[Z_0 - m_0]^{**} = \underline{k}_Y^T \cdot \underline{\underline{L}}^{-1} \cdot \underline{R}_\alpha = \underline{k}^T \cdot \left(\underline{\underline{U}}^{-1} \underline{\underline{L}}^{-1}\right) \cdot \underline{R}_\alpha = \underline{k}^T \cdot \underline{\underline{K}}^{-1} \cdot \underline{R}_\alpha$
i.e. $[Z_0 - m_0]^{**} \equiv [Z_0 - m_0]^*$, as defined in (15).

Thus, the SK algorithm as described by the system (14) is the sequence of two steps:

(i) an orthogonalization of the data

$$\underline{Y}_\alpha = \underline{\underline{L}}^{-1} \cdot \underline{Z}_\alpha \text{ , with : } \underline{\underline{K}} = \underline{\underline{L}} \cdot \underline{\underline{L}}^T$$

(ii) a straight weighting of the orthogonal data Y_α by their correlation with the unknown Z_0:

$$[Z_0 - m_0]^* = k_Y^T \cdot [Y_\alpha - m_{Y_\alpha}]$$

The first step accounts for the redundancy between the data Z_α. Once that redundancy is corrected, a straightforward correlation weighting, of type (16), is performed.

Dual kriging

Going back to the matrix expression (15) of the SK estimate:

$$[Z_0 - m_0]^* = \sum_{\alpha=1}^{n} \lambda_\alpha [Z_\alpha - m_\alpha] = \underline{k}^T \cdot \underline{\underline{K}}^{-1} \cdot \underline{R}_\alpha \ ,$$

it can be rewritten as:

$$\begin{aligned} [Z_0 - m_0]^* &= \left(\underline{k}^T \cdot \underline{\underline{K}}^{-1} \cdot \underline{R}_\alpha\right)^T = \underline{R}_\alpha^T \cdot \underline{\underline{K}}^{-1} \cdot \underline{k} \\ &= \underline{b}^T \cdot \underline{k} = \textstyle\sum_{\alpha=1}^{n} b_\alpha C_{\alpha 0} \end{aligned} \qquad (26)$$

with: $\underline{b}^T = [b_1, \cdots, b_n] = \underline{R}_\alpha^T \cdot \underline{\underline{K}}^{-1}$

Thus: $\underline{b} = \underline{\underline{K}}^{-1} \cdot \underline{R}_\alpha$ and the coefficients b_α are given by the system, called "dual SK" system:

$$\underline{\underline{K}}\, \underline{b} = \underline{R}_\alpha \ , \text{ i.e.} \qquad (27)$$

$$\sum_{\beta=1}^{n} b_\beta C_{\alpha\beta} = z_\alpha - m_\alpha \ , \quad \alpha = 1, \cdots, n.$$

The dual expression (26) and the dual system (27) are both extremely similar to the original (primal) expression (15) and system (14). Instead of being read as a linear combination of the n data $(z_\alpha - m_\alpha)$, the SK estimate is read as a linear combination of n interpolation functions which happen to be the n data-to-unknown covariance functions $C_{\alpha o}$.

In the case (18) of a single attribute $Z(x)$ distributed over various locations of a formation, relation (26) is written:

$$Z^*(x_0) - m = \sum_{\alpha=1}^{n} \lambda_\alpha [Z(x_\alpha) - m] = \sum_{\alpha=1}^{n} b_\alpha C(x_\alpha - x_0) \qquad (28)$$

and the dual system (27) amounts to ascertain that the interpolated surface $Z^*(x_0)$, $x_0 \in A$, identifies the data values $z(x_\alpha)$ at data locations, i.e. when $x_0 = x_\alpha$:

$$\begin{aligned} Z^*(x_\alpha) - m = \; & \textstyle\sum_{\beta=1}^{n} b_\beta C(x_\beta - x_\alpha) = r_\alpha = z_\alpha - m \\ & \text{for all } \alpha = 1, \cdots, n \end{aligned} \qquad (29)$$

Spline interpolation amounts to replace in expression (28) the covariance functions $C(x_\alpha - x)$ by extremely regular and smooth functions $f(x_\alpha - x)$, then ensure data identification through a system of type (29):

$$Z^*(x) = b_0 + \sum_{\alpha=1}^{n} b_\alpha f(x_\alpha - x) \qquad (30)$$

with:

$$Z^*(x_\alpha) = b_0 + \sum_{\beta=1}^{n} b_\beta f(x_\beta - x_\alpha) = \text{ datum } z(x_\alpha) \text{ for all } \alpha = 1, \cdots, n \qquad (31)$$

Then one may argue on the arbitrariness of choosing, in the SK case, covariance functions for interpolation functions. As was shown before, the choice of covariance functions allows to minimize the resulting error variance, Var $\{Z_0 - Z_0^*\}$.

Selected Bibliography

Aboufirassi, M. and Mariño, M., 1984, "Cokriging of aquifer transmissivity and specific capacity," Math. Geology, v. 16, no. 1, pp. 19-36.

Buxton, B., 1982, Coal Reserves Assessment: A geostatistical case study (M.S. thesis), Branner Earth Sciences Library, Stanford University, 144 p.

David, M., 1976, "The practice of kriging," in Advanced Geostatistics in the Mining Industry, ed. Guarascio, et al., publ. Reidel, Holland, pp. 31-48.

Davis, M. and Culhane, P., 1984, Contouring very large data sets using kriging, in Geostatistics for Natural Resources Characterization, ed. Verly et al., publ. Reidel, Holland, Part 2, 599-619.

Delhomme, J. P., 1978, Kriging in hydroscience, in Advanced Water Resources 1(5), pp. 251-266.

Doyen, P., 1986, "Porosity mapping from seismic velocity data: An approach based on spatial correlation functions," submitted to Geophysics.

Draper, N. and Smith, H., 1966, Applied regression analysis, Wiley & Sons, 709 p.

Golub, G. and Van Loan, C., 1983, Matrix computations, John Hopkins Univ. Press, 476 p.

Journel, A. and Huijbregts, C., 1978, Mining Geostatistics, Chap. 5, Academic Press, 600 p.

Lapin, L., 1983, Probability and Statistics for Modern Engineering, Brooks/Cole publ., 624 p.

Myers, D. E., 1984, Cokriging - New developments, in Geostatistics for Natural Resources Characterization, ed. Verly, et al., publ. Reidel, Holland, Part 1, pp. 295-305.

Weisberg, S., 1947, Applied Linear Regression. Revised ed. 1985, Wiley & Sons, 324 p.

Lesson III: Linear Regression Under Constraint(s) and Ordinary Kriging (OK)

In all the simple kriging (SK) developments of the previous Lesson II, the mean(s) were supposedly known, cf. relations (9) and (18). Consider, for example, the SK estimator (18) of an unknown $Z(x_0)$ from n data related to the same attribute but at different locations $x_\alpha, \alpha = 1, \cdots, n$:

$$Z^*(x_0) - m = \sum_{\alpha=1}^{n} \lambda_\alpha [Z(x_\alpha) - m] \qquad (1)$$

The common (stationary) mean m is supposed known. If the process $Z(x)$ was sampled at all locations x within the formation A considered, that mean m would be known but also there would not be any estimation problem left. Thus, in practice, either m is to be estimated prior to the SK algorithm, or an estimation algorithm should be designed which does not require prior estimate of that mean.

Consider the linear estimator:

$$Z^*(x_0) = \lambda_0 + \sum_{\alpha=1}^{n} \lambda_\alpha Z(x_\alpha)$$

The error mean is:

$$E\{Z(x_0) - Z^*(x_0)\} = m - \lambda_0 - \sum_{\alpha=1}^{n} \lambda_\alpha m = -\lambda_0 + m\left(1 - \sum_{\alpha} \lambda_\alpha\right)$$

and should be set to zero, whatever the unknown value of m. This can be achieved only if:

$$\begin{cases} \lambda_0 = 0 \\ \sum_{\alpha=1}^{n} \lambda_\alpha = 1 \end{cases}$$

Thus, an unbiased linear estimator of $Z(x_0)$ is written as:

$$Z^*(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(x_\alpha) \ , \ \text{with}: \ \sum_{\alpha=1}^{n} \lambda_\alpha = 1 \qquad (2)$$

Note. Although for reason of convenience, we have used the same notation λ_α for the SK weights of relation (1) and the weights of relation (2) they are not equal.

The error variance is written, as in Lesson II (11):

$$\sigma_E^2 = \text{Var}\{Z(x_0) - Z^*(x_0)\} = \sum_{\alpha=0}^{n} \sum_{\beta=0}^{n} a_\alpha a_\beta C(x_\alpha - x_\beta) \qquad (3)$$

with: $\begin{cases} a_0 = 1 \\ a_\alpha = -\lambda_\alpha, \alpha = 1, \cdots, n \end{cases}$

and: $C(x_\alpha - x_\beta) = \text{Cov}\{Z(x_\alpha), Z(x_\beta)\}$

We wish to minimize that error variance, still ensuring unbiasedness of the estimator, i.e., ensuring the constraint: $\sum_{\alpha=1}^{n} \lambda_\alpha = 1$. This amounts to an optimization under linear constraint. The Lagrange formalism will be used whereby a function, of the $(n + 1)$ parameters

$\lambda_\alpha, \alpha = 1, \cdots, n$ and 2μ being the Lagrange parameter, is defined as:

$$S(\lambda_\alpha, \alpha = 1, \cdots, n\,;\,\mu) = \sigma_E^2 + 2\mu \left[\sum_{\alpha=1}^{n} \lambda_\alpha - 1 \right]$$

The extreme of that function S is obtained by setting to zero its $(n+1)$ partial derivatives:

$$-\tfrac{1}{2}\,\frac{\partial S}{\partial \lambda_\alpha} = \textstyle\sum_{\beta=0}^{n} a_\beta C(x_\alpha - x_\beta) - \mu = 0 \;,\quad \alpha = 1, \cdots, n$$

$$-\tfrac{1}{2}\,\frac{\partial S}{\partial \mu} = \textstyle\sum_{\alpha=1}^{n} \lambda_\alpha - 1 = 0$$

The n first equations are rewritten:

$$C(x_\alpha - x_0) - \sum_{\beta=1}^{n} \lambda_\beta C(x_\alpha - x_\beta) - \mu = 0 \;,\ \text{i.e.,}$$

$$\sum_{\beta=1}^{n} \lambda_\beta C(x_\alpha - x_\beta) + \mu = C(x_\alpha - x_0) \;,\ \alpha = 1, \cdots, n.$$

Finally, the $(n+1)$ unknowns, the λ_α's and μ, are given by a system of $(n+1)$ linear equations, known as the "constrained normal" system of equations, belatedly renamed "ordinary kriging" (OK) system:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta C(x_\alpha - x_\beta) + \mu = C(x_\alpha - x_0) \;,\ \alpha = 1, \cdots, n \\ \sum_{\beta=1}^{n} \lambda_\beta = 1 \end{cases} \tag{4}$$

The corresponding minimized error variance, also called "ordinary kriging" (OK) variance is written:

$$\begin{aligned} \sigma_{OK}^2 &= E\{[Z(x_0) - Z^*(x_0)]^2\} = \\ &= C(0) - \textstyle\sum_{\alpha=1}^{n} \lambda_\alpha C(x_\alpha - x_\beta) - \mu \geq 0 \end{aligned} \tag{5}$$

Just like the SK system of Lesson II (12), the OK system (4) presents one and only one solution as soon as the data covariance matrix $\underline{K} = [C(x_\alpha - x_\beta)]$ is positive definite, in practice as soon as:

(i) the covariance function $C(h)$ is licit, i.e., is a positive definite function, see Lesson I condition (34)

(ii) there are no two data locations totally redundant, i.e.,

$$C(x_\alpha - x_\beta) = C(x_{\alpha'} - x_\beta) \text{, for all } \beta = 1, \cdots, n \text{,}$$
$$\text{if and only if: } \alpha = \alpha'$$

Independence case

If the n RV data $Z(x_\alpha)$ are independent (or uncorrelated) one from another, the data covariance matrix $[C(x_\alpha - x_\beta)]$ reduces to a diagonal matrix with all elements of the diagonal equal to the common variance $C(0) = \text{Var}\,\{Z(x)\}$, for all x.
The OK system (4) becomes:

$$\begin{cases} \lambda_\alpha C(0) + \mu = C(x_\alpha - x_0) \;,\ \alpha = 1, \cdots, n \\ \sum_\alpha \lambda_\beta = 1 \end{cases} \tag{6}$$

i.e.

$$\begin{cases} \lambda_\alpha = \frac{C(x_\alpha - x_0)}{C(0)} - \frac{\mu}{C(0)} = \rho(x_\alpha - x_0) - \frac{\mu}{C(0)} \\ -\mu = \frac{C(0)}{n}\,[1 - \sum_{\alpha=1}^{n} \rho(x_\alpha - x_0)] \end{cases}$$

$$\begin{aligned} \sigma_{OK}^2 &= C(0)\,[1 - \textstyle\sum_{\alpha=1}^{n} \rho^2(x_\alpha - x_0)] \\ &\quad -\mu\,[1 - \textstyle\sum_{\alpha=1}^{n} \rho(x_\alpha - x_0)] \geq 0 \end{aligned}$$

with: $\rho(h) = C(h)/C(0) \in [-1, +1]$ being the standardized covariance, or "correlogram," measuring the correlation between two values $Z(x), Z(x+h)$ distant of vector h.

If, moreover, the n data are uncorrelated with the unknown RV $Z(x_0)$:
$\rho(x_\alpha - x_0) = 0 \;,\ \alpha = 1, \cdots, n$, the OK system (6) then yields:

$$\begin{cases} \lambda_\alpha = 1/n \;,\ \alpha = 1, \cdots, n \\ -\mu = C(0)/n \end{cases} \tag{7}$$

Thus: $Z^*(x_\alpha) = \frac{1}{n} \sum_{\alpha=1}^{n} Z(x_\alpha)$, whatever the location x_0 and: $\sigma_{OK}^2 = C(0) + \frac{C(0)}{n} \geq 0$

In the case of total spatial independence, the OK estimate reduces to the arithmetic mean of the n data retained.

In the SK case Lesson II (16), the SK estimate did reduce to the known mean m_0. In the OK case, that mean is unknown and is estimated by the mean of the n data.

The exactitude property

If one datum, say, $Z(x_{\alpha'})$, is considered to be the unknown $Z(x_0)$, then:
$C(x_{\alpha'} - x_\beta) = C(x_0 - x_\beta)$, for all $\beta = 1, \cdots, n$, and the OK system (4) is written:

$$\begin{cases} \sum_{\substack{\beta=1 \\ \beta \neq \alpha'}}^{n} \lambda_\beta C(x_\alpha - x_\beta) + \lambda_{\alpha'} C(x_\alpha - x_0) + \mu = C(x_\alpha - x_0) \;, \\ \alpha = 1, \cdots, n \\ \sum_{\beta=1}^{n} \lambda_\beta = 1 \end{cases}$$

The unique solution is: $\lambda_{\alpha'} = 1, \lambda_\beta = 0 \beta \neq \alpha', \mu = 0$.
Thus, the OK estimate identifies the datum value $Z(x_{\alpha'}) = Z(x_0)$, and:

$$\sigma_{OK}^2 = C(0) - \lambda_{\alpha'} C(0) = 0 \tag{8}$$

As SK, the OK algorithm features the exactitude property: the OK surface, $Z^*(x_0)$, $x_0 \in$ Domain, honors the data values at data locations.

Similarity SK-OK

Rewriting the expression Lesson II (18) of the SK estimator as:

$$Z^*_{SK}(x_0) = \left(1 - \sum_{\alpha=1}^{n} \lambda_\alpha\right) m + \sum_{\alpha=1}^{n} \lambda_\alpha Z(x_\alpha) \qquad (9)$$

It appears as a linear combination of $(n+1)$ data, including the known mean m, with the sum of the $(n+1)$ weights being equal to 1 by construction. In the OK case, the mean is unknown, thus there are only n weights which must sum up to 1, calling for a constrained optimization and a Lagrange parameter.

Example. $n = 1$ (1 single datum)

The unbiasedness condition (2) fully determines the unique weight: $\lambda_1 = 1$, and: $Z^*(x_0) = Z(x_1)$.

The OK system (4) determines the Lagrange parameter value:

$$\lambda_1 C(0) + \mu = C(x_1 - x_0) \Rightarrow \mu = C(x_1 - x_0) - C(0)\,,$$

Thus:

$$\sigma^2_{OK} = E\{[Z(x_0) - Z(x_1)]^2\} = 2\,[C(0) - C(x_1 - x_0)] \geq 0$$

More generally, the basic estimation variance of one value $Z(x)$ by another $Z(x+h)$, a vector h apart, isby definition the "variogram," see Lesson I (31):

$$2\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} = 2[C(0) - C(h)] \quad (10)$$

Example. $n = 2$ (2 data), see Figure 1

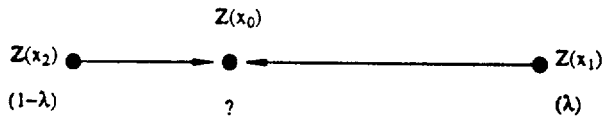


Fig. 1. Regression from 2 data

The unbiasedness condition defines the OK estimator as:

$$Z^*(x_0) = \lambda Z(x_1) + (1-\lambda) Z(x_2)$$

The OK system (4) is written:

$$\begin{cases} \lambda C(0) + (1-\lambda) C(x_1 - x_2) + \mu = C(x_1 - x_0) \\ \lambda C(x_1 - x_2) + (1-\lambda) C(0) + \mu = C(x_2 - x_0) \end{cases}$$

After some elementary algebra, it is found that:

$$\lambda = \frac{1}{2} + \frac{1}{2} \frac{C(x_1 - x_0) - C(x_2 - x_0)}{C(0) - C(x_1 - x_2)}$$

Introducing the correlogram $\rho(h) = C(h)/C(0)$, it comes:

$$\lambda = \frac{1}{2} + \frac{1}{2} \frac{\rho(x_1 - x_0) - \rho(x_2 - x_0)}{1 - \rho(x_1 - x_2)} \qquad (11)$$

As expected, the weight λ given to the datum $Z(x_1)$ increases as

(i) the correlation $\rho(x_1 - x_0)$ of that datum with the unknown $Z(x_0)$ increases

(ii) the correlation $\rho(x_2 - x_0)$ of the other datum with $Z(x_0)$ decreases.

Similarly, it is found:

$$\mu = \frac{C(0)}{2}\,[\rho(x_1 - x_0) + \rho(x_2 - x_0) - \rho(x_1 - x_2) - 1] \qquad (12)$$

$$\sigma^2_{OK}/C(0) = 1 - \lambda\rho(x_1 - x_0) - (1-\lambda)\rho(x_2 - x_0) - \mu/C(0)$$

Thus, and as expected, the estimation variance σ^2_{OK} decreases, i.e., the estimation precision increases as

(i) the two data are better correlated with the unknown $Z(x_0)$, i.e., the two correlations $\rho(x_1 - x_0)$ and $\rho(x_2 - x_0)$ increase

(ii) the two data are less redundant: if $\rho(x_1 - x_2)$ decreases, μ will increase and σ^2_{OK} decreases.

As already pointed out in the previous Lesson II, the regression (kriging) algorithm accounts not only for the respective correlation $\rho(x_\alpha - x_0)$ between each datum and the unknown, but also for the degree of redundancy between the data through the data covariance matrix $\underline{K}$.

If the two data of the last example are uncorrelated (no redundancy): $\rho(x_1 - x_2) = 0$, and expression (11) gives:

$$\lambda = \frac{1}{2} + \frac{1}{2}\,[\rho(x_1 - x_0) - \rho(x_2 - x_0)]$$

The weight λ increases directly as the excess of correlation $[\rho(x_1 - x_0) - \rho(x_2 - x_0)]$ of datum $Z(x_1)$ over datum $Z(x_2)$. If, moreover, the two data are equi-correlated with the unknown: $\lambda = (1-\lambda) = 1/2$, and:

$$\sigma^2_{OK}/C(0) = \frac{3}{2} - 2\rho(x_1 - x_0)$$

If, moreover, the two data are uncorrelated with the unknown

$$\rho(x_1 - x_0) = \rho(x_2 - x_0) = 0 \ , \ \text{and :} \qquad (13)$$

$$\sigma^2_{OK} = \frac{3}{2}C(0) = C(0) + \frac{C(0)}{2} \ , \ \text{as it should.}$$

Kriging in presence of a trend

In presence of an unknown, but constant, mean m, the condition $\sum_\alpha \lambda_\alpha = 1$ allows to filter out that constant m from the OK system and in the same time achieves unbiasedness.

One can generalize that concept and seek for conditions on the n weights λ_α that would filter out a more complex trend function $m(x) = E\{Z(x)\}$ that would be location x-dependent.

Consider a trend function, i.e., a smoothly variable function of the coordinate(s) x, which is of known shape but with unknown parameters:

$$m(x) = \sum_{\ell=0}^{L} a_\ell f_\ell(x) \qquad (14)$$

where the $f_\ell(x)$'s are known functions of the coordinate(s) and the a_ℓ's are unknown parameter values making the trend value $m(x)$ unknown.

By convention, the zeroth term is the constant function, i.e.,

$$f_0(x) = 1 \ , \ \text{for all } x$$

The case of a constant mean, $m(x) = m$, would correspond to $L = 0$, and $a_0 = m$ either known (SK case) or unknown (OK case).

Examples. In $\Re^1$ (e.g., 1 dimensional time space), the $f_\ell(t)$ can be cosine functions of given frequencies ω_ℓ:

$$f_\ell(t) = \cos 2\pi\omega_\ell t$$

In $\Re^2$ (e.g., horizontal space), the $f_\ell(x_1, x_2)$ are usually taken as low order polynomials, up to order 2 for a quadratic trend:

$$m(x,y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy$$

corresponding to the choice $L = 5$, and:

$$f_0(x,y) = 1 \ , \ f_1(x,y) = x \ , \ f_2(x,y) = y$$

$$f_3(x,y) = x^2 \ , \ f_4(x,y) = y^2 \ , \ f_5(x,y) = xy$$

A linear trend in $\Re^2$ would be defined as:

$$m(x,y) = a_0 + a_1 x + a_2 y$$

If that linear trend is known to be limited to the 45° direction, the number of parameters can be limited to two, i.e., $L = 1$:

$$m(x,y) = a_0 + a_1(x + y) \qquad (15)$$

At data locations $x_\alpha, \alpha = 1, \cdots, n$, only the $Z(x_\alpha)$ values are known, not the trend values $m(x_\alpha)$, because the notion of trend is usually a model of the experimenter not necessarily intrinsic to the spatial distribution of the attribute Z.

