INTRODUCTION TO GEOSTATISTICAL THEORY AND
EXAMPLES OF PRACTICAL APPLICATIONS

by

Henning Omre
Norwegian Computing Center
Postboks 335 - Blindern
Oslo 3, Norway

~ 1984

ABSTRACT

A brief introduction to geostatistical modeling, estimation and simulation is given. The main characteristics of these parts of geostatistics are emphasized. Application of the methods to mining and petroleum exploration is discussed.

I. INTRODUCTION

Geostatisticians are mainly conserned with solving practical problems arising in the analysis of spatially correlated variables. Examples of variables are oregrade, depth to geological horizon or airpollution. Dr. G. Matheron, a french mining engineer and mathematician, formalized the geostatistical theory in the early nineteen sixties, see Matheron (1965). His primary aim was to understand and solve problems in the mining industry. The geostatistical theory is based on a stochastic process consept and it is closely connected to the theory of random functions, see Yaglom (1962). Matern (1960) also seems to have inspired Dr. G. Matheron in his work. The field of geostatistics has hardly added anything fundamentally new to the disiplin of probability theory and statistics. Its great contribution, however, is that the general theory has been refrased in order to fit real life problems. In time series, a field closely related to geostatistics, Box and Jenkins (1970) similarly refrased the general theory for practical needs.

In geostatistics the variable under study is named the regionalized variable, $\{z(x);x \varepsilon V\}$. Note that it actually takes deterministic values in every point in the reference domain. The reference variable, x, usually is a vector in two or three dimensions. This regionalized variable is considered as one realization of a random function $\{Z(x);x \varepsilon V\}$. Normally this regionalized variable is known only in a finite number of points, x_i, i=1,N. The set of observations is s: $\{z(x_i);x_i \varepsilon V;i=1,N\}$. The corresponding set of random variables is S: $\{Z(x_i);x_i \varepsilon V;i=1,N\}$.

The main objective of geostatistics is to determine the regionalized variable $\{z(x);x\varepsilon V\}$ from the set of observations, s, and available apriori knowledge about the characteristics of the phenomenum under study.

Several aspects of a geostatistical analysis may be unfamiliar to statisticians:

- the objective is to estimate the regionalized