The problem consists of estimating at any unsampled location $x_0 \neq x_\alpha$, the unknown value $Z(x_0)$ ensuring the unbiasedness condition:

$$E\{Z(x_0) - Z^*(x_0)\} = 0 \ , \ \text{i.e.}$$

$$E\{Z^*(x_0)\} = E\{Z(x_0)\} = m(x_0) = \sum_{\ell=0}^{L} a_\ell f_\ell(x_0)$$

Consider the linear estimator:

$$Z^*(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(x_\alpha) \qquad (16)$$

Note. Again, for reason of convenience, we have adopted for the weights the same notation λ_α used for the SK estimator (1) and OK estimator (2). Although sharing the same notation, these weights need not be equal.

The expected value of the RV error is to be set to zero:

$$E\{Z(x_0) - Z^*(x_0)\} = m(x_0) - \sum_{\alpha=1}^{n} \lambda_\alpha m(x_\alpha) =$$

$$\sum_{\ell=0}^{L} a_\ell \left[f_\ell(x_0) - \sum_{\alpha=1}^{n} \lambda_\alpha f_\ell(x_\alpha) \right] = 0$$

Unbiasedness, whatever the unknown parameters a_ℓ's, is thus achieved by imposing the following $(L+1)$ conditions on the n weights λ_α:

$$\sum_{\alpha=1}^{n} \lambda_\alpha f_\ell(x_\alpha) = f_\ell(x_0) \ , \ \text{for all } \ell = 0, \cdots, L \qquad (17)$$

- The OK case corresponds to $L = 0$, $f_0(x) = 1$, thus to the single condition: $\sum_\alpha \lambda_\alpha = 1$
- The case of a diagonal linear drift (15) would call for two unbiasedness conditions:

$$\sum_\alpha \lambda_\alpha = 1 \ , \ \text{and :} \ \sum_\alpha \lambda_\alpha (x_\alpha + y_\alpha) = x_0 + y_0$$

We wish now to minimize the error variance under the previous $(L+1)$ linear constraints (17) on the weights λ_α. The Lagrange formalism, used to establish the OK system under one single constraint, is repeated similarly except for $(L+1)$ Lagrange parameters $\mu_0, \mu_\ell, \ell = 1, \cdots, L$.

The resulting system of $(n + L + 1)$ equations with $(n + L + 1)$ unknowns, the n weights λ_α and the $(L + 1)$

Lagrange parameter μ_ℓ, is but a high order constrained "normal"system, belatedly renamed "universal kriging" (UK) system:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta C(x_\alpha - x_\beta) + \sum_{\ell=0}^{L} \mu_\ell f_\ell(x_\alpha) = C(x_\alpha - x_0) \ , \\ \alpha = 1, \cdots, n \\ \\ \sum_{\beta=1}^{n} \lambda_\beta f_\ell(x_\beta) = f_\ell(x_0) \ , \ \ell = 0, \cdots, L \end{cases} \tag{18}$$

The corresponding minimized error variance, or "kriging" variance, is written:

$$\begin{aligned} \sigma^2_{UK} &= E\{[Z(x_0) - Z^*(x_0)]^2\} \\ &= C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha C(x_\alpha - x_0) - \sum_{\ell=0}^{L} \mu_\ell f_\ell(x_\alpha) \end{aligned} \tag{19}$$

Remarks. The adjective "universal" stems from the $(L+1)$ unbiasedness conditions (17), also called "universality conditions" for they ensure unbiasedness whatever the unknown trend parameters a_ℓ's. This adjective "universal" is unfortunate, for there is nothing universal nor intrinsic to either the conditions (17) or the constrained system (18): they depend on the arbitrary choice of the trend component functions $f_\ell(x)$'s.

A better, although more cumbersome name, for the system (18) and the corresponding algorithm would be "kriging under trend constraints."

The system (18) reduces to the OK system (4) in the case of a constant but known mean, $L = 0$ and $m(x) = m$.

The covariance $C(h)$ used in system (18) is that of the residual data $R(x_\alpha) = Z(x_\alpha) - m(x_\alpha)$. Since the trend values $m(x_\alpha)$ are usually unknown, the residual data are also unknown and their covariance $C(h)$ cannot be inferred directly. The formalism of "intrinsic random functions of order L" has been designed for inference of $C(h)$ directly from the original data $Z(x_\alpha)$, see discussion in the next section of this lesson.

In most practical situations, within a site or formation there exists subzones or privileged directions within or along which the trend effect $m(x+h) - m(x)$ can be ignored, at least for short interdistances h. In such subzones and/or directions, the R-covariance $C(h)$ can be inferred directly from the Z-data.

The system (18) presents one and only one solution if:

(i) the data covariance matrix $\underline{K} = [C(x_\alpha - x_\beta)]$ is positive definite, which requires that the covariance model $C(h)$ be itself positive definite.

(ii) the $(L+1)$ functions $f_\ell(x)$ are linearly independent on the set of n data.

This last condition simply says that, e.g., estimation of a linear drift in, say, the x direction cannot be done if all n data are aligned in the y-direction.

The intrinsic random functions of order L

Consider the random function $Z(x)$, with covariance function $C(h) = \text{Cov}\ \{Z(x), Z(x+h)\}$ and the linear estimator

$$Z^*(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(x_\alpha)$$

The estimation error Y can be seen as a linear combination of $(n+1)$ RV's:

$$Y = Z(x_0) - Z^*(x_0) = \sum_{\alpha=0}^{n} d_\alpha Z(x_\alpha) \tag{20}$$

with:

$$d_\alpha = \begin{cases} 1 \ , \ \text{for } \alpha = 0 \\ -\lambda_\alpha \ , \ \text{for } \alpha = 1, \cdots, n \end{cases}$$

The error variance, i.e., the variance of the linear combination is given by Lesson I (33):

$$\text{Var}\, Y = \sum_{\alpha=0}^{n} \sum_{\beta=0}^{n} d_\alpha d_\beta C(x_\alpha - x_\beta) \geq 0 \tag{21}$$

A variance being necessarily non-negative, the covariance function $C(h)$ must be such that the expression (21) be positive whatever the choices of the $(n+1)$ weights d_α, positive or negative. This is the well known "positive definite" condition for a function $C(h)$ to be a valid covariance function. Zero variance is obtained only when all $(n+1)$ weights d_α are zero.

Calculation of the variance (21) requires existence and knowledge of the covariance function $C(h)$, which required prior knowledge or estimation of the constant mean: $E\{Z(x)\} = m$, for all x.

There are cases, as in (14), where the expected value $E\{Z(x)\} = m(x)$ is neither constant nor known. It would then be useful to express the error variance Var Y as a function of some "generalized" covariance function $K(h)$ that would not require prior knowledge of $m(x)$.

Consider first the case of a constant but known mean: $E\{Z(x)\} = m$, for all x. Under the OK unbiasedness condition (2):

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \ , \text{ entailing } \sum_{\alpha=0}^{n} d_\alpha = 0 \tag{22}$$

the estimation error Y can be written as a linear combination of $(n+1)$ increments:

$$Y = Z(x_0) - Z^*(x_0) = \sum_{\alpha=0}^{n} d_\alpha [Z(x_\alpha - Z(x_0)] \tag{23}$$

and its variance is expressed as:

$$\text{Var } Y = \sum_{\alpha=0}^{n}\sum_{\beta=0}^{n} d_\alpha d_\beta \text{ Cov}\{Z(x_\alpha - Z(x_0)\,,\; Z(x_\beta - Z(x_0)\}$$

Introducing the variogram Lesson 1 (22) defined as the variance of the increment $Z(x+h) - Z(x)$:

$$\begin{aligned} 2\gamma(h) &= \text{Var}\{Z(x+h) - Z(x)\} \\ &= \text{Var}\{[Z(x+h) - Z(x_0)] - [Z(x) - Z(x_0)]\} \\ &= 2\gamma(x+h-x_0) + 2\gamma(x-x_0) \\ &\quad -2\text{Cov}\{Z(x_0) - Z(x+h)\,,\; Z(x_0) - Z(x)\} \end{aligned}$$

Thus:

$$\begin{aligned} &\text{Cov}\{Z(x_0) - Z(x+h)\,,\; Z(x_0) - Z(x)\} \\ &= \gamma(x+h-x_0) + \gamma(x-x_0) - \gamma(h) \end{aligned} \tag{24}$$

The error variance (23) is then rewritten:

$$\begin{aligned} &\text{Var } Y \\ &= \textstyle\sum_{\alpha=0}^{n}\sum_{\beta=0}^{n} d_\alpha d_\beta[\gamma(x_\alpha - x_0) + \gamma(x_\beta - x_0) - \gamma(x_\alpha - x_\beta)] \\ &= [\textstyle\sum_{\alpha=0}^{n} d_\alpha \gamma(x_\alpha - x_0)] \cdot \left[\sum_{\beta=0}^{n} d_\beta\right] \\ &+ [\textstyle\sum_{\alpha=0}^{n} d_\alpha \gamma(x_\alpha - x_0)] \cdot \left[\sum_{\beta=0}^{n} d_\beta\right] \\ &- \textstyle\sum_{\beta=0}^{n} d_\alpha d_\beta \gamma(x_\alpha - x_\beta) \end{aligned}$$

Accounting for relation (22), it comes finally:

$$\begin{aligned} \text{Var } Y = \; & -\textstyle\sum_{\alpha=0}^{n}\sum_{\beta=0}^{n} d_\alpha d_\beta \gamma(x_\alpha - x_\beta) > 0 \\ & \text{under the condition } \textstyle\sum_{\alpha=0}^{n} d_\alpha = 0 \end{aligned} \tag{25}$$

Note the similarity of the two error variance expressions (21) and (25). The negative of the semi-variogram, $-\gamma(h)$, is said to be a "generalized covariance" of order zero. The order zero comes from the order of a polynomial trend $m(x) = P_0(x) = m$ that can be filtered by the error expression (20). Indeed, if a constant m (polynomial trend of order 0) is added to the random function model $Z(x)$, i.e., to the $(n+1)$ values $Z(x_0)$, $Z(x_\alpha)$ in expression (20), it would be cancelled out thanks to the condition $\sum_{\alpha=0}^{n} d_\alpha = 0$.

Any linear combination $Y = \sum_{\alpha=0}^{n} d_\alpha Z(x_\alpha)$, such that $\sum_{\alpha=0}^{n} d_\alpha = 0$, is said to be a "generalized increment" of order 0 and its variance is given by expression (25). In particular, the elementary increment $Z(x+h) - Z(x)$ is an increment of order 0 and acccording to (25):

$$\begin{aligned} &\text{Var}\{Z(x+h) - Z(x)\} = \\ &+ \gamma(x+h-x) + \gamma(x-x-h) - 2\gamma(0) = 2\gamma(h) \\ &\text{since } \gamma(0) = 0 \end{aligned}$$

Generalization. Any linear combination $Y = \sum_{\alpha=0}^{n} d_\alpha Z(x_\alpha)$, such that the weights d_α verify the $(L+1)$ unbiasedness conditions (17), is said to be a "generalized increment" of order L and its variance is given as a double sum of a "generalized covariance" of order L:

$$\text{Var } Y = \sum_{\alpha=0}^{n}\sum_{\beta=0}^{n} d_\alpha d_\beta K_L(x_\alpha - x_\beta) > 0 \tag{26}$$

under the $(L+1)$ conditions (17):

$$\sum_{\alpha=0}^{n} d_\alpha f_\ell(x_\alpha) = 0\,,\; \ell = 0, \cdots, L$$

A generalized increment of order L, $Y = \sum_{\alpha=0}^{n} d_\alpha Z(x_\alpha)$, would filter any polynomial of order L and more generally any function $f(x) = \sum_{\ell=0}^{L} a_\ell f_\ell(x)$ being added to the random function $Z(x)$. Indeed:

$$\sum_{\alpha=0}^{n} d_\alpha \left[Z(x_\alpha) + \sum_{\ell=0}^{L} a_\ell f_\ell(x_\alpha)\right]$$

$= \sum_\alpha d_\alpha Z(x_\alpha) + \sum_{\ell=0}^{L} a_\ell \sum_{\alpha=0}^{n} d_\alpha f_\ell(x_\alpha) = \sum_\alpha d_\alpha Z(x_\alpha)$, whatever the coefficients a_ℓ.

An intrinsic random function of order L (IRF-L) is a class of random functions defined up to an arbitrary polynomial of order L, and more generally up to an arbitrary trend of type $f(x) = \sum_{\ell=0}^{L} a_\ell f_\ell(x)$. The variance of such random functions is then characterized from the "intrinsic" covariance $K_L(h)$.

The variance (26) can be minimized under the $(L+1)$ constraints yielding a system of equations identical to the constrained system (18), except for the residual covariance $C(h)$ being replaced by the generalized covariance $K_L(h)$.

Up to this point, notwithstanding its formal elegance, the formalism of intrinsic random function of order L (IRF-L) has not brought any new insight into the very classical and simpler algorithm of constrained minimization of an error variance.

Inference. Proponents of the IRF-L approach argue that the generalized covariance $K_L(h)$ is easier to infer than the residual covariance $C(h)$.

Inference of the residual covariance $C(h)$ requires the choice of subzones and/or directions where the trend $\sum_{\ell=0}^{L} a_\ell f_\ell(x)$ can be ignored; $C(h)$ is then inferred directly from the original Z-data.

Inference of the generalized covariance $K_L(h)$ requires also a preliminary definition of generalized increments of order, i.e., linear combinations $Y = \sum_{\alpha=0}^{n} d_\alpha Z(x_\alpha)$ whose variances would be insensitive to any trend of type $\sum_{\ell=0}^{L} a_\ell f_\ell(x)$ and expressed as function of $K_L(h)$ through relation (26).

Assume that N such generalized increments $Y_j, j = 1, \cdots, N$ could be defined.

The generalized covariance $K_L(h)$ is arbitrarily assumed to be of a rather restrictive and isotropic form, such as a parametric polynomial form, e.g., for $L = 1$ in $\Re^2$:

$$K_1(h) = b_0(1 - |h|^{0^+}) - b_1|h| + b_2|h|^3 \qquad (27)$$

with: $b_0, b_1, b_2 \geq 0$.

The covariance parameters b_0, b_1, b_2 are identified by a regression-type procedure that seeks to identify the theoretical variances Var $\{Y_j\}$, calculated from the model (27) and relation (26), to the experimental values y_j^2:

$$\sum_{j=1}^{N} \left[y_j^2 - \text{Var}\{Y_j\}\right] \text{ is minimum} \qquad (28)$$

An analog of such procedure in the case of $L = 0$, and $b_2 = 0$, would be a least-squares fit of a linear variogram model to the experimental average squared increments $\frac{1}{N}\sum_{j=1}^{N} y_j^2$ with $y_j = [z(x_j + h) - z(x_j)]^2$. Well established practice has shown that such blind least-squares fit of a parametric variogram model is not recommendable.

Besides the severe restriction to narrow families of parametric polynomial-type covariance models, the regression-type procedure (28) used to identify the parameters b's has little resolution, is highly outlier-data sensitive and allows little experimentor interaction. All practitioners of traditional variogram modeling know the paramount importance of "variogram cleaning."

Moreover, it is this author's understanding that automatic determination of the parameters b's often yields generalized covariance models close to a pure nugget effect, i.e., a model where the first term $b_0(1 - |h|^{0^+})$ is predominant. In which case, the corresponding kriging yields but a traditional least-squares fit of the specified trend, assuming little or no spatial correlation; indeed, a small yield for such an elaborated formalism.

In conclusion, this author does not recommend usage of the IRF-L formalism and the related canned packages. If programs like Bluepack are used wisely, i.e., setting aside all the automatic parameter determination aspects, they do not differ from any other constrained kriging problem, i.e., from a linear system solver.

Selected Bibliography

Bras, R. and Iturbe, I., 1984, Random functions in hydrology, Addison-Wesley, 559 p.

Delfiner, P., 1976, "Linear estimation of non-stationary spatial phenomena," in Advanced Geostatistics in the Mining Industry, ed. Guarascio, et al., publ. Reidel, Holland, pp. 49-68.

De Marsily, G., 1986, Quantitative Hydrogeology, Academic Press, 440 p.

Englund, E. and Sparks, A., 1988, Geo-Eas User's Guide, (Geostatistical Environmental Assessment Software), ed. US EPA, EMSL, Las Vegas, NV 89193, 155 p.

Froidevaux, R., 1988, Geostat-Toolbox Primer, (Geostatistical 3 dimensional software), ed. FSS International, 10 Ch. de Drize, Troinex, Switzerland, 88 p.

Goldberger, A., 1962, "Best linear unbiased prediction in the generalized linear regression model," in JASA, v. 57, pp. 369-375.

Journel, A. and Huijbregts, C., 1978, Mining Geostatistics, Chap. 5, Academic Press, 600 p.

Kitanidis, P., 1986, Parameter uncertainty in estimation of spatial functions: Bayesian analysis, in Water Resources Research, 22(4), 499-507.

Luenberger, D., 1969, Optimization by vector space methods, Wiley & Sons, 326 p.

Matheron, G., 1973, "The intrinsic random functions and their applications," in Adv. Appl. Prob. 5, pp. 438-468.

Whittle, P., 1963, Prediction and Regulation, U. of Minnesota Press, revised ed. (1983), 187 p.

Lesson IV: Non-Parametric Assessment of Local Uncertainty

Uncertainty about any paritcular unknown value is modeled by a probability distribution os that unknown value conditional to available related information. The essence of that model lies in the assumed relations between information and unknown and between the various elements of information available. These conditional distributions are not necessarily related to any particular prior multivariate distribution model, such as Gaussian. Their determination is done prior and independently of the estimate(s) retained, and accounts for the data configuration, data values and data quality. It allows for usage of soft information either of local or global (structural) nature.

Prior to the choice of a goal oriented optimal estimate(s), availability of conditional distributions for the unknown allows mapping of probabilities of exceedence, risks of misclassification and the assessment of need for additional sampling.

Introduction

Consider the general problem of spatial interpolation and uncertainty assessment. The "local" information consists in a data set distributed in space, with not all data of the same quality or relevant to the same attribute. The "global" information consists in usually soft qualitative, appreciation of the patterns of spatial dependence between local data arising from structural, geological in-

terpretation and/or experience from similar and better developed fields.

Deterministic interpolation techniques, including triangulation and inverse distance-weighting, do not consider the possibility of a distribution of potential values for the unknown, they do not provide any measure of the reliability of the estimates. The main advantage of probabilistic interpolation (prediction) techniques, e.g. ordinary kriging as described in Lesson III, is that an error variance is attached to each estimate. Unfortunately, unless a parametric distribution of spatial errors is assumed, an error variance falls short of providing confidence intervals and the error probability distribution required for risk assessment. Most often symmetric, Gaussian-related, distribution models are assumed for the errors; these models are fully characterized by only two parameters, the error mean and the error variance. Such congenial models are perfectly reasonable for the distribution of, say, measurement errors in the highly controlled environment of a laboratory. However, they are questionable when used for spatial interpolation errors, the kind considered in this text.

It is argued that uncertainty about an unknown is intrinsic to the level of information available and to a prior model for the relations between that information and the unknown. Thus assessment of uncertainty should not be done around a particular estimate, if only because many optimality criteria can be defined resulting in different, yet all "optimal", estimates. Non-parametric geostatistics put as priority, not the derivation of a "optimal" estimator, but the modeling of the uncertainty. That uncertainty model takes the form of a probability distribution of the unknown rather than that of an estimation error.

The modeling of the probability that the unknown at any unsampled location is greater than a given threshold value capitalizes on the proportion of neighboring data valued above that same threshold. That modeling accounts for the proximity of each datum to the unsampled location, the particular threshold considered and the quality of the data being considered.

In this Lesson, the concepts of non-parametric Geostatistics are presented with a minimum of probabilistic formalism. For example, stationarity is introduced as the decision to proceed with averaging and statistical inference over a pre-determined population or area in space; when it comes to implementation, the hypothesis of stationarity boils down to that, no more, no less.

Probabilistic assessment of uncertainty

Consider an attribute value z to be measured. Any particular measurement would provide an estimated value z_j likely to be in error. Hence and whenever possible, several measurements are performed yielding several estimates, $z_j,\ j = 1, \cdots, n$. Provided that the measurement device and its accuracy have remained constant over the sequence $j = 1, \cdots, n$, (which implies in particular that there is no potential for trend in the n outcomes z_j), the distribution of these n outcomes z_j can be used to model the uncertainty about the model z.

For example it may be said that "the probability that the unknown z is lesser or equal to a threshold z_0" is modeled by the corresponding proportion of outcomes $z_j \leq z_0$. By so saying the unknown value z has been elevated to the status of a random variable (RV) Z, the cumulative distribution function (cdf) of which is modeled by:

$$Prob\{Z \leq z_0|(n)\} = proportion\ of\ z_j \leq z_0,\ j = 1, \cdots, n \tag{1}$$

Introducing the indicator function (of the threshold value z_0):

$$i(z_0; z_j) = \begin{cases} 0, & \text{if } z_0 < z_j \\ 1, & \text{if } z_0 \geq z_j \end{cases} \tag{2}$$

The previous model for uncertainty is written:

$$F(z_0|(n)) = Prob\{Z \leq z_0|(n)\} = \frac{1}{n}\sum_{j=1}^{n} i(z_0; z_j) \in [0, 1] \tag{3}$$

Again the notation $F(z_0|(n))$ serves to recall that this probability model is a function of both the threshold value z_0 and the information set (n) constituted by the n outcome values $z_j, j = 1, \cdots, n$.

The model (3) corresponds to an equal weighted average (by $\frac{1}{n}$) of the indicator data $i(z_0; z_j)$. Under the previous hypothesis that the measurement accuracy has remained stable over the sequence $j = 1, \cdots, n$, there is indeed no reason to over/under weight any particular outcome. Otherwise one may consider an alternative model corresponding to an unequal weighting scheme:

$$F(z_0|(n)) = Prob\{Z \leq z_0|(n)\} = \sum_{j=1}^{n} a_j\ i(z_0; z_j) \in [0, 1] \tag{4}$$

with $a_j \geq 0$, for all j, and $\sum_{j=1}^{n} a_j = 1$.

Remarks. Expressions (3) and (4) are different models of the uncertainty about the unknown value z. They are not different estimates of some elusive "true" probability distribution. In particular, one cannot say that model (4) is better than model (3), before having defined what a "good" model should feature. Also such definition would be needed to determine the set of weights $\{a_j,\ j = 1, \cdots, n\}$.

The conditions, $a_j \geq 0$ and $\sum_j a_j = 1$, ensure that the function $F(z_0|(n))$ is an acceptable cumulative distribution function (cdf), i.e. a non-decreasing function valued between [0,1].

Probability intervals. Availability of cdf's as models of uncertainty allows the derivation of probability intervals:

$$Prob\{Z \in]a,b]|(n)\} = F(b|(n)) - F(a|(n)) \quad (5)$$

Again such a probability interval is but a model figuring the uncertainty around the unknown z. Questions such as "How reliable is this probability interval?" amounts to asking "How reliable is the model $F(z_0|(n))$?", and cannot be answered unless a model for the distribution of cdf models is built. Statisticians do have such second level models, but most often they do not bother qualifying the uncertainty attached to models of uncertainty.

Note that these probability intervals can be established prior to the choice of any particular estimate for the unknown value z.

Estimates for z. Beyond the assessment of uncertainty, a unique estimated value for z may be required, say, for decision-making or engineering design purposes. If there is no reason to over/under weight any of the n outcomes, a "reasonable" estimate is the equal-weighted arithmetic average:

$$\hat{z}^{(1)} = \frac{1}{n}\sum_{j=1}^{n} z_j \quad (6)$$

An equally "reasonable" estimate is the median value of the n outcomes, i.e. the value $\hat{z}^{(2)} = z_{.5}$ that would leave approximately one half of the outcome values below it and one half above it:

$$\hat{z}^{(2)} = F^{-1}(0.5|(n)), \text{ such that: } F(\hat{z}^{(2)}|(n)) \approx .5 \quad (7)$$

If the n outcomes are not equally accurate, they should be weighted differently. Using, for example, the weights a_j retained for the cdf model (4), the two following "reasonable" estimates for z are derived:

$$z^{*(1)} = \sum_{j=1}^{n} a_j z_j \quad (8)$$

$$\text{and } z^{*(2)} = F^{-1}(0.5|(n)). \quad (9)$$

However, other estimates can be derived independently of the cdf model. Examples are:

$$z^{*(3)} = max\{z_j,\ j = 1,\cdots,n\}, \text{ for conservative reasons} \quad (10)$$

$$z^{*(4)} = \frac{max(z_j) + min(z_j)}{2} \quad (11)$$

$$z^{*(5)} = \hat{z}^{(1)} \text{ equal-weighted arithmetic average} \quad (12)$$

$$z^{*(6)} = \frac{1}{n-2}\sum_{j=1}^{n-2} z_j \text{ arithmetic average eliminating the lowest and highestobserved outcomes} \quad (13)$$

Remarks. All previous estimates can be considered as "reasonable", although they can be quite different one from another. There is thus a need to go beyond the adjective "reasonable" and define precisely worded criteria for retaining a single value for estimate of the unknown value z.

Because there is no unique "best in all cases", estimate for the unknown z, assessments of uncertainty such as probability intervals of type 5 should not depend on the particular estimate chosen but solely on the information available (n). In other words, the uncertainty linked to an unknown value z is a function of the information available, not the estimate retained.

Indeed, consider n analyses z_k of a particularly toxic (lethal) substance; for cleaning decisions, the maximum value estimate $z^{*(3)}$ may be retained. Such an estimate however reasonable will be outside most probability intervals of type 5 based on the same information set (n).

A corollary of the previous remark is that traditional 95% probability intervals, leaving 2.5% probability below and above respectively, need not be centered or even contain the estimated value retained.

We have thus established the need for:

- defining a model for the uncertainty about the unknown value, given the available information (n).
- defining criteria for retaining a unique value for estimate of the unknown.

These two tasks need not be related, nor should they call for any Gaussian assumption.

But first we need to broaden our discussion on probabilistic assessment of uncertainty to the case of non-repetitive measurements at different locations of a given space.

The case of spatial distributions.

In the previous section, the case of repeated measurements, z_j, $j = 1,\cdots,n$, of a unique attribute value z has been considered. The cdf model $F(z_0|(n))$, cf, relations (3) or (4), provided an assessment of the measurement uncertainty.

We will consider now the uncertainty linked to spatial interpolation when an attribute value $z(x)$ at location x is to be estimated from measurements $z(x_j)$, $j = 1,\cdots,n$ made at different locations $x_j \neq x$.

For the sake of simplicity we will consider first that the data $z(x_j)$ are error free. This limitation is removed in the later section "Soft Kriging".

The previous space of measurement variability is now replaced by the physical space of locations x, but otherwise the approach used is the same.

Define "spatial independence " as the state whereby the attribute value $z(x)$, at any location x, is not influenced in any way by the attribute values $z(x_j)$ at other locations $x_j \neq x$, no matter how close they are. If the

(n+1) values $z(x_j)$ and $z(x)$ relate to the same attribute Z, but otherwise could be considered as independent one from another, a possible model for the uncertainty about the unknown value $z(x)$ given the n data $z(x_j)$ is the distribution of these data. More precisely it is said that the probability of the unknown $z(x)$ be lesser or equal to a given threshold z is the corresponding proportion of such data values $z(x_j)$. Again by so saying the unknown value $z(x)$ has been elevated to the status of a RV. $Z(x)$, the cumulative distribution of which is modeled by:

$$Prob\{Z(x) \leq z|(n)\} = \begin{array}{l}\text{proportion of data} \\ z(x_j) \leq z,\ j = 1, \cdots, n\end{array} \quad (14)$$

or introducing the indicator variables, one for each datum location x_j and each threshold value z:

$$i(z; x_j) = \begin{cases} 0, & if\ z < z(x_j) \\ 1, & if\ z \geq z(x_j) \end{cases} \quad (15)$$

the previous relation is written similar to relation (3).

$$F(z; x|(n)) = Prob\{Z(x) \leq z|(n)\} = \frac{1}{n}\sum_{j=1}^{n} i(z; x_j) \in [0, 1] \quad (16)$$

Remark. By pooling all n data $z(x_j)$ into the same cdf (16), we are assuming that, although independent from one another, they still have something in common: the fact that they relate to the same attribute Z. This is the source of the stationarity hypothesis, which states that the (n+1) random variables, $Z(x)$, $Z(x_j)$, $j = 1, \cdots, n$, have the same distribution which can then be estimated by an average of type (16) considering data taken at different locations.

Dependence. In the general case there will be some pattern of dependence between the attribute values $z(x)$, $z(x')$, $z(x'')$ at different locations x, x', x''. Thus the datum value which is the closest or more generally the most related to the unknown value should be weighted more, leading to an unequal weighted average:

$$\begin{array}{rl} F(z; x|(n) = & Prob\{Z(x) \leq z|(n)\} \\ = & \sum_{j=1}^{n} a_j(x)\ i(z; x_j) \in [0, 1] \end{array} \quad (17)$$

with $a_j(x) \geq 0$ for all j, and $\sum_{j=1}^{n} a_j(x) = 1$.

Since the probability value $F(z; x|(n))$ needs to be calculated for each different threshold value z, the weights can be made also dependent on z, defining the new cdf model:

$$F(z; x|(n)) = \sum_{j=1}^{n} a_j(z; x)\ i(z; x_j) \quad (18)$$

Remarks. The weights $a_j(z; x)$ need not be any more non-negative nor sum up to one, as long as expression (18) remains that of a cdf, i.e. provided that the following order relations are verified:

$$\begin{array}{ll} F(z; x|(n)) & \in [0, 1] \\ F(z; x|(n)) & \geq F(z'; x|(n)),\ \text{for all } z \geq z' \end{array} \quad (19)$$

Probability Intervals. The uncertainty about the value $z(x)$ can be assessed through probability intervals derived directly from any of the cdf models (16), (17), and (18):

$$Prob\{Z(x) \in]a, b]|(n) = F(b; x|(n)) - F(a; x|(n)) \quad (20)$$

Of particular importance for many applications is the probability of exceedance of a threshold b:

$$Prob\{Z(x) > b|(n)\} = 1 - F(b; x|(n)) \quad (21)$$

Again note that such probability intervals and probabilities of exceedance can be established prior to the choice of any particular estimate of the unknown value $z(x)$.

Isopleth curves for the probability of exceedance (21) can be contoured (for x varying in space and b fixed). For many decision making processes, these iso-probability maps suffice; thus there would be no need for the additional hypotheses required to retrieve unique-valued estimates of $z(x)$.

Exactitude Requirement. Whatever cdf model $F(z; x|(n))$ is retained to assess the uncertainty about the unknown $z(x)$, it should be consistent with the fact that at data locations there is no uncertainty, i.e. the model should be as such

$$\begin{array}{rl} F(z; x_j|(n)) = & Prob\{Z(x_j) \leq z|(n)\} \\ \equiv i(z; x_j) = & \begin{cases} 0, & if\ z < \text{ datum value } z(x_j) \\ 1, & \text{otherwise} \end{cases} \end{array} \quad (22)$$

for all data locations x_j, $j = 1, \cdots, n$.

Remarks. Relation (22) assumes that the data $z(x_j)$ are exact. If they are not, then the cdf model $F(z; x|(n))$ need only be consistent with what is actually known at location x_j, possibly only a prior distribution for the value $z(x_j)$, see section on Soft Kriging.

Relations (18) and (22) show that the cdf model $F(z; x|(n))$ may be seen as an estimate of the unknown indicator function $i(z; x)$ using weighted averages of the known indicator values $i(z; x_j)$:

$$F(z; x|(n)) = [i(z; x)]^* = \sum_{j=1}^{n} a_j(z; x)\ i(z; x_j) \quad (23)$$

Under a rigorous probability setting one can show that the uncertainty model (18) or (23) is, indeed, a linear estimate of the conditional distribution of $Z(x)$ given the n data $Z(x_j) = z(x_j)$, [Isaaks 1984, p.26] or [Journel 1986 - appendix].

Determination of the cdf uncertainty model

The very reason for considering an unequal weighting of the data, as in relations (4), (17), (18), or (23), is the recognition of the existence of patterns of spatial dependence between the data $z(x_j)$ and the unknown $z(x)$, thus entailing some dependence between the indicator data $i(z;x_j)$ and the unknown indicator function $i(z;x)$. Hence these patterns must be identified and first, the notion of spatial dependence must be defined and measured.

Euclidian Measure. A datum value $z(x_j)$ can be seen as a pair of information, one related to the datum location x_j, the second related to the attribute value. Common measures of proximity between two values $z(x)$ and $z(x+h)$ are linked to the euclidian distance $|h|$, or modulus of the interdistance vector h; examples are:

- Constant - $|h|$
- or the reciprocal of $|h|$ to some power: $\frac{1}{|h|^w}$, with usually $w = 1$ or 2.

Such measures account for only the location information x_j of a datum $z(x_j)$, not for the attribute-information z. Indeed, the Euclidian distance $|h|$ is attribute Z independent; it is the same whether very continuous layer thickness or very erratic transmissivity values are considered. Also the Euclidian distance is the same for two indicator pairs $[i(z;x_j), i(z;x_j+h)]$ and $[i(z';x), i(z';x_j+h)]$, although $z \neq z'$.

Variogram Distance Measure. The classical, although not unique distance measure used in geostatistics is the variogram function $2\gamma(h)$ modelled from the experimental average squared discrepancy between the $n(h)$ data pairs separated by approximately the same vector h, see Lesson I (29):

$$2\gamma_Z(h) \text{ modeled from } \frac{1}{n(h)} \sum_{j=1}^{n(h)} [z(x_j) - z(x_j+h)]^2$$

$$2\gamma_I(z;h) \text{ modeled from } \frac{1}{n(h)} \sum_{j=1}^{n(h)} [i(z;x_j) - i(z;x_j+h)]^2 \tag{24}$$

The average squared discrepancy usually increases with the interdistance $|h|$. However, as opposed to the Euclidian distance $|h|$, the variogram distance $2\gamma(h)$ is attribute specific with in particular: $\gamma_I(z;h) \neq \gamma_I(z';h)$ if $z \neq z'$. The variograms are dependent on both the modulus $|h|$ and the direction of the interdistance vector h. They are said to be anisotropic, if they actually depend on the direction of vector h otherwise they are said to be isotropic. As an example, porosity data present greater spatial dependence (smaller variogram values) in directions of deposition (horizontal and along ancient channels).

The corresponding measure of proximity or dependence is the covariance:

$$\begin{aligned} C_Z(h) &= Constant - \gamma_Z(h) \\ C_I(z;h) &= Constant - \gamma_I(z;h) \end{aligned} \tag{25}$$

The arbitrary constant is filtered out from all subsequent utilizations of these proximity measures.

The Indicator Kriging Algorithm. Once a proximity measure is available for the attribute being estimated, a straightforward generalized linear regression algorithm also called "ordinary kriging" can be applied to determine the weighting system, Lesson III (2). For example interpreting relation (23) as an estimation of the unknown indicator value $i(z;x)$:

$$\begin{aligned} F(z;x|(n)) &= [i(z;x)]^* \\ &= \sum_{j=1}^{n} a_j(z;x)\, i(z;x_j) \end{aligned} \tag{26}$$

The n weights $a_j(z;x)$ are given by a constrained system of normal equations, also called the ordinary kriging system, Lesson III (4).

$$\begin{aligned} &\sum_{j'=1}^{n} a_{j'}(z;\underline{x}) \cdot C_I(z;\underline{x}_{j'} - \underline{x}_j) + \mu(z;\underline{x}) = C_I(z;\underline{x} - \underline{x}_j) \\ &\qquad \text{for all } j = 1, \cdots, n \\ &\sum_{j'=1}^{n} a_{j'}(z;\underline{x}) = 1 \end{aligned} \tag{27}$$

Remarks. System (27) appears as a system of (n+1) linear equations with (n+1) unknowns, the n weights $a_j(z;x)$ and a Lagrange parameter $\mu(z;x)$ associated to the condition that the weights sum up to 1. A sufficient condition for the system (27) to provide one and only one solution is that the proximity model $C_I(z;x)$ be positive definite, i.e. be a covariance function, and that no two data be at the exact same location, i.e. $x_j \neq x_{j'}$ for all $j \neq j'$. There is one such system for each threshold value z and for each unsampled location x. In practice the interval of variability of the attribute Z is discretized by a small number of K threshold values, $z_k, k = 1, \cdots, K$. Thus at each location x, K systems of type (27) need to be solved to provide an assessment of the uncertainty through K discrete cdf values $F(z_k;x|(n))$. Cdf values for intermediary threshold values $z \in]z_k, z_{k+1}]$ can be obtained by interpolation, e.g. linear:

$$\begin{aligned} F(z;\underline{x}|(n)) = \; & F(z_k;\underline{x}|(n)) \\ & + \tfrac{z-z_k}{z_{k+1}-z_k} \cdot [F(z_{k+1};\underline{x}|(n)) - F(z_k;\underline{x}|(n))] \end{aligned} \tag{28}$$

Other non-linear interpolation procedures might also be considered.

Exactitude. The ordinary kriging systems (27) provide weights such that the exactitude requirement (22) is met.

Order Relations. The K systems (27) do not ensure that the resulting K cdf values verify the order relations

(19). In practice a piecewise linear model is fitted to the K values $F(z_k;x|(n))$ so that to guarantee the order relations, and it is this model which is used to characterize the uncertainty about the unknown. For more details on such corrections see [Sullivan 1984 p 36-42].

The Probability Kriging Algorithm. The indicator estimate (23) uses only the indicator part of exceedence rather than the whole information available $z(x_j)$. That estimate is enhanced if all the information available is used, defining the new weighted average:

$$[i(z;x)]^* = \sum_{j=1}^{n} a_j(z;x) \cdot i(z;x_j) + \sum_{j=1}^{n} b_j(z;x) \cdot z(x_j)$$

This new estimate mixes indicator data valued either 0 or 1 and data $z(x_j)$ valued in the possibly widely different measurement unit of attribute Z. This scaling problem is solved by considering instead of the data $z(x_j)$ their rank order transforms $r(x_j)$.

If the n data $z(x_j)$ are ranked in increasing order and $r(x_j) \in [1,n]$ is the rank of the datum $z(x_j)$, the transformed datum $\frac{1}{n}r(x_j)$ is valued between [0,1]. The improved indicator estimate is written and is taken as model for the uncertainty about $z(x)$:

$$\begin{aligned} F(z;x|(n)) = & \ [i(z;x)]^* \\ = & \ \textstyle\sum_{j=1}^{n} a_j(z;x) \cdot i(z;x_j) \\ + & \ \tfrac{1}{n} \textstyle\sum_{j=1}^{n} b_j(z;x) \cdot r(x_j) \end{aligned} \qquad (29)$$

Relation (29) can be read as a cokriging estimate of the indicator value $i(x;z)$ by two sets of data, the n indicator data $i(z;x_j)$ and the n rank order data $r(x_j)$. The corresponding $2n$ weights $a_j(z;x)$ and $b_j(z;x)$ are given by a cokriging system of the type described in Lesson II (21-22), see also [Isaaks, 1984, p. 16-27], [Journel, 1984-a].

It is of paramount importance to understand that expressions (26) or (29) are not used as estimates of the local indicator value $i(z;x)$, although they are built as such. These expressions are used as model values for the conditional probability $F(z;x|(n)) = Prob\{Z(x) \leq z|(n)\}$, [Journel, 1986, Appendix].

The Soft Kriging Generalization

In all the preceding developments, the data available were all considered "hard data", that is without uncertainty, generating indicator data that are valued either 0 or 1.

However in real practice, data are never exactly hard if only because of measurement errors. It would be appropriate to enter actual data either under the form of constraint intervals:

$$z(x_j) \in]a(x_j), b(x_j)] \qquad (30)$$

or under the form of prior probability distributions:

$$Prob\{Z(x_j) \leq z\} = F(z;x_j) \in [0,1] \qquad (31)$$

For a given datum location the indicator information can be viewed as a column of K indicator data, one indicator datum for each of the K threshold values z_k. In the case (30) of a constraint interval, the indicator data column is informed only outside that interval with:

$$\begin{aligned} i(z_k;x_j) = & \ 0 \text{ for all } z_k \leq a(x_j) \\ i(z_k;x_j) = & \ \text{unknown for all } z_k \in]a(x_j), b(x_j)] \\ i(z_k;x_j) = & \ 1, \text{ for all } z_k > b(x_j) \end{aligned} \qquad (32)$$

A "hard" datum is but a particular case of a constraint interval (31) with zero amplitude: $a(x_j) = b(x_j) = z(x_j)$. In the case (32) of local information given as a probability distribution, the indicator datum column is filled-in with values between [0,1] instead of equal to either 0 or 1.

$$i(z_k;x_j) = F(z_k;x_j) \in [0,1] \qquad (33)$$

The indicator algorithm (26) or (29) thus allows pooling and joint weighting of data of several types,

- hard data.
- inequality constraints of type (30) with possibly
  1. $a(x_j) = -\infty$ corresponding to $z(x_j) \leq b(x_j)$
  2. $b(x_j) = +\infty$ corresponding to $z(x_j) > a(x_j)$
- soft information presented under the format of a prior distribution of type (31).

At a given location x_j where no core sample is available, well log data or simply geological interpretation may indicate that a particular rock type prevails. At that location x_j, no hard data is available but one may consider the soft information constituted by, say, the distribution (31) of all core porosity data available within the indicated rock type.

The result of indicator kriging is a probability column $F(z_k;x|(n))$ valued between [0,1] and used as a model for the uncertainty about the unknown value $z(x)$.

Exactitude. If at a location x, there exists prior soft information of type (33), the exactitude property of kriging entails that this prior information is restituted unchanged:

$$F(z_k;x|(n)) \equiv F(z_k;x), \text{ for all } z_k, \qquad (34)$$

whatever is the information at the n other locations. In other words the process of indicator/probability kriging

does not update prior information, it completes it by interpolation.

In most practical cases at every location there exists minimum prior information such as a large constraint interval [0, 100%] for porosity or saturation. Consequently at any location, the initial indicator column is only partially filled with:

$$i(z_k;x) = \text{ 0 for all } z_k \leq 0$$
$$i(z_k;x) = \text{ 1 for all } z_k > 100\%$$

The process of probability/indicator kriging does not update these prior values but completes the indicator column with probability values for those threshold values $z_k \in [0, 100\%]$.

Optimal estimate(s) for the unknown

The probabilistic assessment of the uncertainty about an unknown value is completed with determination of a probability distribution of type (23). As discussed before, knowledge of that distribution suffices to the derivation of type (18).

It is recalled that the determination of a probability distribution for an unknown does not, and should not, require any preliminary choice of a unique estimate for $z(x)$. Similarly, a probability-oriented oriented decision making process should not require more than the assessment of the uncertainty about $z(x)$.

However it is traditional to require the derivation of such unique estimated values $z^*(x)$. Since the unknown value could be any value within the interval [0,100%] for concentrations, there is clearly a need to make precise the criterion for selecting a unique estimated value. Plausible criteria are numerous, each leading to different estimates. Different uses of the estimates may call for different criteria, and there is no such thing as an "optimal for all purposes" estimate or estimation algorithm. As an example, most estimates featuring good local accuracy properties, including the kriging estimate, are smoothed in their spatial variability; thus they may prove inadequate for mapping purposes if the map is to be used to evaluate spatial variability.

An incisive layman-oriented discussion of optimality for decision making can be found in [Srivastava 1987]. This is not the place to survey the vast array of diverse optimality criteria leading to as vast an array of estimates, all optimal. We will limit our discussion to those estimates that can be derived straightforwardly from the uncertainty model $F(z;x|(n))$. These estimates are but a subset of the former vast array of optimal estimates and may not include the one needed for a specific application.

Loss Functions and L-Optimal Estimates.

When estimating an unknown value by a single value $z^*(x) = z^*$, a non-zero error $z^* - z(x)$ is likely to occur. Assume that the impact or loss attached to each level of error can be assessed by a function $L(z^* - z(x))$. The function $L(e)$ is known, e.g. $L(e) = e^2$, but the argument $e = z^* - z(x)$ is unknown. However an uncertainty model $F(z;x|(n))$ which represents the distribution of the unknown $z(x)$ is available. Thus the idea is to use that distribution model to determine the expected loss value:

$$E\{L(z^* - Z)|(n)\} = \int_{-\infty}^{+\infty} L(z^* - z) \cdot dF(z;x|(n)) \quad (35)$$

in practice approximated by the discrete sum:

$$\approx \sum_{k=1}^{K} L(z^* - z_k') \cdot [F(z_{k+1};x|(n)) - F(z_k;x|(n))] = \varphi(z^*;x) \quad (36)$$

with, e.g. $z_k' = \frac{z_{k+1}+z_k}{2}$, if the attribute Z interval of variability has been discretized by K threshold values $z_k, k = 1, \cdots, K$, and usually: $F(z_{K+1};x|(n)) = 1$.

The expected loss (35) calculated from the model $F(z;x|(n))$, appears as a function $\varphi(z^*;x)$ of the particular estimated value z^* retained. The optimal estimate for the loss function L is the value $z_L^*(x)$ minimizing the expected loss $\varphi(z^*;x)$:

$$z_L^*(x) = \text{ value } z^* \text{ minimizing } \varphi(z^*;x) \quad (37)$$

Derivation of the optimum $z_L^*(x)$, for any particular loss function L, can be done through iterative calculations of the discrete sums approximating the Stieltjes integral (35); such calculations do not present any difficulty.

For some particular loss functions, the solution is straightforward,

- if $L(e) = e^2$, i.e. for the least squares criterion, the best estimate z^* is the expected value of the probability distribution $F(z;x|(n))$, also called E-type estimate:

$$z_E^*(x) = \int_{-\infty}^{+\infty} z \, dF(z;x|(n)) \quad (38)$$

see relation (43) for an approximation of this integral.

- if $L(e) = |e|$, i.e. for the mean absolute deviation criterion the best estimate is the median of the distribution $F(z;x|(n))$ defined as:

$$q_{.5}(x) = F^{-1}(.5;x|(n)) \text{ such that } F(q_{.5}(x);x|(n)) = .5 \quad (39)$$

- if

$$L(e) = \begin{cases} w_1 e & for \quad e \geq 0 \quad \text{(overestimation)} \\ w_2 |e| & for \quad e < 0 \quad \text{(underestimation)} \end{cases} \tag{40}$$

i.e. for an asymmetric linear loss function, the best estimate is the p-quantile of the distribution $F(z; x|(n))$, see [Journel 1984 -b]:

$$q_p(x) = F^{-1}(p; x|(n)) \text{ with } p = \frac{w_2}{w_1 + w_2} \in [0, 1] \tag{41}$$

- if

$$L(e) = \begin{cases} 0, \text{ for } e = 0 \\ \text{constant, otherwise} \end{cases} \tag{42}$$

the best estimate is the most plausible outcome of the distribution $F(z; x|(n))$, i.e. the mode of the corresponding density function $f(z; x|(n)) = \partial F(z; \underline{x}|(n))/\partial z$

Remarks. Spatial distributions of toxic chemical concentrations are usually highly skewed and would generate strongly asymmetric pdf models $f(z; x|(n))$ for the uncertainty about an unknown. In such cases the optimal estimates (38) and (39) are quite different, and the impact of using one criterion rather than the other can be dramatic.

It bears repeating that nature is rarely Gaussian, and that uncontrolled spatial distributions of earth/environmental sciences attributes are almost always non-Gaussian, as opposed to distributions of repetitive measurements in a highly controlled, carefully planned, laboratory experiment.

Most decision making processes, if properly analyzed would call for an asymmetric loss function of type (40), whether linear or not. For example, for mud-loading purposes, the impact of pressure underestimation (with blow-out risk) is usually much higher than that of an overestimation (using too much mud). Thus for such decision a standard least-squares type estimate such as (38) would be inappropriate; fortunately decision makers are aware of such shortcomings and apply various sorts of safety factors which amount to considering asymmetric loss functions. The rigorous concept of loss function and optimal L-estimate may allow greater consistency in the decision making process, once a particular and possibly subjective loss function has been chosen.

The E-Type Estimate. The E-type estimate (40), although a least squares type estimate, is usually different from the direct ordinary kriging estimate, $z^*_{krig}(x)$ using only the original data $z(x_j)$ and the corresponding proximity measure model $C_Z(h)$ as defined in relation (25). As opposed to the approach yielding the E-type estimate, the ordinary kriging approach does not qualify the ordinary kriging estimate for uncertainty (probability intervals), unless a Gaussian distribution for interpolation errors is assumed.

The Stieltjes integral (38) defining the E-type estimate is, in practice, approximated by a discrete sum:

$$z^*_E(\underline{x}) \approx \sum_{k=1}^{K} z_k' \cdot [F(z_{k+1}; \underline{x}|(n)) - F(z_k; \underline{x}|(n))] \tag{43}$$

with: $z_k' = \int_{z_k}^{z_{k+1}} z \, dF(z; \underline{x}|(n)) \in]z_k, z_{k+1}]$ being the k^{th} class mean

and z_k, $k = 1, \cdots, K+1$ being (K+1) class bounds discretizing the Z attribute interval of variability.

The indicator/probability kriging process provides for the probability values $F(z_k; x|(n))$ but not for the class means z_k'. Estimation of these class means, except for the last one, usually does not pose any problem. For example a class mean can be estimated by the equal weighted average of all data falling into the corresponding class. Whenever the class mean z_k' relates to a narrow class amplitude $z_{k+1} - z_k$, there is no need to consider sophisticated unequal-weighted averages.

However, for heavily skewed distributions the last class mean may have a disproportionate influence on the estimate (43). This last class mean is highly dependent on the largest datum observed and thus on the decision to keep it, cull it, or set it back to a lower value . A conservative decision consists of using the class median for the estimate. Another solution calls for a parametric model of the distribution within that class: for example a 2 parameter lognormal model can be fitted to two robust statistics of the highest data $z(x_j) \in]z_K, z_{K+1}]$, such as z_K being the $F(z_K; x|(n))$ - quantile, and the median of these high data being the $[1 + F(z_K; x|(n))]/2$- quantile, then the last class mean of that lognormal model is taken for the estimate of z_K'.

The poor outlier resistance of the E-type estimate (43) is shared by all mean type estimates, including the direct ordinary kriging estimate. At least in the case of an E-type estimate , that outlier-sensitivity is confined to the last class. Quantile type estimates such as (39) and (41) have much better outlier resistance and should be considered whenever their underlying optimality criteria are acceptable.

Risk-qualified maps

Availability of cdf models $F(z; x|(n))$, one such function for each location x within a given area A, allows contouring of isopleth curves of:

- the optimal estimate(s) $z^*(x)$ retained. These estimates can be derived independently of the uncer-

tainty model $F(z;x|(n))$, see for examples relations (6) to (13).

It is recalled that most estimated values $z^*(x)$ derived from a local accuracy criterion, including all kriged estimates and the E-type estimates (38), tend to present an oversmooth image of the spatial distribution of the actual attribute values. A solution to this problem is proposed in the next Lesson.

- probabilities that the actual unknown value $z(x)$ exceeds any given threshold, such as:

$$Prob\{Z(x) > z_0|(n)\} = 1 - F(z_0;x|(n)) \quad (44)$$

Considering for threshold value, e.g., a high permeability value z_0, a zone where all such probabilities of exceedence are high may indicate a volume of high pay or a flow path. Correspondingly, if a low threshold value z_0 is considered, mapping isopleth values of probabilities of non-exceedence $1 - Prob\{Z(x) \leq z_0|(n)\}$ indicate probable zones of low pay or flow barriers.

- risk of misclassification:

$$\begin{aligned}\alpha(x) &= Prob\{Z(x) \leq z_0|z^*(x) > z_0,(n)\} \\ &= F(z_0;x|(n))\end{aligned} \quad (45)$$

for those locations x such that $z^*(x) > z_0$, or:

$$\begin{aligned}\beta(x) &= Prob\{Z(x) > z_0|z^*(x) \leq z_0,(n)\} \\ &= 1 - F(z_0;x|(n))\end{aligned} \quad (46)$$

for those locations x such that $z^*(x) \leq z_0$. Again, the estimate $z^*(x)$ used for the classification could be any.

- p-quantiles, $q_p(x), x \in A$, p fixed in [0,1], of the distributions $F(z;x|(n))$, i.e. contour maps of the values $q_p(x)$ for which the probability of exceedance is $1-p$. If p is chosen low, e.g. $p = .1$, the probability of exceedence of $q_p(x)$ at location x is high: $1 - p = .9$, thus those areas where the values $q_{.1}(x)$ are high are high z-values areas with a large confidence. Conversely, if p is chosen high, e.g. $p = .9$, the probability of non-exceedence of $q_p(x)$ at location x is high: $p = .9$, thus those areas where $q_{.9}(x)$ are low are low z-values areas with a large confidence. Note that determination of p-quantiles and their mapping do not require the choice of a particular estimator $Z^*(x)$.

It is suggested that a probabilistic assessment of additional sampling need not require a prior map of estimated values $z^*(x)$, i.e. a prior, often arbitrary, decision about an optimality criterion. Additional data are needed in zones that are not only poorly sampled but also critical for the design at hand. For example, if a string of locations (x) with high probability of exceedence of a high permeability threshold, say $Prob\{Z(x) > z_0|(n)\} > .8$, is interrupted by some low such probability values $Prob\{Z(x') > z_0|(n)\} < .5$, the latter locations x' are prime candidates for additional sampling if ascertaining a continuous flow path along the string (x) is critical to the design at hand. Similarly if a string of locations (x) with high probability of non exceedence of a low permeability threshold, $Prob\{Z(x) \leq z_0|(n)\} = .8$, is broken by a few locations (x') with lower such probability, these latter locations (x') are candidates for additional sampling if the continuity of that flow barrier (low permeability string) is critical. Similar assessments can be done from quantile maps $\{q_p(x), x \in A\}$.

Warning: Quantile maps $q_p(x)$ or exceedence maps $Prob\{Z(x) > z_0|(n)\}$, for very large p and threshold values z_0, may depend almost totally on the model extrapolating the last calculated value $F(z_K;x|(n))$ towards the maximum value 1. If $q_p(x)$ is larger than z_K, or p is larger than $F(z_K;x|(n))$, they are not any more actual-data related. This model dependence for extreme values occurence is not particular to the indicator kriging formalism, it is shared (although possibly not stated) by all other algorithms, e.g. when used a standard normal 95% or 99% confidence interval. Non-parametric geostatistics is no replacement for actual data.

Conclusions

The available information whether hard data, constraint intervals, or soft information under the form of prior probability distributions, is coded as a series of indicator columns. There is one such indicator column for each location, whether a datum location or an unsampled location. The indicator/probability kriging process consists in filling-in the missing entries of the indicator columns with probability values. These probability values are obtained by weighting neighboring and related indicator data.

This process yields at each location a probability distribution for the unknown attribute. The process is non-parametric in the sense that it does not assume any particular prior model for these distributions. These probability distributions characterize the uncertainty about the unknowns, independently of the particular estimates retained; they are only information-dependent and are fully so in the sense that they are not only data-configuration

dependent, but also data-values and data-quality dependent.

If a loss function measuring the impact of any given magnitude of the interpolation error is available an optimal estimate minimizing that loss can be derived. Various optimal estimates may thus be considered depending on their particular utilizations. However a single measure of uncertainty applies to all such estimates, for a given information set.

The availability of distributions of the unknowns allows contouring isopleth maps of

- conditional quantile values
- probability of exceedence of any given threshold
- risks α and β of misclassification

Such maps can be used for various decision-making processes, including the assessment for additional data.

Implementation of the indicator/probability kriging technique requires but linear geostatistical tools, i.e. a variogram and an ordinary kriging (normal equations) software. However, as in many statistical applications, the necessary art of approximations requires some experience.

Typical of non-parametric techniques, the elaboration through indicator/probability kriging of the distributions for the unknowns does require an appreciable amount of data. This requirement is in fact a positive aspect, for it indicates whenever data are clearly insufficient for the goal at hand, the alternatives being to acquire more data or rely on a model which is thus not data-based. The general rule is to exhaust the information available through non-parametric tools, and only then revert to parametric modeling.

Selected Bibliography:

Berger, J.O., 1980, Statistical decision theory and Bayesian analysis. ed. Springer-Verlag, 617 p.

Hohn, M.E., 1988, Geostatistics and Petroleum Geology, Van Nostrand, 264 p.

Isaaks, E.H., 1984, Risk qualified mappings for hazardous wastes a case study in non-parametric geostatistics, MSc Thesis, 85p, Branner Earth Sciences Library, Stanford.

Journel, A.G., 1984-a, The place of non-parametric Geostatistics, in "Geostatistics for Natural Resources Characterization", ed. Verly et al., public. Reidel, Dordrecht, Part 1, p. 307-335.

Journel, A.G., 1984-b, mAD and Conditional quantiles estimators, in ibid, Part 1, p. 261-270.

Journel, A.G., 1984-c, Indicator approach to toxic chemical sites, EPA, EMSL-Las Vegas, Report of Project No. CR-811235-02-0

Journel A.G., 1986, Constrained interpolation and soft kriging, in Proceed. of the 19th APCOM Symp. publ. Soc. Min. Eng., Littleton, Co. p15-30.

Luenberger, D.L., 1969, Optimization by vector space methods, ed. Wiley and Sons, NY, 326p.

Lemmer I.C., 1986, Mononodal indicator variography- Part 1-2 Math Geology, Vol. 18, no7, p.589-623.

Srivastava, R.M., 1987, Minimum variance or maximum profitability in CIMM, Vol 80, no.901, pp.63-68.

Sullivan, J.A., 1984, Non-parametric estimation of spatial distributions, PHD Thesis, Branner Earth Sciences Library, Stanford University, 368 p.

Lesson V: Stochastic Simulations for Imaging Spatial Uncertainty

In the previous lessons we strove to model the uncertainty about one unknown at a time, for example, the unsampled porosity value $Z(x)$ at location x. In Lesson IV that uncertainty is modeled by a posterior (conditional) distribution:

$$F(z;x|(n)) = \text{Prob}\{Z(x) \leq z|(n)\} \qquad (1)$$

However, that posterior distribution does not provide information about *joint* spatial uncertainty, e.g. the probability that a string of locations $x_j, j = 1, \cdots, N$ be valued above a certain threshold value z_0 characterizing, say, a fracture or a preferential flow path. Such joint spatial uncertainty would require the joint N-variate posterior distribution:

$$K(z_0; x_j, j = 1, \cdots, N|(n)) = \qquad (2)$$
$$= \text{Prob}\{Z(x_j) > z_0, j = 1, \cdots, N|(n)\}$$

where (n) represents the set of prior (conditioning) data.

Only when the N RV's $Z(x_j)$ can be assumed independent one from another, the joint probability (2) appears as the product of the N probabilities of type (1):

$$K(z_0; x_j, j = 1, \cdots, N|(n)) = \Pi_{j=1}^{n} [1 - F(z_0; x_j|(n))]$$

When N is large, the previous product would yield an artificially small probability of joint exceedence of the threshold z_0. Spatial dependence of the RV's $Z(x_j)$ could make this joint probability much higher, i.e., the possibiliy of a flow path along the string x_j more likely.

Similarly, a map of local estimates $Z^*(x_j)$, whether ordinary kriging estimates (Lesson III, (2)) or E-type estimates (Lesson IV, (40)), may fail to reveal the possibility of such a flow path along the string x_j. Local estimation algorithms are designed to minimize a certain local error impact, whether an error variance in the OK case or a conditional variance in the case of the E-type. They are not designed to reproduce patterns of spatial continuity, less those of extreme values such as $Z(x_j) > z_0$. Kriged maps are notoriously smoothed in their spatial variation, compare Figures 3 and 5 which fail to reflect the proportion of extreme values (both high and low) as seen on the data.

For reservoir characterization where detection of patterns of occurrence of extreme values (permeability-porosity) are more critical than local accurate estimation or global estimation of mean values, we need maps that would honor patterns of spatial continuity and which provide an assessment of joint spatial uncertainty.

Simulation vs. Estimation

One may see the process of reservoir characterization as that of providing the "correct" input to a transfer function representing some aspect of the reservoir engineering, see Figure 1. For example, the transfer function may be a flow simulator requiring for input an unsmoothed spatial distribution of permeability/porosity/saturations, the output being breakthrough times, production rates and sweeping efficiency.

Short of a unique perfect input map of permeabilities which would require exhaustive sampling, the idea is to propose equiprobable alternative maps as input. All such maps should share in common whatever information is available, hard and soft, numerical and interpretative, local and structural. Their differences then provide an image if not a measure of spatial uncertainty about the input. These alternative inputs can then be processed through the flow simulator to yield the corresponding measure of uncertainty on the response function, and answers to such questions as, "What is the probability of an early breakthrough?", see Figure 1.

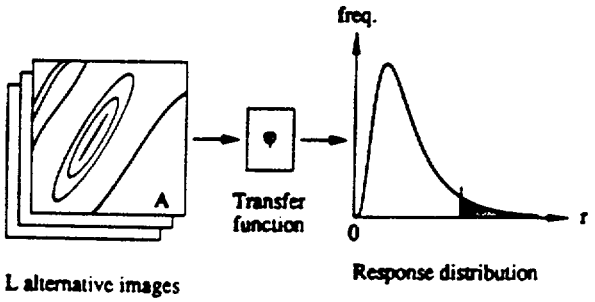


Fig. 1. Processing uncertain spatial information

This approach requires running the transfer function on each alternative input spatial distribution and could be tedious. A shortcut consists in selecting among the alternative input images some deemed diametrically opposed, say, a most favorable image, a median case and a worst case, then run a limited sensitivity analysis on these images. Other shortcuts do exist, most of them specific to the particular transfer function under consideration.

In an estimation approach, the unknown spatial distribution $\{z(x), x \in A\}$ is replaced by a unique set of estimated values $\{z^*(x), x \in A\}$. In a (stochastic) simulation approach, joint spatial uncertainty is characterized by L alternative sets of simulated values $\{z_\ell(x), x \in A\}, \ell = 1, \cdots, L$.

In the first case, one gets a unique response with no measure of its uncertainty:

$$r^* = \varphi(z^*(x), x \in A) \tag{3}$$

Knowing the estimation variance or even the posterior cdf about each of the unknown $z(x)$ does not provide any measure of uncertainty about R or R^*, unless the transfer function φ is linear.

In the second case, one gets a distribution of responses:

$$r_\ell = \varphi(z_\ell(x), x \in A), \quad l, \cdots, L \tag{4}$$

That distribution of responses provide a measure of uncertainty and an assessment of such critical probability of exceedence as Prob $\{R > r_0\}$ is estimated by the proportion of observed responses $r_\ell > r_0$.

Remarks. No matter the unbiasedness or the local accuracy of each estimate $z^*(x)$, the average response $\bar{r}$ need not be close to the response (3) based on the unique estimated map:

$$\bar{r} = \frac{1}{L}\sum_{\ell=1}^{L} r_\ell \neq r^* = \varphi(z^*(x), x \in A) \tag{5}$$

Local accuracy does not entail that the map is any good as a whole. In all rigor, estimates based on a local accuracy criterion such as kriging should only be tabulated; to map them is to entice the user to read on the map patterns of spatial continuity that may be artifacts of the data configuration and the kriging smoothing effect. Where data are abundant the kriged map would appear more variable, where data are sparse the kriged map is essentially flat. The previous remark would apply equally well to most traditional interpolation algorithms, including spline or polynomial fitting and inverse distance weighting.

Each set of simulated values $\{z_\ell(x), x \in A\}$ should reproduce those spatial statistics of the unknown true values $\{z(x), x \in A\}$ which are deemed the most consequential for the transfer function φ. For example, flow in a fractured reservoir is essentially dependent on the pattern of connectivity within A of the high permeability values $(z(x) > z_0)$, not so much on whether at any particular location x the permeability is high or low, nor even on the proportion within A of these high values. In such a case, the $z_\ell(x)$'s should reproduce patterns of connectivity of extreme values even if this means trading off such property as local unbiasedness or reproduction of the overall histogram of the $z(x)$'s over A.

A set of simulated values $\{z_\ell(x), x \in A\}$ is good inasmuch as it is drawn from a model that identifies the relevant spatial characteristics of the unknown true values $\{z(x), x \in A\}$.

The random function $Z(x)$, function of the coordinate(s) x within a domain A, can be seen as a set of spatially dependent random variables $\{Z(x), x \in A\}$, one at each location x. Stochastic simulation amounts to draw joint realizations from all random variables. The essence of any stochastic simulation is the particular spatial dependence model that is imposed on the random function and thus imparted to each of the simulated sets $\{z_\ell(x), x \in S\}, \ell = 1, \cdots, L$.

Spatial connectivity of extreme values between any two locations x and $x + h$ is a highly multivariate property that would require establishing the probability that, along any path linking x and $x + h$, all values $z(u)$ exceed jointly a certain threshold. Notwithstanding the innumerable number of possible paths linking two locations in a 3-dimensional space, such multivariate statistics is never accessible from usually available data. Short of a full multivariate connectivity measure one can define a two-points connectivity function as the joint probability that at two locations x and $x + h$ the attribute does not exceed the threshold values z, z':

$$F(z, z'; h) = Prob\{Z(x) \leq z, Z(x + h) \leq z'\} \quad (6)$$

This two-points connectivity function was defined in Lesson I (21) as the bivariate distribution function characterizing spatial dependence between any two RV's $Z(x)$ and $Z(x + h)$, separated by the same vector h. The bivariate cdf (6) can be seen as an non-centred indicator covariance, function of the separation vector h:

$$F(z, z'; h) = E\{I(x; z) \cdot I(x + h; z')\} \quad (7)$$

with the indicator transform random function defined as in Lesson IV (15):

$$I(z; Z(x)) = I(z; x) = \begin{cases} 0, \text{ if } z < Z(x) \\ 1, \text{ if } z \geq Z(x) \end{cases} \quad (8)$$

Note that the complement indicator transform: $J(z; x) = 1 - I(z; x)$ can be used to define the joint probability of exceedence:

$$\begin{aligned} D(z, z'; h) &= Prob\{Z(x) > z, Z(x + h) > z'\} \\ &= E\{J(x; z) \cdot J(x + h; z')\} \end{aligned} \quad (9)$$

A full characterization of the spatial dependence between two RV's $Z(x), Z(x + h)$ is given by the bivariate cdf (6), thus correspondingly by the set of all indicator covariances for all choices of the pair of parameters z, z'. Such indicator covariances can be inferred from experimental covariances run from the corresponding indicator data. For example, the two-points connectivity function $F(z, z; h)$ can be inferred by the proportion of data pairs, in number $n(h)$, approximately distant of vector h and jointly not exceeding x:

$$\hat{F}_A(z, z; h) = \frac{1}{n(h)} \sum_{\alpha=1}^{n(h)} i(z; x) \cdot i(z; x + h) \quad (10)$$

Alternatively, the function $F(z, z; h)$ can be synthesized from "soft" structural information as provided, e.g. by digitized geologically interpreted and drawn maps based on outcrop sections and/or data from similar depositional environments. Also the model $F(z, z; h)$ can be borrowed from a similar but more mature field where information is more abundant.

Consider for example the indicator covariance (9) defined for a high threshold value $z = z'$: if the two-points connectivity function $D(z, z; h)$ is valued high for all $h \leq h_0$, one can expect a significant probability that two locations x and $x + h_0$ be linked by one or more strings (paths) of low values $(< z)$. However, it bears repeating that $D(z, z; h)$ is not the multivariate connectivity function that one would ideally wish to have.

The two-points measure of connectivity, whether $F(z, z; h)$ or $D(z, z; h)$, need not be isotropic (the same in all directions of h), nor the same for all threshold values z. For example, very high permeability values (open fractures) may be well connected in a given direction and not in others, resulting in a strongly anisotropic covariance $D(z, z; h)$; whereas low permeability values (e.g. non-pay shales) may be more isotropic in their horizontal variability.

The covariance of the original random function $Z(x)$, as defined in Lesson I(32), is but a moment of the bivariate cdf $F(z, z'; h)$:

$$C(h) = E\{Z(x)Z(x + h)\} - [E\{Z(x)\}]^2 \quad (11)$$

Just like the variance of an univariate cdf $F(z)$ does not characterize fully that cdf, the covariance $C(h)$ does not characterize fully the bivariate cdf $F(z,z';h)$ i.e. does not reflect the various connectivity functions $F(z,z;h)$ or $D(z,z;h)$. Consequently, a stochastic simulation $\{z_\ell(x), x \in A\}$ that reproduces only the covariance $C(h)$, or equivalently only the semi-variogram $\gamma(h)$, does not necessarily identify any of the important connectivity functions $F(z,z;h)$ or $D(z,z;h)$. There are cases where identification of one or more critical connectivity functions, i.e. indicator covariances $F(z,z;h)$ or $D(z,z;h)$, is more relevant than reproduction of their indifferentiated average $C(h)$.

The Gaussian model.

The Gaussian RF model is remarkable and particularly congenial in the sense that it is fully determined by the single covariance $C(h)$ of type (11). Conversely, the Gaussian model does not leave any flexibility for the reproduction of more than one indicator covariance model, [Journel and Alabert, 1988].

Consider a Gaussian RF $Y(x)$, with zero mean and correlogram $\rho_Y(h) = C_Y(h)/C_Y(o)$. Any indicator covariance of type (7) or (10) is fully determined from $\rho_Y(h)$, [Xiao 1985], e.g.,

$$\text{Prob}\{Y(x) > y_p\ ,\ Y(x+h) > y_p\} = \qquad (12)$$

$$= (1-p)^2 + \frac{1}{2\pi}\int_0^{\arcsin \rho_Y(h)} \exp\left(-\frac{y_p^2}{1+\sin\theta}\right) d\theta$$

with $y_p = G_0^{-1}(p)$ being the p-quantile of the standard normal cdf, see Lesson I (14). For example, at the median threshold: $y_{.5} = 0$:

$$\text{Prob}\{Y(x) > 0\ ,\ Y(x+h) > 0\} = \frac{1}{4} + \frac{1}{2\pi}\arcsin \rho_Y(h)\ , \qquad (13)$$

[Abramovitz and Stegun 1964, p. 936.]

In practice, the original variable $Z(x)$ may not be normally distributed, thus the Gaussian model would apply to its normal score transform $Y(x) = T(Z(x))$, e.g., $Y(x) = \ln Z(x)$. However, such univariate-type transform does not change in essence the bivariate cdf. Indeed,

$$\text{Prob}\{Z(x) > z_p\ ,\ Z(x+h) > z_{p'}\} \equiv \qquad (14)$$

$$\text{Prob}\{Y(x) > y_p\ ,\ Y(x+h) > y_{p'}\}\ ,\ \text{for all}\ p, p' \in [0,1]\ ,$$

whatever the monotone transform T.

$(z_p, z_{p'})$ and $(y_p, y_{p'})$ are the p, p'-quantiles of the univariate cdf's of respectively $Z(x)$ and $Y(x)$.

Thus, a normal score transform does not introduce any additional degrees of freedom for modeling of specific indicator covariances, i.e., connectivity functions of type (7) or (9).

Maximum entropy.

Not only do Gaussian model-related simulations lack the flexibility to identify actual indicator covariances, they impart to the realizations $\{z_\ell(x), x \in A\}$ a severe maximum entropy character. It can be shown [Jones 1979, p. 137], that among all possible random function models with an imposed covariance $C(h)$ or correlogram $\rho(h)$ the Gaussian model maximizes entropy, i.e., spatial disorder.

Indeed, in expression (12) for extreme threshold values y_p, whether extremely low or high, y_p^2 is large and the integral is close to zero leaving a constant value for the bivariate probability.

$$\text{Prob}\{Y(x) > y_p\ ,\ Y(x+h) > y_p\} \simeq \qquad (15)$$

$$\text{Prob}\{Y(x) > y_p\} \cdot \text{Prob}\{Y(x+h) > y_p\} = (1-p)^2$$

when p tends towards 0 or 1.

Thus, the Gaussian model does not allow for spatial connectivity of extreme occurrences: the two indicators $I(z;x)$ and $I(z;x+h)$ become independent as the threshold value z becomes extreme (departs more from the median).

If high permeability values might connect each other in space generating preferential flow paths, a Gaussian-related model (e.g., a lognormal model) would be inadequate, because it would fail to reveal the possibility of occurrence of such preferential flow paths. Maximum entropy, i.e., minimum ordering of extreme values, does not represent a conservative (i.e., less favorable) model for flow simulations and reservoir management.

The extreme convenience of the Gaussian model is no sufficient justification for adopting a model that may dangerously understate the potential for unfavorable spatial distributions. It behooves the geostatistician to come up with non-Gaussian alternatives allowing patterns of connectivity of high/low values, if only for worst cases analysis.

Simulation features.

Stochastic simulation provides alternative, equiprobable sets of simulated values $\{z_\ell(x), x \in A\}$, $\ell = 1, \cdots, L$ mimicking part or all of the bivariate spatial variability of the unknown true values $\{z_\ell(x), x \in A\}$.

The vast majority of simulation algorithms are Gaussian-related. They yield simulated sets values $\{z_\ell(x), x \in A\}$ which features:

(i) reproduction of an univariate cdf, usually some histogram deemed representative of the domain A

(ii) honoring the data values at data locations, i.e.

$$z_\ell(x_\alpha) = z(x_\alpha) \qquad (16)$$

for all realization ℓ, at any datum location x_α

(iii) reproduction of the covariance $C(h)$ as defined in relation (11)

(iv) a multivariate Gaussian distribution for the normal score transforms, [Journel and Huijbregts 1978], [Mantoglou and Wilson 1981], [Luster 1985], and [Alabert 1987a].

Because of the limitations of the Gaussian model, for some applications we may trade the last two features (iii) and (iv) for:

(iii-a) reproduction of any number K indicator covariances $F(z_k, z_k; h)$ of type (6) or (10)

(iv-a) honoring not only the hard data values $z(x_\alpha)$ but also any number of soft local data, such as constraint intervals $z(x_\alpha) \in]a_\alpha, b_\alpha]$ and/or prior probability distributions for the datum value $z(x_\alpha)$, see Figures 2b and 2c.

Sequential indicator simulation

Consider K threshold values $z_k, k = 1, \cdots, K$ discretizing the range of variability of the attribute $Z(x)$. K can be made as large as needed, however, in practice a small number from 3 to 9 would be sufficient, defining from 4 to 10 classes for $Z(x)$. Indeed, in many applications the exact value of $z(x)$ with 2 or 3 digits of precision is not as important as the fact that it is low, median or high and whether those lows, medians or highs are spatially connected or not.

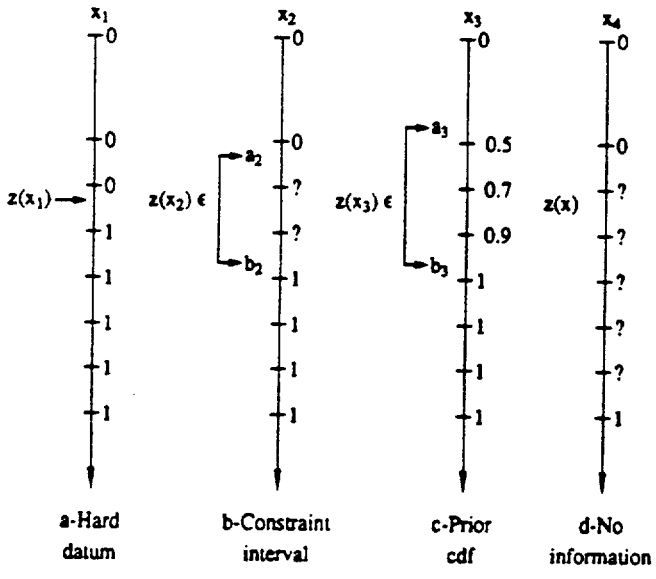


Fig. 2. Indicator coding of prior information

Indicator coding.

Any particular hard (or assumed so) datum $z(x_1)$ can be coded into an indicator data column $\{i(z_k; x_1), k = 1, \cdots, K\}$, with the indicator data defined as in (8), see Figure 2a.

At the location x_2 of a soft datum known only to be in an interval, $z(x_2) \in]a_2, b_2]$, the indicator data column is incomplete: the values $i(z_k; x_2)$ for $z_k \in]a_2, b_2]$ are unknown, see Figure 2b. Such constraint intervals may arise from, say, soft porosity data deduced from well logs or seismic data, the bounds a_α, b_α being deduced from calibration using neighboring hard data, [Thadani, et al. 1987].

At some other locations x_3 the soft datum $z(x_3)$ may be known to lie within an interval $]a_3, b_3]$ with some prior probability distribution. Thus, at that location:

$$\begin{aligned} i(z_k; x_3) &= 0 \text{ , for all } z_k < a_3 \\ &= \text{Prob}\{Z(x_3) \le z_k\} \text{ , for all } z_k \in]a_3, b_3] \\ &= 1 \text{ , for all } z_k \ge b_3 \end{aligned} \qquad (17)$$

Figure 2c and [Journel 1986].

At any unsampled location, where no soft information pre-exists, the indicator data column is void, see Figure 2d.

The idea of indicator/probability kriging (IK/PK) is to complete all indicator columns with a probability valued between [0, 1], see Lesson IV (26, 29).

The idea of sequential indicator simulation is:

(i) to complete a first indicator column for all K threshold values z_k, say, at location x by IK or PK.

(ii) to draw from that indicator column (posterior cdf) a value, say, $z_\ell(x)$

(iii) to move that simulated value $z_\ell(x)$ into the conditioning set, $\cdots$

and repeat the process at another location x', until all required locations are informed by a simulated value.

At each step, the process is conditional to the initial information and all previously simulated values. As the sequential process moves forward, only the nearest conditioning data are retained for the IK/PK algorithm deter-

mining the posterior cdf from which the simulated value is drawn.

Exactitude property.

Since the IK/PK algorithm is class-exact, see Lesson IV (35), the simulated values $z_\ell(x)$ will also be class-exact, in the sense that, e.g., at a location of a prior interval $x(x_2) \in]a_2, b_2]$, all simulated values $z_\ell(x_2), \ell = 1, \cdots L$, will honor that interval up to the class z_k, z_{k+1} resolution. Thus, conditions (ii) and (iv-a) are verified up to the class resolution of the initial indicator coding.

Indicator covariance reproduction.

Consider the simple indicator kriging estimate used to determine a posterior cdf model for the RV $Z(x)$:

$$i^*(z_k, x) - p_k = \sum_{\alpha=1}^{n} \lambda_\alpha [i(z_k; x_\alpha) - p_k] \qquad (18)$$

with: $p_k = F(z_k) = E\{I(z_k; x)\}$ being the mean indicator value to be reproduced.

The corresponding simple kriging (SK) system is written, see Lesson II (12):

$$\sum_{\beta=1}^{n} \lambda_\beta C_k(x_\beta - x_\alpha) = C_k(x - x_\alpha) \qquad (19)$$

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

if the n nearest data locations x_α are retained for IK.

$C_k(h)$ is the centered indicator covariance model for the threshold z_k, i.e., according to relation (6):

$$C_k(h) = F(z_k, z_k; h) - p_k^2 \qquad (20)$$

By construction of the SK estimate (18), the indicator estimate is unbiased:

$$E\{I^*(z_k; x)\} = p_k = F(z_k)\,,$$

thus condition (i) is honored: the model histogram will be reproduced by the simulated values $z_\ell(x)$.

Consider now the covariance of the indicator estimator $I^*(z_k, x)$ with anyone of the n indicator data:

$$E\{[I^*(z_k; x) - p_k][I(z_k; x_\alpha) - p_k(x)]\} =$$

$$\sum_{\beta=1}^{n} \lambda_\beta E\{[I(z_k; x_\beta) - p_k][I(z_k; x_\alpha) - p_k]\} =$$

$$\sum_{\beta=1}^{n} \lambda_\beta C_k(x_\beta - x_\alpha) = C_k(x - x_\alpha)$$

according to (20) for any $\alpha = 1, \cdots, n$.

Thus, the SK indicator estimator $I^*(z_k; x)$ is such that it honors all the indicator covariance model values, $C_k(x - x_\alpha)$. The condition (iii-a) is therefore verified.

Implementation problems.

Implementation of a sequential indicator simulation exercise requires the following five steps:

1. Indicator coding of all local prior information, including possibly soft data. This coding requires determining the minimum resolution acceptable and the K threshold values z_k.

2. Indicator covariance/variogram inference. The K indicator covariances $C_k(h)$ are either modeled from the corresponding experimental covariances or synthesized from geological interpretation and drawings. It is often preferable to use geological interpretation, i.e., a form of "soft" structural information, rather than accept blindly implicit Gaussian-type indicator covariance models with their maximum entropy character.

3. Define a random path through all N nodes to be simulated.

4. At each node x_j of the simulation grid, perform the (up to) K indicator/probability krigings to determine the posterior cdf of the unknown $Z(x_j)$ at that location:

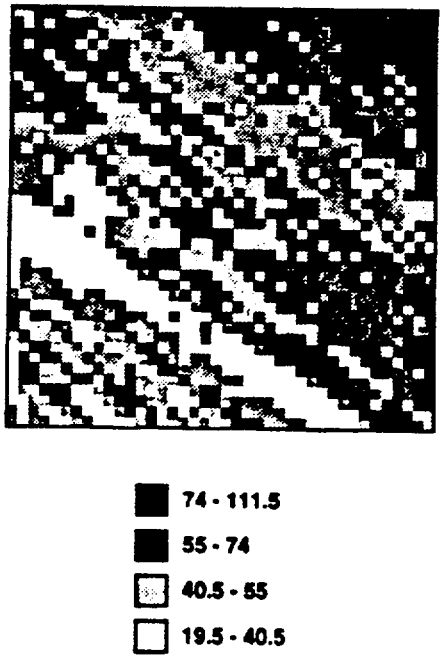
$$i^*(z_k; x_j) = \text{Prob}\{Z(x_j) \leq z_k | \text{Information}\}$$

5. Draw from that posterior cdf the simulated value $z_\ell(x_j)$ and add that value to the list of conditioning data, to impart the necessary correlation structure to the next simulated values.

6. Return to step 3 and loop over all N nodes $x_j, j = 1, \cdots, N$.

Generation of another set of simulated values requires starting the process again from step 3 or 4. Starting again from step 4 amounts to keep the same random path, in which case considerable CPU time can be saved: indeed all indicator data configurations remain the same and thus the kriging weights at each node can be saved for reuse at that same node in other simulations.

The main advantage of the sequential indicator simulation algorithm over other more traditional, and Gaussian-related algorithms, is the possibility to control K spatial covariances instead of a single one. Of course, this advantage is real only if these K indicator covariances are



number of data	1600
mean	55.53
variance	249.
coefficient of variation	0.284
minimum	19.5
maximum	111.5
prob. $[z \leq 40.5]$	0.17
prob. $[40.5 < z \leq 55.]$	0.36
prob. $[55. < z \leq 74.]$	0.34
prob. $[z > 74]$	0.13

Fig. 3. Statistics and greyscale map of the exhaustive data set

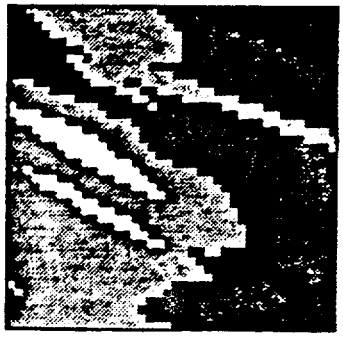


Fig. 4. Greyscale map of 1600 kriged values from 64 data

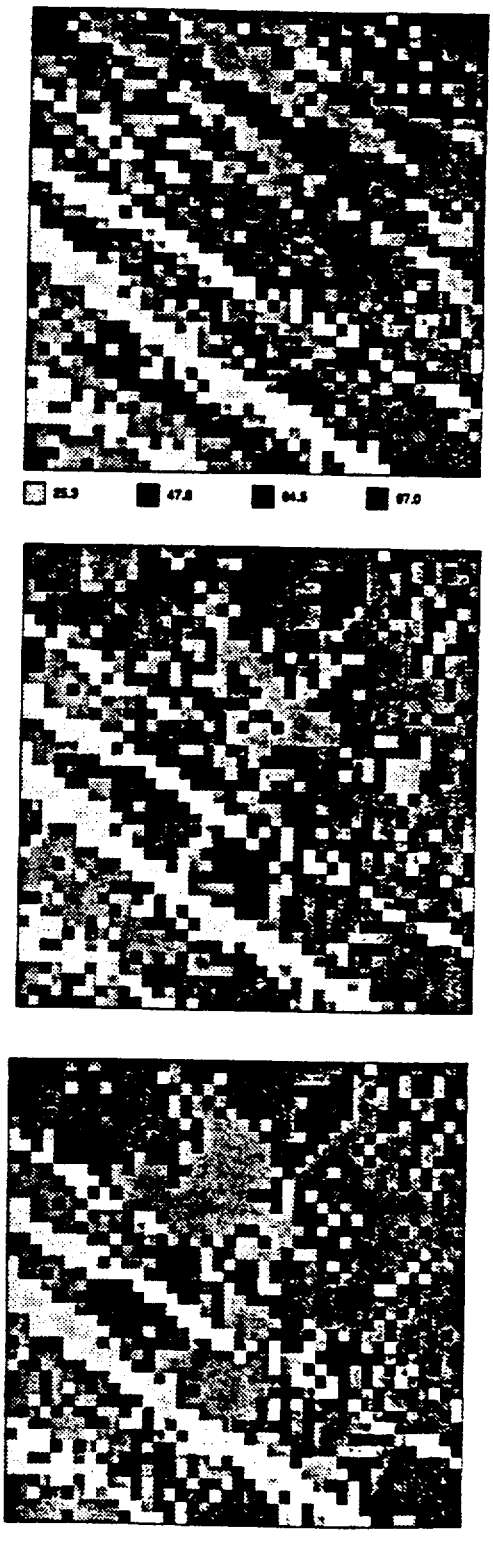


Fig. 5. Greyscale maps of the three sets of simulated values

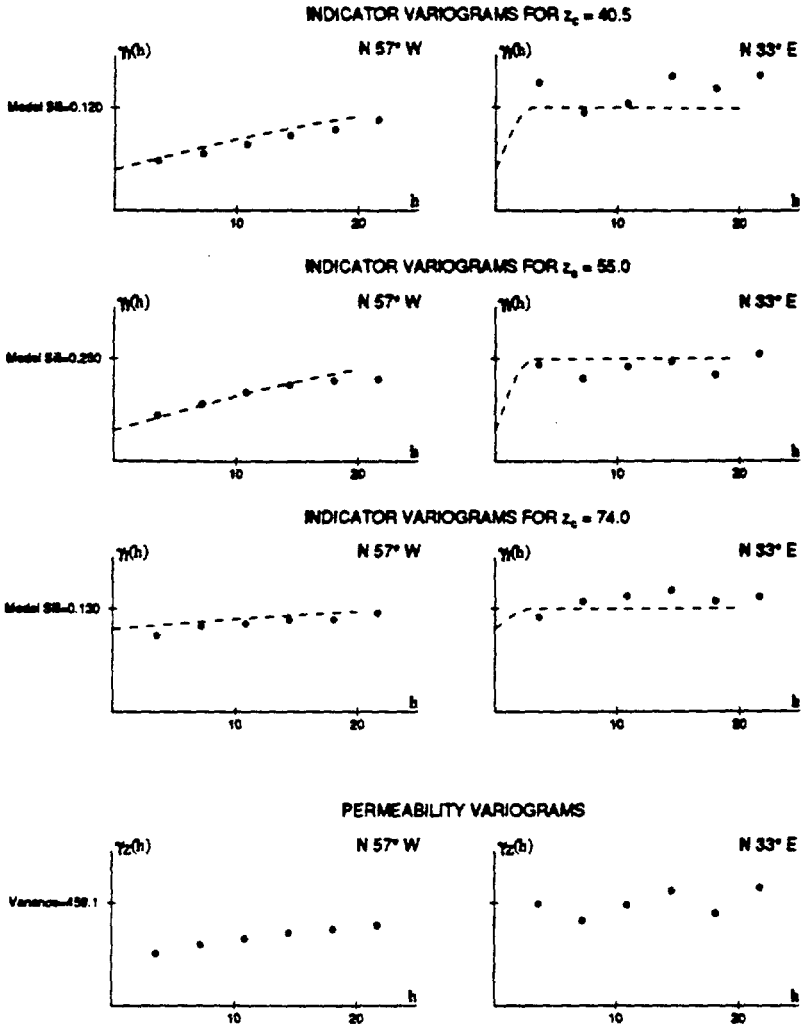


Fig. 6. Indicator variograms of the simulation of figure 5a (dots .) Dash line is the model

indeed representative of the spatial distribution of $Z(x)$ over the domain A.

The Berea case study.

The greyscale map of Figure 3 images the spatial distribution of 40 x 40 air permeameter measurements taken on a vertical slab of Berea sandstone of dimensions two feet x two feet. This data set is described in [Giordano, et al. 1985] and made available by Arco. The corresponding statistics of the 1600 data values are attached to Figure 3. Note the diagonal banding, particularly strong for the low permeability values.

Typically, the image of Figure 3 is not available and must be reconstituted with much less data. Figure 4 gives a greyscale representation of the kriged map (ordinary kriging) using 64 data taken at random locations. Comparing Figures 3 and 4, note the strong smoothing effect induced by the kriging algorithm which is in essence a weighted moving average.

Three sets of simulated values $\{z_\ell(x_j), j = 1, \ldots, 1600\}$, $\ell = 1, \ldots, 3$, using the sequential indicator simulation algorithm are given on Figure 5. Only four classes ($K = 3$) were considered, thus yielding a rather gross resolution. The three threshold values considered are: $z_1 = 40.5$, $z_2 = 55.0$, $z_3 = 74.0$ millidarcies, corresponding to the .14, .5 (median), and .84 quantiles, respectively, of the 64 data distribution.

Figure 6 gives, as an example, the indicator variograms calculated from the simulated set of Figure 6a together with the corresponding models. The variogram models are seen to have been reasonably well-reproduced by the simulated data set.

The banding of the original Figure 3 is well reproduced by all three simulated images of Figure 5. In particular, the main streaks of low values are reproduced.

The variance of the 1600 simulated values of Figure 5a is $\sigma_l^2 = 229$, a value close to the variance $\sigma^2 = 249$ of the original 1600 permeability values, whereas the variance of the 1600 kriged values of Figure 5 is only $\sigma^2 = 175$.

Selected Bibliography

Abramovitz, M. and Stegun, I., 1964, Handbook of mathematical functions, Dover Publ., 9th print, 1046 p.

Alabert, F., 1987a, "The practice of fast conditional simulations through the LU decomposition of the covariance matrix," in Math. Geol., v. 19, n. 5, pp. 369-387.

Alabert, F., 1987b, Stochasting imaging of spatial distributions using hard and soft information, M.S. thesis, Stanford University, 197 p.

Anderson, T., 1958, An introduction to multivariate statistical analysis, Wiley and Sons, 374 –

Fogg, G., 1986, Groundwater and sand body interconnectedness in a thick multiple-aquifer system, in Water Resources Research, 22(5), 679-694.

Giordano, R., S. Salter, and K. Mohanty, 1985, "The effects of permeability variations on flow in porous media," SPE Paper 14365, 60th SPE Annual Conference, Las Vegas.

Hewett, T., 1986, "Fractal distributions of reservoir heterogeneity and their influence on fluid transport," SPE Paper 15386, 61st SPE Annual Conference, New Orleans.

Jones, D., 1979, Elementary information theory, Clarendon Press, Oxford, 182 p.

Journel, A. and C. Huijbregts, 1978, Mining Geostatistics, Academic Press, Chap. 7, 600 p.

Journel, A., 1986, "Constrained interpolation and qualitative information," in Math. Geol., v. 18, n. 3, pp. 269-286.

Journel, A.G. and Alabert, F., 1988, Focusing on spatial connectivity of extreme-values attributes: Stochastic Indicator models of reservoir heterogeneities, SPE paper 18324

Luster, G., 1985, Raw materials for Portland cement: Applications of conditional simulations of coregionalization, Ph.D. thesis, Stanford University, 531 p.

Mantoglou, A., and J. Wilson, 1981, Simulation of random fields with the turning band method, Report No. 264, Department of Civil Engineering, M.I.T., 199 p.

Thadani, S., Alabert, F. and Journel, A.G., 1987, An integrated geostatistical/pattern recognition technique for the characterization of reservoir spatial variability, in Proc. of SEG Annual Meeting, New Orleans, 1987

Xiao, H., 1985, A description of the behavior of indicator variograms for a bivariate normal distribution, M.S. thesis, Stanford University, 51 p.

SUBJECT INDEX

TERRA REVIEW

Non-Gaussian data expansion in the Earth Sciences

A.G. Journel and F. Alabert
Stanford Center for Reservoir Forecasting, Applied Earth Sciences Department, Stanford University

ABSTRACT A formalism is proposed to generate alternative equiprobable images of an underlying population spatial distribution. The resulting images honour data values at their locations and reflect important characteristics of the data such as patterns of spatial connectivity of extreme-values. The formalism capitalizes on a coding of all information available into bits (0–1), which are then processed all together accounting for their patterns of correlation in space. Such common coding allows accounting for qualitative information, possibly of an interpretative nature, to complement the usually sparse hard data available in Earth Sciences applications. The approach proposed, although of a probabilistic nature, does not call for any Gaussian-type modelling or hypothesis.
TERRA Nova 1989, 1, 123–134.

INTRODUCTION

Popular understanding tends to limit statistics to a data reduction (descriptive) rôle: large tables are summarized into a few numbers or graphs such as mean, median, histograms and pie charts. In fact in many applications statistics are taken towards a data expansion (inference) objective. In Earth Sciences applications, for example, the ratio of sample size to population size is usually extremely low, from 10^{-6} to 10^{-9} or below. The problem is not to reduce the information but to expand it towards a picture of the considerably larger population. At the heart of that data expansion process is a probablistic model that allows for a number of controlled interpolations between the data available. The critical question is 'how good is the model and have enough alternative models been tried?'

Consider the grey-scale map of Fig. 1. It presents the spatial distribution of 40×40 air permeameter measurements taken on a vertical slab of Berea sandstone of dimensions 2×2 foot (Giordano *et al.*, 1985). These 1600 horizontal permeability measurements taken on a regular grid will be considered as providing an exhaustive sampling of the slab population at the scale considered. Note the clear diagonal banding and the strong connectivity of low values.

Typically, the image of Fig. 1, or at least its important features, must be reconstituted from a much smaller sample. Sixteen randomly located sample values were drawn from the 1600 reference database. Their locations are given by crosses on the location map in Fig. 12.

Figure 2 gives a first reconstitution of the reference map of Fig. 1. A generalized linear regression algorithm, also known as ordinary kriging (Goldberger, 1962; Journel and Huijbregts, 1978) is used to interpolate from the 16 data. The kriging algorithm is 'exact' in the sense that it honours the data values at their locations. The algorithm accounts for data spatial correlation through a covariance model. This covariance model is usually inferred from the experimental correlation observed between the data available. However, since statistical inference is not the subject of this paper, all covariance models used in this paper were derived from all 1600 values of the reference database.

Figure 2 reveals the well-known spatial smoothing characteristic of all regression and moving average techniques. The initial 16 data are honoured through local discontinuities whose amplitude depends on the high frequency content of the covariance model used.

Figure 3(a) gives a second reconstitution drawn from a stochastic Gaussian-related model (Journel, 1974). A Gaussian-related model is fully characterized by a marginal distribution or histogram, not necessarily normal, and by a single covariance function. Then realizations are drawn

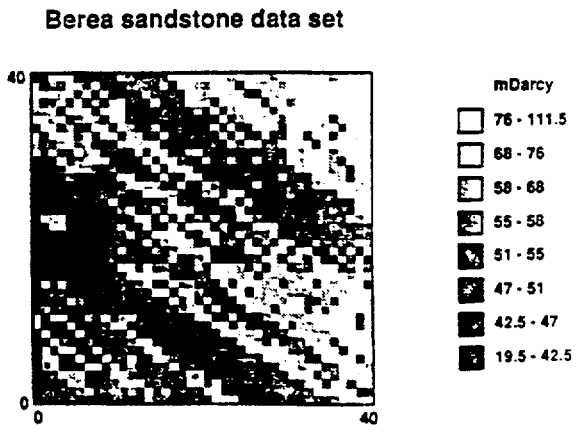


Fig. 1. The Berea data set. Colour-coded spatial distribution of 1600 permeability data.

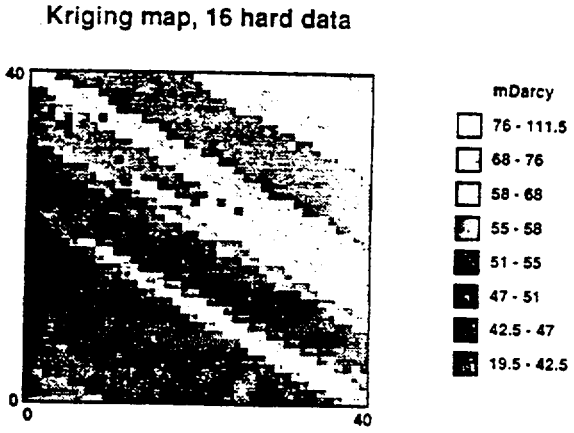


Fig. 2. A regression-type estimated map. (Ordinary kriging using 16 data.)

from it through a generalized Monte Carlo procedure. Figure 3(a) corresponds to one such drawing. Each realization is 'conditional' to the data (16) available, in the sense that it honours the data values at their locations. In addition, the simulated values of each realization are related in a way which ensures reproduction of the input covariance model.

If Fig. 3(a) reproduces reasonably well the diagonal banding of the exhaustive map of Fig. 1, it does so with excess local noise. This noise tends to blur the strong connectivity of low values. At a larger scale such connectivity of extreme-valued permeability may be the most important feature, controlling fluid flow, say in an oil reservoir or, alternatively a nuclear repository.

Figure 3(b) provides a third reconstitution based on a non-Gaussian stochastic model. This model accounts for the spatial connectivity of extreme-valued permeabilities through a sequence of indicator covariance models. As defined hereafter, an indicator covariance is a measure of the joint probability of two values in space exceeding a given permeability threshold. An indicator covariance is thus a model of two-point spatial connectivity. As was the case for the two previous reconstitutions of Figs 2 and 3(a), the 16 data values used are reproduced at their locations. However, the diagonal banding and particularly the connectivity of low values is reproduced with much less noise: compare with the exhaustive map of Fig. 1.

Figure 4 is based on the same non-Gaussian model and the same 16 data as used for Fig. 3(b). However, the anisotropy of the indicator covariance model (connectivity model) for the high permeability values has been intentionally inverted, generating banding of high values orthogonal to the true banding. The same 16 data, however, are honoured at their locations.

Since all four images of Figs 2–4 honour identically the same 16 data, they cannot be differentiated from statistics

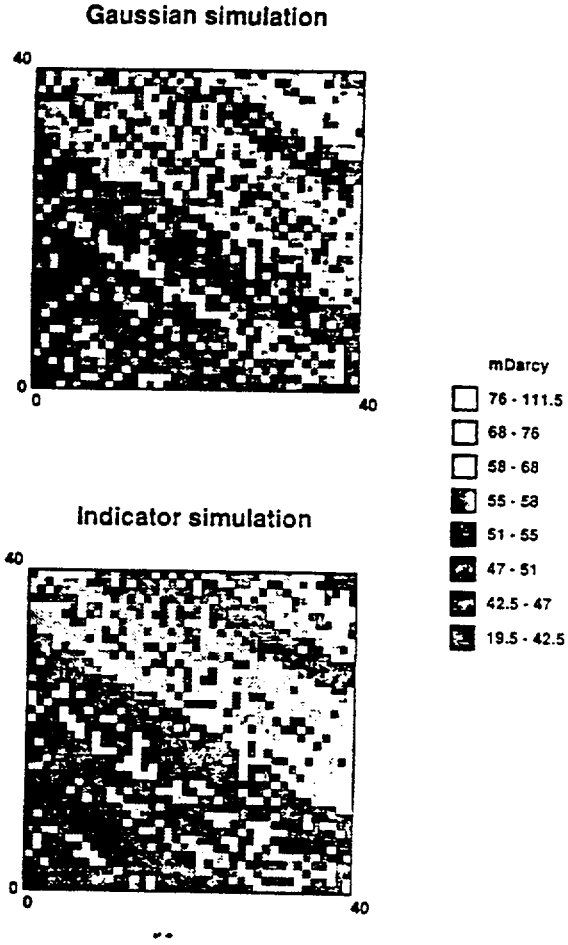


Fig. 3. Two reconstitutions using 16 data. (a) By drawing from a Gaussian-related model (b) Using an indicator simulation approach (non-Gaussian).

based on these data. Yet these four images appear quite different, and any planning based on them might yield vastly different decisions.

Imaging of spatial uncertainty

None of the three images of Figs 2 and 3 identifies the 'true' distribution of Fig. 1. In any process of data expansion, whether interpolation or simulation, there is necessarily some error involved and the resulting spatial uncertainty should be assessed.

A unique estimated map such as that of Fig. 2, no matter the criterion for which it is optimal, does not carry any measure of spatial uncertainty. A measure of spatial uncertainty is necessarily multivariate: it should not only address the uncertainty at any particular location $\mathbf{u}$ within a site but also the *joint* uncertainty along any string of locations $\mathbf{u}_l$, $l = 1,\ldots,L$ within that site. For example, if a string of high values is observed on any one of the reconstituted

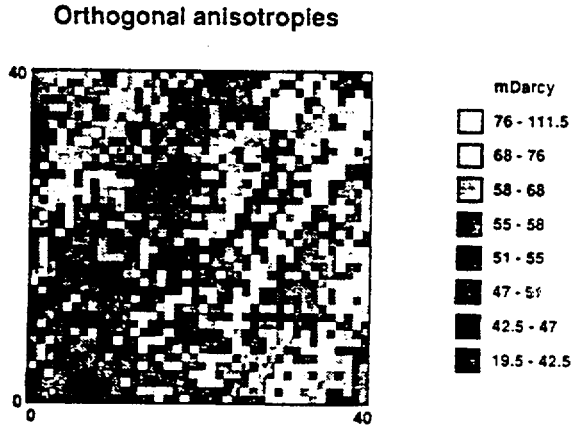


Fig. 4. A reconstitution with inverted anisotropy for the high values (non-Gaussian model and 16 data).

images of Figs 2 and 3 and the existence of such a string is highly consequential for the project at hand, the question arises as to how reliable that string is.

Stochastic simulation allows drawing several equiprobable realizations all honouring the same initial data at their locations. Figure 5 gives three such equiprobable simulations (reconstitutions), among many more possible, of the reference image of Fig. 1. All such simulations use the same non-Gaussian model and the same 16 conditioning data as used for the fourth simulation of Fig. 3(b). Spatial features observed consistently on all simulated images may be considered as reliable, i.e. representative of the true reference map, whereas features present on some simulations but not on others would be deemed unreliable.

These introductory examples emphasize the uncertainty faced in data expansion, and how the results depends on the prior model explicitly or implicitly adopted through the choice of an interpolation or simulation algorithm. The process of inducing a model, be it deterministic or probabilistic, from prior information is usually more important than the details of how to extrapolate (deduction process) from that model.

It behoves the cartographer to provide a readily usable assessment of spatial uncertainty, and the statistician to come up with stochastic models with greater flexibility than Gaussian-related models.

MODELLING SPATIAL DEPENDENCE

In most Earth Sciences related applications spatial dependence is an irrefutable hard fact that cannot be ignored. Identification and reproduction of patterns of dependence in space and or time might be the very goal of the statistical analysis. In mapping problems spatial dependence is a blessing for it allows building from the correlation of nearby data to estimate (predict) a particular unsampled value.

Consider the distribution in space of an attribute z, say that of horizontal permeability $z(\mathbf{u})$, $\mathbf{u}$ being the location on the vertical sandstone slab of Fig. 1. The spatial dependence between any two values $z(\mathbf{u})$ and $z(\mathbf{u}+\mathbf{h})$ separated by vector h is modelled by the bivariate probability distribution:

$$P\{Z(\mathbf{u}) \leq z,\ Z(\mathbf{u}+\mathbf{h}) \leq z'\}. \qquad (1)$$

A stationarity decision over a given domain (population) S allows inference of the probability distribution (1) or part thereof. The stationarity decision defines a space where repetitiveness can be found, giving physical significance to the probability (1) and allowing its inference from corresponding proportions of data pairs.

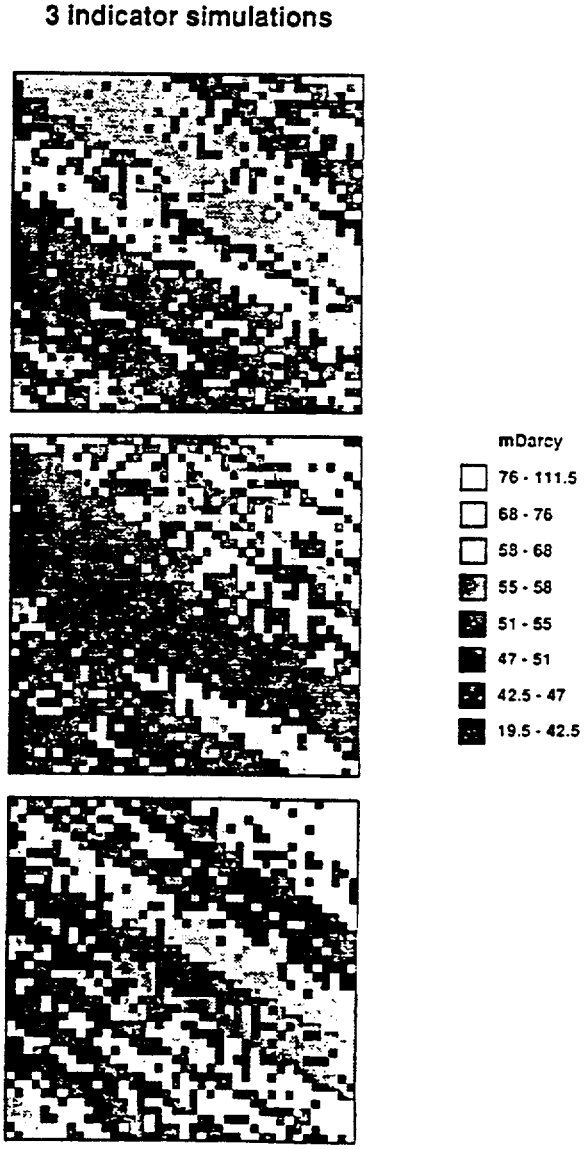


Fig. 5. Three equiprobable reconstitutions (using 16 data and the same non-Gaussian model as for Fig. 3b).

Indicator statistics

Indicator coding of information and indicator statistics are particularly convenient to infer characteristics of spatial dependence such as the bivariate probability distribution (1) (Switzer, 1977; Journel, 1983).

The indicator transform of the spatially distributed attribute $z(\mathbf{u})$ is defined as the binary variable:

$$i[z(\mathbf{u}); z] = i(\mathbf{u}; z) = \begin{cases} 1, \text{ if } z(\mathbf{u}) \leq z \\ 0, \text{ if not} \end{cases} \quad (2)$$

At any location $\mathbf{u}_j$ where a datum $z(\mathbf{u}_j)$ is available, the whole series of indicator data $i(\mathbf{u}_j; z)$ for all threshold values z is also available.

The indicator values $i(\mathbf{u}; z)$ are interpreted as outcomes of indicator random variables $I(\mathbf{u}; z)$. The marginal distribution of the binary random variable $I(\mathbf{u}; z)$ is fully characterized by a single parameter, its mean or expected value, which is seen to be the marginal probability distribution of $Z(\mathbf{u})$ for the threshold value z:

$$E\{I(\mathbf{u}; z)\} = 1 \times P\{Z(\mathbf{u}) \leq z\} + 0 \times P\{Z(\mathbf{u}) > z\}$$
$$= P\{Z(\mathbf{u}) \leq z\} = F(z). \quad (3)$$

Following from the decision of stationarity the distribution model $F(z)$ is made independent of the location $\mathbf{u}$ and is inferred from a spatial average of the corresponding indicator data available.

Similarly, the bivariate distribution (1) of any two random variables $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ is seen to be the expected value of the product of two indicator random variables, e.g. in the case $z = z'$:

$$E\{I(\mathbf{u}; z) \cdot I(\mathbf{u}+\mathbf{h}; z)\} = P\{Z(\mathbf{u}) \leq z, Z(\mathbf{u}+\mathbf{h}) \leq z\}.$$
$$= F(\mathbf{h}; z). \quad (4)$$

Expression (4) appears to be the non-centred covariance of the indicator process $I(\mathbf{u}; z)$. There is one such indicator covariance for each new threshold value z.

The indicator correlogram, or measure of linear correlation between two indicators $I(\mathbf{u}; z)$ and $I(\mathbf{u}+\mathbf{h}; z)$ is the centred standardized covariance:

$$\rho_I(\mathbf{h}; z) = \frac{F(\mathbf{h}; z) - F^2(z)}{F(z)[1-F(z)]} \in [-1, +1] \quad (5)$$

with: $\text{Var}\{I(\mathbf{u}; z)\} = F(z)[1-F(z)]$.

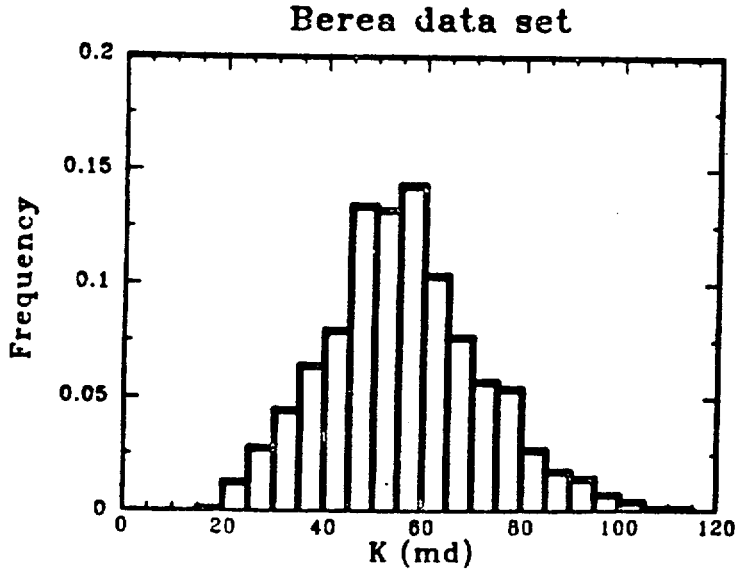
The threshold parameter z can be written as a p-quantile z_p such as: $F(z_p) = p \in [0, 1]$, in which case the previous indicator statistics (4) and (5) are parametrized with the probability value p.

$$E\{I(\mathbf{u}; z_p)\} = E\{I(\mathbf{u}; p)\} = p$$
$$\text{Var}\{I(\mathbf{u}; p)\} = p(1-p)$$
$$\rho_I(\mathbf{h}; z_p) = \rho_I(\mathbf{h}; p). \quad (6)$$

The indicator formalism allows inferring Z-probability distributions through indicator statistics: mean indicator data for the marginal distribution $F(z)$, indicator covariance for the bivariate distribution $F(\mathbf{h}; z)$, (Journel, 1983). For a given set of data locations, inference of the indicator covariances is usually easier than inference of the traditional Z-covariance, at least for not too extreme threshold values of z. Also indicator covariance models can be synthesized, e.g. from interpretative hand-drawn maps, to reflect patterns of spatial continuity of the process $Z(x)$.

Besides its relation to the bivariate probability $F(\mathbf{h}; z)$ the indicator correlogram $\rho_I(\mathbf{h}; p)$ can be read as a unit-free measure of two-points spatial connectivity: the higher $\rho_I(\mathbf{h}; p)$ the greater the probability of having two values $z(\mathbf{u})$

Table 1. Statistics of the 1600 reference permeability data.



number of data	1600
mean	55.526
variance	249.
coefficient of variation	0.284
skewness	0.379
kurtosis	3.127
minimum	19.5
maximum	111.5
first quartile	45.0
median	55.0
third quartile	65.0

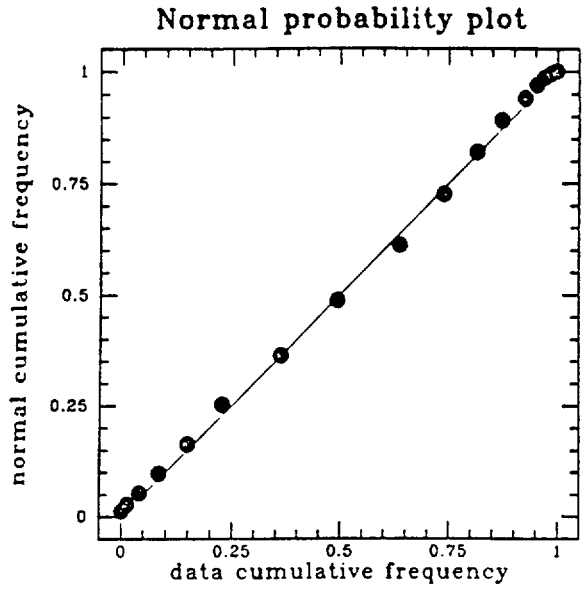


Fig. 6. Normal probability plot of the reference 1600 data (note the quasi-normality of that data set).

and $z(\mathbf{u}+\mathbf{h})$ jointly not exceeding (or equivalently exceeding) the same threshold value z_p. Conversely, when $\rho_I(\mathbf{h}; p) = 0$ from relation (5) it appears that: $F(\mathbf{h}; z_p) = F^2(z_p) = p^2$, i.e. the two events $Z(\mathbf{u}+\mathbf{h}) \leq z$ and $Z(\mathbf{u}) \leq z$ are independent.

The quantification of spatial connectivity using indicator covariances has been anticipated by a few authors, see in particular (Haralick *et al.*, 1973), but not actually put to work for stochastic imaging.

Table 1 gives the statistics of the 1600 permeability data shown in Fig. 1. The corresponding cumulative histogram, i.e. the indicator mean $E\{I(\mathbf{u}; p)\} = F(z_p)$ is plotted on a normal graph in Fig. 6.

The quasi-straight line indicates a good fit by a Gaussian (normal) model. However this marginal Gaussian character does not imply any multivariate Gaussian character as shown by the following figures.

Figures 7–9 give the indicator correlogram maps, $\rho_I(\mathbf{h}; p)$ corresponding to the $p = 0.1$, 0.5, and 0.9 quantile threshold values of z_p. The experimental correlograms calculated from the 1600 indicator data are given on the left of the maps while the models calculated from a bi-Gaussian hypothesis are given on the right. The maximum correlation value $\rho_I(0; p) = 1.0$ is plotted at the centre of each map. Each two-dimensional correlogram value $\rho_I(h_u, h_v; p)$, with $\mathbf{h} = (h_u, h_v)$, is plotted at a point h_u-units left of and h_v-units above the centre point. Such correlation maps allow a global appreciation of directional patterns of spatial dependence.

The generation of the bi-Gaussian models on the right-hand side of Figs 7–9 is discussed in the next section.

All three experimental correlogram maps exhibit strong anisotropy: the connectivity measure $\rho_I(\mathbf{h}; p)$ depends not only on the length $|\mathbf{h}|$ of the vector $\mathbf{h}$ but also on its direction. Maximum continuity is obtained in the N57°W direction of banding: refer to Fig. 1. Most striking are the differences in correlation behaviour as the threshold value z_p increases.

- At the low threshold value $z_{0.1} = 35.5$ *md*, indicator correlation ($\rho > 0.5$) persists beyond the map limits in the direction of banding but vanishes very quickly in the direction across banding.
- At the median threshold value $z_{0.5} = 55.0$ *md*, the indicator correlation anisotropy is less marked.

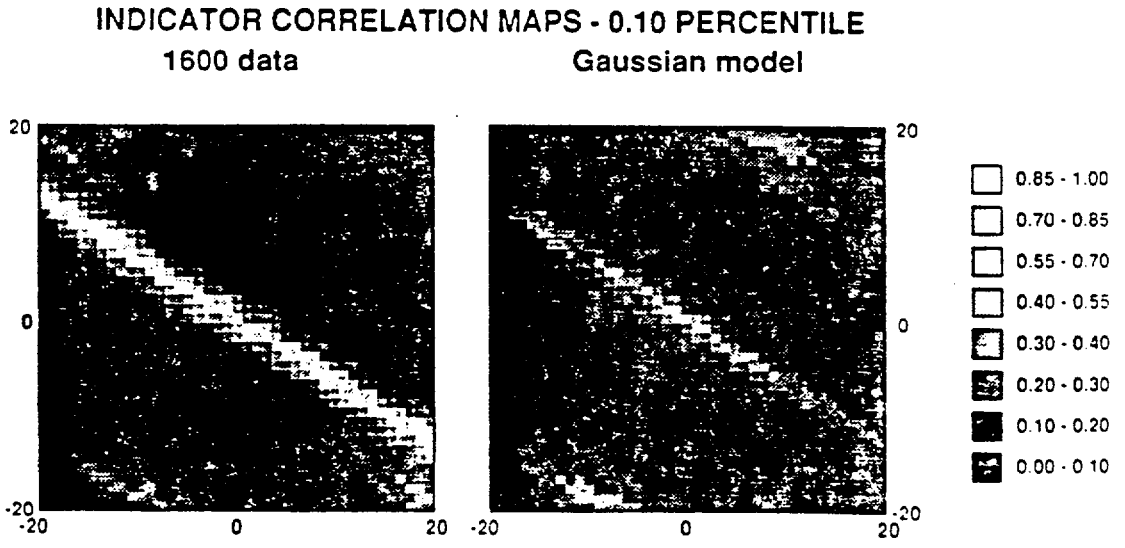


Fig. 7. Berea indicator correlogram $\rho_I(\mathbf{h}; z_{0.1} = 35.55\ md)$. (a) Experimental values (from 1600 data). (b) Gaussian model.

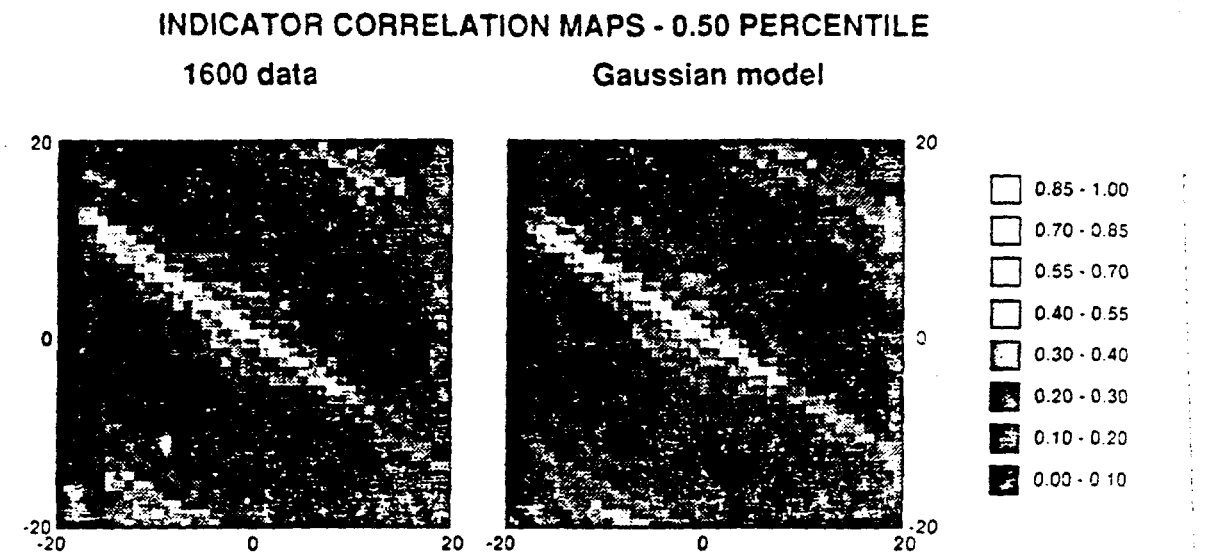


Fig. 8. Berea indicator correlogram $\rho_I(\mathbf{h}; z_{0.5} = 55.0\ md)$.(a) Experimental values (from 1600 data). (b) Gaussian model.

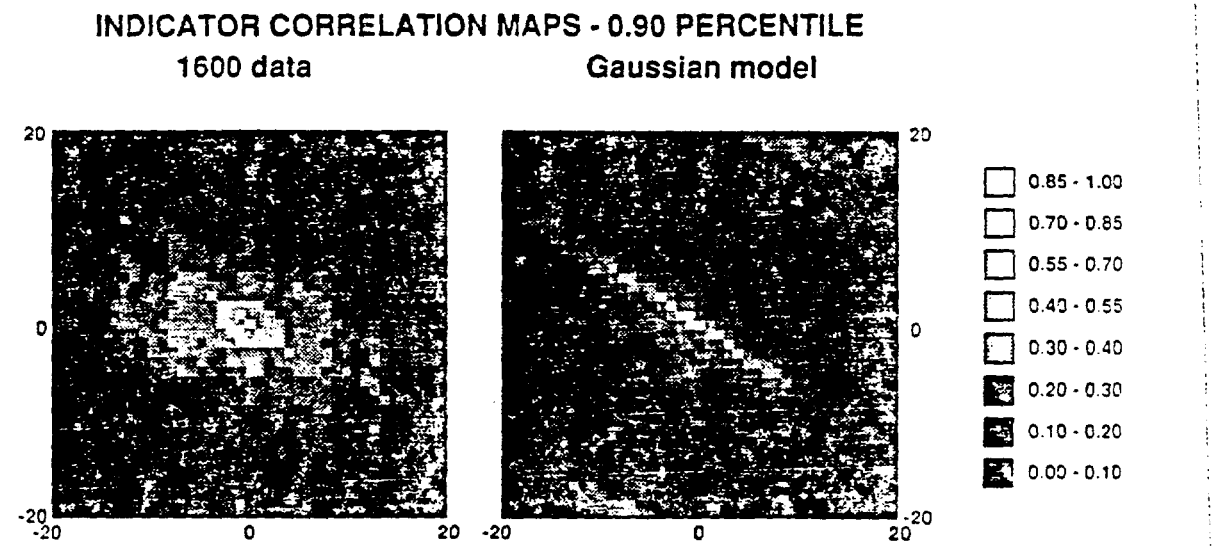


Fig. 9. Berea indicator correlogram $\rho_I(\mathbf{h}; z_{0.9} = 76\ md)$.(a) Experimental values (from 1600 data). (b) Gaussian model.

- At the high threshold $z_{0.9} = 76\ md$, leaving only 10% of Z-data above it, experimental indicator correlation becomes almost isotropic: there is no strong banding of high permeability values as can be checked on Fig. 1. Indicator correlation in the direction across banding is much stronger at high threshold values than it is at low threshold values.

The three experimental indicator correlation maps show that the pattern of spatial connectivity (correlation and anisotropy) is clearly different for different classes of permeability values.

As a measure of comparison, Fig. 10 gives the Z-correlogram map, $\rho_Z(\mathbf{h})$, of the z-permeability attribute itself, again calculated from all 1600 data. This Z-correlogram map appears as intermediary between the three indicator correlogram maps at the left of Figs 7–9. This is no surprise since it is known that an attribute covariance is the average of all its indicator covariances and cross-covariances (Alabert, 1987).

$$C_Z(\mathbf{h}) \int_{z\,min}^{z\,max} dz \int_{z\,min}^{z\,max} C_I(\mathbf{h}; z, z')dz' \qquad (7)$$

with $C_I(\mathbf{h}; z, z') = \mathrm{Cov}\{I(\mathbf{u}; z), I(\mathbf{u}+\mathbf{h}; z')\}$.

However, the Z-correlogram map of Fig. 10 understates severely the spatial connectivity and anisotropy of low permeability values as evidenced in Fig. 1 and at the left of Fig. 7. Conversely, Fig. 10 overstates the banding (connectivity

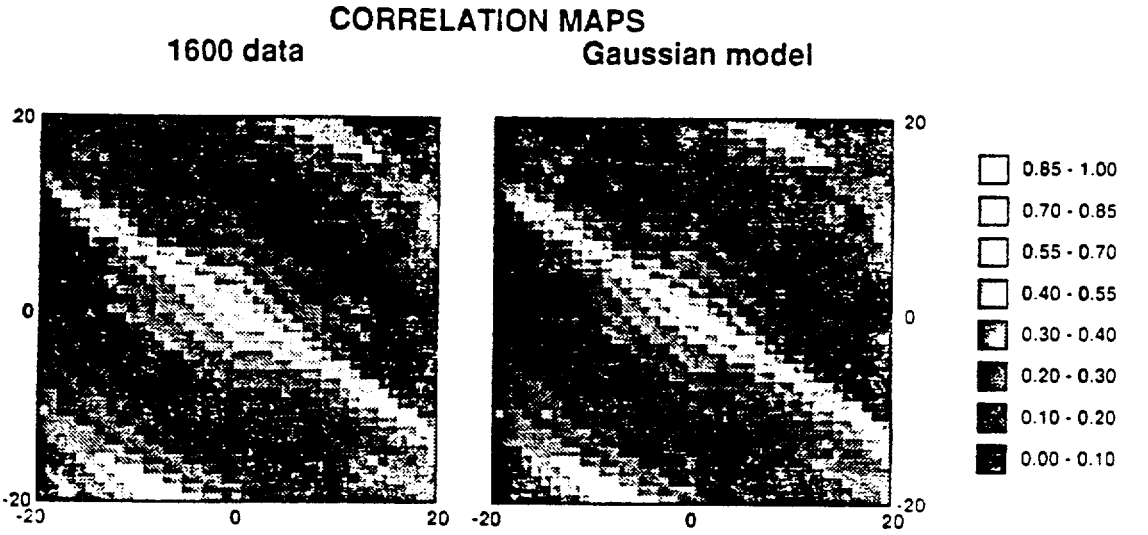


Fig. 10. Berea permeability correlogram (from original 1600 permeability data). (a) Experimental values (from 1600 data). (b) Gaussian model.

and anisotropy) of high permeability values as seen on the left of Fig. 9.

If detection of such connectivity of low values (flow barriers or aquitards) is critical, the average correlation image given by the Z-correlogram of Fig. 10 is inadequate.

The Gaussian attraction

The Gaussian model holds a dominant, though questionable, place in statistical theory for it provides a full multivariate probability distribution at very little inference cost.

First a normal score transform is applied to the original z-data. This transform allows the defining of a new set of z-data with standard normal marginal distribution (Anderson, 1958):

$$y(\mathbf{u}) = T[z(\mathbf{u})], \text{ and: } z(\mathbf{u}) = T^{-1}[y(\mathbf{u})]. \quad (8)$$

The transform T is usually made monotone and invertible so that any proposition made on the ys can be transformed back into a similar proposition made on the original zs. In particular, the multivariate distribution of the process $Z(\mathbf{u})$ is expressed in terms of the multivariate distribution of the Gaussian process $Y(\mathbf{u})$, itself fully determined from the Y-correlogram $\rho_Y(\mathbf{h})$:

$$P\{Z(\mathbf{u}_j) \leq z_{p_j},\ j = 1,\ldots,n\} = P\{Y(\mathbf{u}_j) \leq y_{p_j},\ j = 1,\ldots,n\}, \quad (9)$$

with z_{p_j} and y_{p_j} being the p_j-quantiles of their respective marginal distributions.

The bivariate case, $n = 2$ in expression (9), is particularly illustrative because analytical expressions are available (Anderson, 1958; Xiao, 1985; Abramovitz and Stegun, 1964). The Z-bivariate distribution, expressed in terms of indicator correlogram as in relations (5) and (6), is written:

$$\rho_I(\mathbf{h};\ p) = \frac{1}{2\pi p(1-p)} \int_0^{\arcsin \rho_Y(\mathbf{h})} \exp\left(-\frac{y_p^2}{1+\sin\theta}\right) d\theta, \quad (10)$$

with y_p being the standard normal p-quantile, and: $\rho_Y(\mathbf{h}) = E\{Y(\mathbf{u}) \cdot Y(\mathbf{u}+\mathbf{h})\}$ being the Y-correlogram.

The Gaussian model (10) possesses an extreme, and unfortunately not always well understood, property of maximum entropy (Jones, 1979) i.e. of maximum disorder for a given covariance $\rho_Y(\mathbf{h})$. This maximum entropy character is shown by the behaviour of the Gaussian indicator correlogram (10) for extreme values of the threshold y_p: when p tends towards zero or one, the normal p-quantile tends towards $\pm\infty$, and the expression under the integral sign (10) tends towards zero leaving:

$$\rho_I(\mathbf{h};\ p) \to 0. \quad (11)$$

The Gaussian model does *not* allow for any indicator correlation at extreme threshold values, i.e. for any spatial connectivity of extreme values. Moreover, the pattern of loss of indicator correlation is symmetric as the threshold value z_p moves away from the median value $z_{0.5}$. Indeed the expression (10) is symmetric about $p = 0.5$, since $y_p^2 = y_{1-p}^2$.

Figures 7–9 provide side by side the experimental Berea sand indicator correlogram maps and the corresponding Gaussian models as calculated from expression (10). The experimental correlogram maps show patterns of correlation that are neither symmetric about $p = 0.5$, nor decreasing as p comes closer to zero or one. In fact, the best indicator correlation is found for an extreme value $p = 0.1$ in the direction of low-permeability banding. The divergence between the experimental indicator correlations and the Gaussian models is seen to be larger for the two extreme thresholds ($p = 0.1$ and 0.9).

Multiple steps connectivity

Just as the two points (**u**) and (**u**+**h**) connectivity is characterized by the mean of the product of two indicator variables $I(\mathbf{u}; z)$ and $I(\mathbf{u}+\mathbf{h}; z)$ as in relation (4), a three-point connectivity can be characterized by the mean of the product of three indicator variables $I(\mathbf{u}; z)$, $I(\mathbf{u}+\mathbf{h}; z)$, $I(\mathbf{u}+2\mathbf{h}; z)$:

$$E\{I(\mathbf{u}; z) \cdot I(\mathbf{u}+\mathbf{h}; z) \cdot I(\mathbf{u}+2\mathbf{h}; z)\}$$

Similarly, n-point connectivity in a particular direction can be characterized by the mean of the product of n indicator variables:

$$E\left\{\prod_{j} I[\mathbf{u}+(j-1)\mathbf{h}; z]\right\} = \phi(n).$$

The elementary lag vector h considered hereafter has for direction N57°W the direction of banding on Fig. 1; the modulus $|\mathbf{h}|$ is the basic step between two pixels in that direction. The threshold value considered is the low 0.1-quantile $z_{0.1} = 35.5$ *md* used to specify connectivity of low values. Figure 11 gives three experimental connectivity functions $\phi(n)$, for up to $n = 16$ steps, in the N57°W direction. The upper curve corresponds to the 1600 data of the reference map shown in Fig. 1. The middle curve corresponds to the simulated map of Fig. 3(b) built from a non-Gaussian model. The lower curve corresponds to the simulated map of Fig. 3(a) using a Gaussian-related model.

The reference map (upper curve) exhibits the strongest connectivity followed by the simulated map using a non-Gaussian model. The Gaussian-related model resulted in the lowest connectivity.

A map based on spatial independence would entail:

$\phi(n) = [F(z_p)]^n = p^n$, for all n.

On Fig. 11, for $n = 2$, it is read:

$\phi(2) = 0.065$ for both the reference map and the non-Gaussian simulated map, Figs 1 and 3(b) respectively,
$\phi(2) = 0.038$ for the Gaussian-related map of Fig. 3(a),
$\phi(2) = p^2 = 0.01$ would correspond to spatial independence.

When used for data expansion a maximum entropy model, such as the Gaussian model, need not be the most conservative. In the Earth Sciences and civil engineering, low entropy patterns such as connected strings of extreme values often correspond to hazardous features whose potential occurrence should not be understated. Areas with potential clusters of low soil strength are hazardous pile sites for construction; strings of high permeability values can be leakage conduits hazardous for the design of a nuclear repository or detrimental to oil sweep efficiency in a reservoir. When mapping in the field or drawing on a board geologists seek low entropy features. Conversely statisticians prefer not to commit themselves and would go for maximum entropy models, not always realizing the adverse consequence of such models for certain applications.

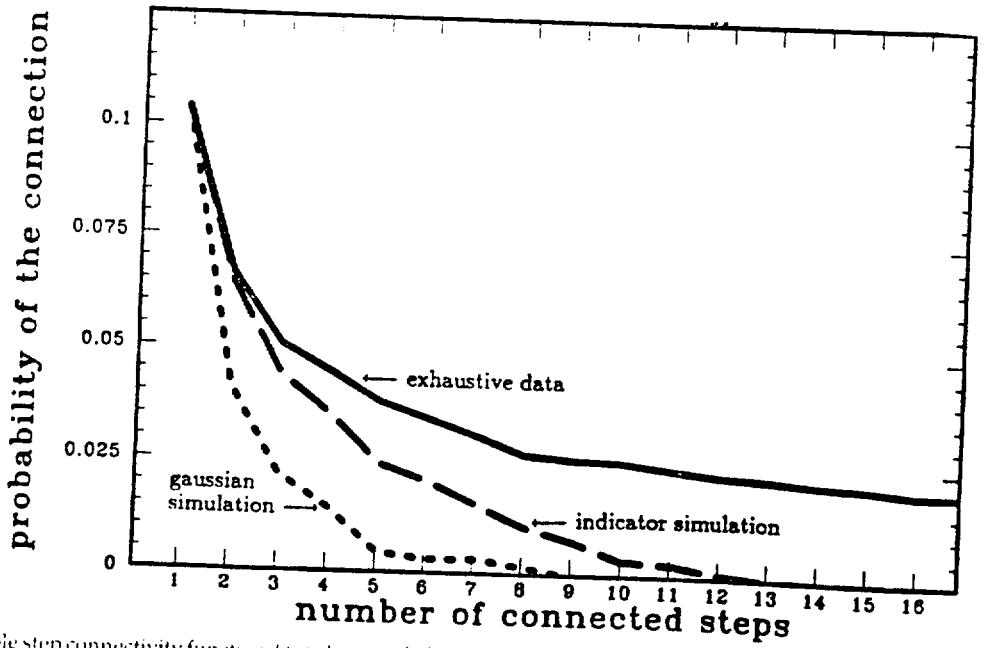


Fig. 11. Multiple step connectivity function $\phi(n)$: $\phi(n)$: probability that n pixels in a row be valued less than or equal to $z_{0.1} = 35.5$ *md* in the direction N57°W. (a) Map corresponding to Fig. 1. (b) Map corresponding to Fig. 3(b) (indicator simulation). (c) Map corresponding to Fig. 3(a) (Gaussian simulation).

Tractable non-Gaussian models for data expansion which can account for low entropy characteristics of the data spatial distribution should be proposed.

SEQUENTIAL INDICATOR SIMULATION

Consider any two random events, A_1 and A_2, with joint probability denoted by $P(A_2, A_1)$. For example, A_1 can be the event '$Z(\mathbf{u}) \leq z$' related to the permeability value $z(\mathbf{u})$ at location $\mathbf{u}$ or it could be the event 'earthquake of magnitude greater than 7', or it could be 'rock type i is present at location $\mathbf{u}$'. The two events A_2 and A_1 need not relate to the same variable, for example A_2 may correspond to the question 'is $z(\mathbf{u}) \leq z$' whereas A_1 may correspond to an information of interpretative nature 'sand is present at location u.'

The conditional probability of event A_2 knowing that event A_1 has occurred is given by Bayes postulate (Thomas, 1986):

$$P(A_2|A_1) = P(A_2, A_1) / P(A_1),$$

with $P(A_1)$ being the marginal probability of event A_1, and $P(A_2, A_1)$ being the joint probability of events A_1 and A_2.

Joint simulation of the two events can be done by simulating first the event A_1 from its marginal probability distribution $P(A_1)$, then simulating the event A_2 from the conditional probability distribution $P(A_2|A_1)$. This sequence requires inference of the conditional probability $P(A_2, A_1)$.

More generally, any number of dependent events A_j, $j = 1,\ldots,N$, can be sequentially simulated using the expression (Devroye, 1986):

$$P(A_j,\ j = 1,\ldots,N) =$$
$$P(A_N|A_j,\ j = 1,\ldots,N-1) \cdot P(A_{N-1}|A_j,\ j = 1,\ldots,N-2). \quad (12)$$
$$P(A_{N-2}|A_j,\ j = 1,\ldots,N-3) \cdots P(A_2|A_1) \cdot P(A_1).$$

The technique requires inference of the successive $(N-1)$ conditional probability distributions $P(A_2|A_1),\ldots, P(A_N|A_j,\ j = 1,\ldots,N-1)$.

Relation (12) is absolutely general, in the sense that the N event A_j can be of any nature. In particular it can be applied to the simulation of joint realizations of N-dependent continuous random variables $Z(\mathbf{u}_j)$, $j = 1,\ldots,N$ modelling the spatial variability of an attribute $z(\mathbf{u})$ over a grid with N nodes $\mathbf{u}_j$. The sequence is:

1. Starting at *any* node, say $\mathbf{u}_1$, derive the conditional distribution of $Z(\mathbf{u}_1)$ given any available information. This available information, denoted (n), usually consists in known outcome values $Z(\mathbf{u}_\alpha) = z_\alpha$ at data locations $\mathbf{u}_\alpha$. The conditional or posterior distribution of $Z(\mathbf{u}_1)$ is denoted:

$$P\{Z(\mathbf{u}_1) \leq z|(n)\} = P\{Z(\mathbf{u}_1) \leq z|Z(\mathbf{u}_\alpha) = z_\alpha,\ \alpha \in (n)\}.$$

2. Draw a realization of $Z(\mathbf{u}_1)$, say $z_1^{(l)}$, from that distribution. Transfer that realization into the data set which is now of dimension $(n+1)$.

3. Move *at random* to a second node, say $\mathbf{u}_2$. Derive the conditional distribution of $Z(\mathbf{u}_2)$ given the information $(n+1)$, i.e. the distribution:

$$P\{Z(\mathbf{u}_2) \leq z|(n+1)\} = P\{Z(\mathbf{u}_2) \leq z|Z(\mathbf{u}_\alpha) = z_\alpha,\ \alpha \in (n),$$
$$Z(\mathbf{u}_1) = z_1^{(l)}\}.$$

Draw a realization of $Z(\mathbf{u}_2)$, say $z_2^{(l)}$, from that distribution. Transfer that realization $z_2^{(l)}$ into the data set which then becomes of dimension $(n+2)$.

Loop at random over all N nodes, until each node is informed with a simulated value $z_j^{(l)}$.

The set of N values $\{z_j^{(l)},\ j = 1,\ldots,N\}$ represents one realization of the random field $\{Z(\mathbf{u}),\ \mathbf{u} \in S\}$ over the N nodes $\mathbf{u}_j$. Any number L of such simulations $\{z_j^{(l)},\ j = 1,\ldots,N\}$, $l = 1,\ldots L$, can be obtained by repeating the entire sequential process L times.

Two major implementation problems must be addressed:

- the derivation of the $(N-1)$ successive conditional probability distributions of type $P\{Z(\mathbf{u}_j) \leq z|(n+j-1)\}$
- the increasing size of the conditioning information from (n) to $(n-N+1)$.

Derivation of conditional distributions (Journel, 1983; Sullivan, 1985; Suro Perez, 1988)

The event $Z(\mathbf{u}) \leq z$ can be characterized by a random binary indicator variable, as defined in (2):

$$I(\mathbf{u}; z) = \begin{cases} 1, \text{ if } Z(\mathbf{u}) \leq z \\ 0, \text{ if not} \end{cases}$$

Any conditional probability for $Z(\mathbf{u})$ can be written as a conditional expectation of an indicator random variable, similarly to relation (3):

$$P\{Z(\mathbf{u}) \leq z|(n)\} = E\{I(\mathbf{u}; z)|(n)\}. \quad (13)$$

Consider K threshold values z_k, $k = 1,\ldots,K$ providing a discrete approximation of the range of variability of the attribute z. Each conditioning datum $Z(\mathbf{u}_\alpha) = z_\alpha$ can be coded into an indicator data column with K members valued either zero or one:

$$Z(\mathbf{u}_\alpha) = z_\alpha \Rightarrow \{i(\mathbf{u}_\alpha; z_k),\ k = 1,\ldots,K\}. \quad (14)$$

The previous probability distribution can be seen as conditional to the n indicator data columns, i.e. to the $n \times K$ indicator data:

$$P\{Z(\mathbf{u}) \leq z_k|Z(\mathbf{u}_\alpha);\ = z_\alpha,\ \alpha \in (n)\}$$
$$= E\{I(\mathbf{u}; z_k)|I(\mathbf{u}_\alpha; z_k) = i(\mathbf{u}_\alpha; z_k),$$
$$k = 1,\ldots,K;\ \alpha \in (n)\}, \quad (15)$$

with z_{k_0} being one of the K threshold values z_k considered.

Arguing or assuming that $I(\mathbf{u}; z_{k_0})$ is more correlated with $I(\mathbf{u}_\alpha; z_{k_0})$ than with any of the other indicator data $I(\mathbf{u}_\alpha; z_k)$ with $z_k \neq z_{k_0}$, the probability (15) can be approximated by:

$$P\{Z(\mathbf{u}) \leq z_{k_0} | (n)\} = E\{I(\mathbf{u}; z_{k_0}) | I(\mathbf{u}_\alpha; z_{k_0}) = i(\mathbf{u}_\alpha; z_{k_0}), \alpha \in (n)\}. \quad (16)$$

This approximation amounts to ignoring all cross-correlation between indicators defined at different threshold values. The result is a considerable reduction in the number of conditioning indicator data, n instead of Kn. In each particular application, the indicator cross-correlations can be calculated and compared to the indicator auto-correlations of type (5). In cases when these indicator cross-correlations cannot be ignored, a more demanding formalism using indicator principal components can be called for (Suro Perez, 1988).

A conditional expectation can be written as a function ϕ of the conditioning information, i.e. for the conditional expectation (16):

$$\begin{aligned} &E\{I(\mathbf{u}; z_{k_0}) | I(\mathbf{u}_\alpha; z_{k_0}) = i(\mathbf{u}_\alpha; z_{k_0}), \alpha \in (n)\} \\ &= \phi\{i(\mathbf{u}_\alpha; z_{k_0}), \alpha \in (n)\} \\ &= a_0 + \sum_{\alpha \in (n)} a_1(\alpha) i(\mathbf{u}_\alpha; z_{k_0}) \\ &+ \sum_{\alpha \in (n)} \sum_{\alpha' \in (n); \alpha \neq \alpha'} a_2(\alpha, \alpha') i(\mathbf{u}_\alpha; z_{k_0}) \cdot i(\mathbf{u}_{\alpha'}; z_{k_0}) + \ldots \\ &+ a_n \cdot \prod_{\alpha \in (n)} i(\mathbf{u}_\alpha; z_{k_0}), \end{aligned} \quad (17)$$

where the symbol Π indicates multiplication.

Indeed any function of one single indicator variable can be written as a linear combination of that indicator $\phi(i) = a_0 + ai$. Similarly any function of any number n of indicator variables can be written as a linear combination of products of these indicators (Journel, 1986). Then, a first order approximation to the conditional probability (16) amounts to retaining only the first $(n+1)$ terms of the expansion (17):

$$P\{Z(\mathbf{u}) \leq z_{k_0} | (n)\} \simeq a_0 + \sum_{\alpha \in (n)} a_1(\alpha) \cdot i(\mathbf{u}_\alpha; z_{k_0}). \quad (18)$$

The argument is that introducing more terms of the expansion (17), such as the products $i(\mathbf{u}_\alpha; z_{k_0}) \cdot i(\mathbf{u}_{\alpha'}; z_{k_0})$, would require the use of trivariate and higher order statistics for which inference is rarely possible. Limiting the expansion (17) to functions of one indicator at a time as in (18) reduces the inference needed to only bivariate distribution functions expressed as indicator covariances, see definition (4).

It remains to determine the set of $(n+1)$ coefficients a_0, $a_1(\alpha)$ for the approximation (18). These coefficients are obtained by linear regression of the unknown indicator $i(\mathbf{u}; z_{k_0})$ on the n indicator data $i(\mathbf{u}_\alpha; z_{k_0})$. That regression calls only for the indicator correlogram $\rho_I(h; z_{k_0})$ as defined in relation (5). In its simplest form, the regression estimate is written (Journel, 1983):

$$[P\{Z(\mathbf{u}) \leq z_{k_0} | (n)\} - F(z_{k_0})]^* = \sum_{\alpha \in (n)} \lambda_\alpha(k_0) \cdot [i(\mathbf{u}_\alpha; z_{k_0}) - F(z_{k_0})] \quad (19)$$

with the n weights $\{\lambda_\alpha(k_0), \alpha \in (n)\}$ given by a system of normal equations.

$$\sum_{\beta \in (n)} \lambda_\beta(k_0) \cdot \rho_I(\mathbf{u}_\alpha - \mathbf{u}_\beta; z_{k_0}) = \rho_I(\mathbf{u} - \mathbf{u}_\alpha; z_{k_0}), \text{ for all } \alpha \in (n). \quad (20)$$

The choice of a linear regression procedure to determine the weights a_α of expression (18) is not arbitrary. It can be shown (Journel, 1986) that a conditional expectation being by definition a $L2$-norm projection on the information space, a regression technique *must* be used to determine it. Also the normal system (20) can be seen as an identification of the correlation of the estimator (19) with any one of the n indicator data $I(\mathbf{u}_\alpha; z_{k_0})$ to the model indicator correlogram value $\rho_I(\mathbf{u} - \mathbf{u}_\alpha; z_{k_0})$. The K systems (20) for $k_0 = 1, \ldots, K$ are interpreted then as the conditions for the simulated values $\{z^{(j)}, j = 1, \ldots, N\}$ to reproduce the K indicator correlogram models $\rho_I(\mathbf{h}; z_{k_0})$, $k_0 = 1, \ldots, K$ (Alabert, 1987).

Dimension reduction

In practice, at each node $\mathbf{u}$, K normal systems of type (20) must be solved to generate a K-discrete version of the conditional probability distribution $P\{Z(\mathbf{u}) \leq z_{k_0} | (n)\}$, $k_0 = 1, \ldots, K$. The factor K is not a problem if the systems (20) are not of too large dimension n. But precisely, in the sequential simulation process over the N nodes $\mathbf{u}_j$, $j = 1, \ldots, N$, the number of conditioning data keeps growing from n for the first node being simulated to $n+N-1$ for the last node. A grid with $N = 10^6$ nodes is not extraordinary in Earth Sciences applications and thus would require the solving of unreasonably large systems.

The solution consists of retaining only those data locations $\mathbf{u}_\alpha$, closest to the node $\mathbf{u}$ being simulated. The argument is that the information carried by the further away conditioning data is 'screened' by the information content of the nearest data. For example, if the five nearest data are retained in each of four quadrants with apex at the node $\mathbf{u}$ being simulated, a normal system of maximum dimension 20 would have to be solved for each node, no matter how advanced is the sequential process $j = 1, \ldots, N$.

Conditioning to soft information

In many applications, particularly in the Earth Sciences, 'hard' numerical data of type (14): $Z(\mathbf{u}_\alpha) = z_\alpha$ are sparse. The reconstitution of the spatial distribution $\{z(\mathbf{u}), \mathbf{u} \in S\}$ is thus likely to be poor. Fortunately, hard data are usually complemented by a host of 'soft' data stemming from different and less accurate measurement devices or from interpretative origin (e.g. geological interpretation). The reconstitution can be dramatically improved if that soft information is accounted for.

The indicator formalism offers great flexibility for coding information of various sources and qualities into a common format; that of zero/one bits (indicator data):

- A hard datum $Z(\mathbf{u}_\alpha) = z_\alpha$ generates a complete indicator data column as in expression (14).
- A constraint interval generates an incomplete indicator data column:

$$Z(\mathbf{u}_\alpha) \in [a, b] = i(\mathbf{u}_\alpha; z_k) = \begin{cases} 0, \text{ if } z_k < a \\ 1, \text{ if } z_k \geq b \\ \text{undefined otherwise.} \end{cases} \quad (21)$$

- Local prior distributions are coded as 'fuzzy' indicator data valued between 0 and 1, as opposed to 'hard' indicator data valued either zero or 1:

$$Z(\mathbf{u}_\alpha) \in [a, b] = i(\mathbf{u}_\alpha; z_k) = \begin{cases} 0, \text{ if } z_k < a \\ \in [0, 1], \text{ if } z_k \in [a, b] \\ 1, \text{ if } z_k \geq b \end{cases} \quad (22)$$

For example, at a location $\mathbf{u}_\alpha$ where a hard measurement $z(\mathbf{u}_\alpha)$ could not be taken, information about the prevailing rock type may be available. That rock type information may indicate in which interval $[a, b]$ the attribute value $z(\mathbf{u}_\alpha)$ resides. In addition the distribution (histogram) of z-values within that rock type may be available and that information can be coded as a prior local distribution of type (22) and put to use.

The indicator data regardless of their origins, are then processed all together for:

- Inference of the required indicator correlograms $\rho_I(\mathbf{h}; z_k)$, $k = 1, \ldots, K$.
- Conditioning the probability distribution (13) required for generating the simulated maps.

In addition to the 16 hard data, 144 constraint intervals of type (21) were given at the nodes of a regular 12×12 grid, (see the location map of Fig. 12). At each of these 144 locations soft information indicates that the permeability value $z(\mathbf{u}_\alpha)$ belongs to one of the three intervals [0, 35.5], [35.5, 76], [76, 150]. This soft information mimicks the imprecise permeability information one might be able to extract from well logs or, in the best case, from dense seismic lines.

One simulated field using that additional soft information is shown in Fig. 13, and should be compared to the simulated fields of Figs 3(b) and 5 which use only the 16 hard data. The improvement achieved by the incorporation of the soft information is appreciable: compare Figs 5 and 13 to the reference Fig. 1.

In addition to allowing a coding of soft local information, the indicator formalism lends itself to utilizing soft information of a structural nature. Indeed, the indicator simulation approach utilizes K connectivity functions $\rho_I(\mathbf{h}; z_k)$, $k = 1, \ldots, K$, as opposed to a single one $\rho_Y(\mathbf{h})$ for the Gaussian approach. Some aspects, such as spatial anisotropy, of these connectivity functions may be obtained directly from geological (soft structural) information. For example, it may be known beforehand that low permeability values are connected along N57°W whereas high permeability values are connected in the orthogonal direction N33°E.

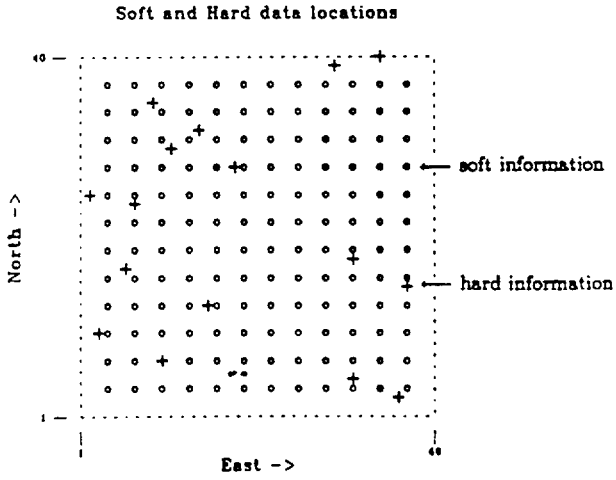


Fig. 12. Location of available information. + 16 hard data. ○ 144 constraint intervals (soft data).

Simulation with soft data

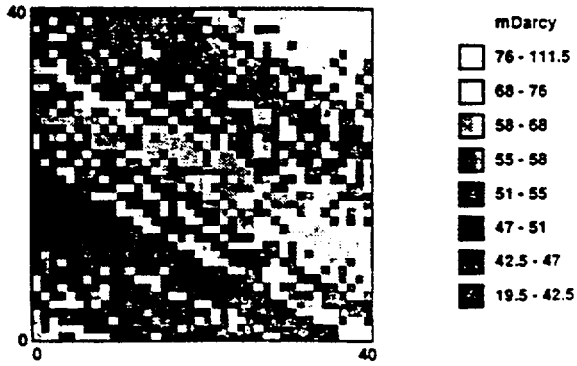


Fig. 13. A reconstitution using both hard and soft information (compare with Figs 3b and 5).

The anisotropy modelled for $\rho_I(\mathbf{h}; z_{c1})$ at the low threshold value z_{c1} may then be inverted for the model $\rho_I(\mathbf{h}; z_{c2})$ at the high threshold value z_{c2}. A resulting simulated field with this inverted anisotropy has been given in Fig. 4, to be compared with the realization of Fig. 3(b); in both cases the same 16 hard data were used.

CONCLUSIONS

In the Earth Sciences, where data are typically scarce, statistics are often taken towards a goal of data expansion. Models of dependence in space are built from sample statistics and allow expansion of the original information into a detailed picture of the population of interest. The reconstructed maps or numerical models are considered good to the extent that they reproduce certain characteristics of the spatial distribution of the underlying population.

Important characteristics to be honoured include:

- the sample values and qualitative information at their locations,
- patterns of connectivity in space, particularly those of extreme-valued attributes.

Geological mapping is aimed at detecting and enhancing low entropy (organized) patterns of spatial connectivity, whereas traditional interpolation techniques tend to smooth out extreme values, and statistical models lean towards maximum entropy (maximum disorganization) whenever a doubt arises. Yet, strings or patches of connected extreme values might be the most consequential features of the map depending on the intended usage of that map.

Spatial connectivity models defined as probabilities for any two attribute values, separated by a vector $\mathbf{h}$, to jointly exceed any given threshold value, can be inferred and imposed on the reconstitution exercise. The resulting maps will reflect these connectivity models, in addition to honouring local information of any type. These connectivity functions can be in number $K > 1$, and can image patterns of low values different from patterns of median or high values, they can be inferred from hard data or synthesized from soft interpretative information. Thus, they allow the generation of maps with a greater and more valuable information content, maps that better reflect the actual population distribution.

Typically in the Earth Sciences, hard numerical data are scarce and the reconstitution exercise must rely on soft information of various sources. The indicator formalism calls for a common coding of all information available at any location into a series of elementary bits (0 or 1). These bits of information are then processed together, independently of their origins. The results are a series of equiprobable maps honouring all prior information, whether hard or soft. The differences between these equiprobable maps image the prevailing spatial uncertainty. That stochastic imaging of spatial uncertainty can be put to different uses, including the assessment of consequent risks and the need for additional information.

REFERENCES

Abramovitz M. and Stegun I. (1964) *Handbook of mathematical functions*, Dover Publ., 1046 pp.

Alabert F. (1987) Stochastic imaging of spatial distributions using hard and soft information, *MSc thesis*, Branner Earth Sciences Library, Stanford Univ., pp. 169.

Anderson T. (1958) *An introduction to multivariate statistical analysis*, Wiley & Sons, 374 pp.

Devroye L. (1986) *Non-Uniform random variate generation*, Springer Verlag, 843 pp.

Giordano R., Salter S. and Mohanty K. (1985) *The effects of permeability variations on flow in porous media*, *SPE paper 14365*, 60th SPE Annual Conf., Las Vegas.

Goldberger A. (1962) Best linear unbiased prediction in the generalized linear regression model, *J. Am. stat. Assoc.*, **57**, 369–375.

Haralick R.M., Shanmugam K. and Dinstein I. (1973) Textural features for image classification, *IEEE Trans.*, Nov. 1973, 610–621.

Jones D. (1979) *Elementary information theory*, Clarendon Press, Oxford, pp. 137.

Journel A. (1974) Geostatistics for conditional simulation of orebodies, *Econ. Geol.*, **69**, 673–80.

Journel A.G. (1983) Non parametric estimation of spatial distributions, *Math Geol.*, **15**, 445–468.

Journel A. (1986) Constrained interpolation and soft kriging. In: *Proc. 19th APCOM Symp.* Soc. of Min. Eng., Littleton, CO. pp. 15–30.

Journel A. and Huijbregts Ch. (1978) *Mining geostatistics*, Chapter 5, Academic Press, 600 pp.

Sullivan J. (1985) Conditional recovery estimation through probability kriging: Theory and practice. In: *Geostatistics for natural resources characterization* (ed. by G.Verly, M. David, A. Journel and A. Marechel), Reidel, Dordrecht, pp. 365–384.

Suro Perez V. (1988) Indicator kriging based on principal component analysis, *MSc thesis*, Branner Earth Sciences Library, Stanford Univ., 168 pp.

Switzer P. (1977) Estimation of spatial distributions from point sources with application to air pollution measurement, *Bull. int. Stat. Inst.*, **47**, 123–137.

Thomas J. (1986) *Introduction to probability*, Springer-Verlag, 247 pp.

Xiao H. (1985) A description of the behavior of indicator variograms for a bivariate normal distribution, *MSc thesis*, Branner Earth Sciences Library, Stanford Univ., 51 pp.

Received 6th December 1988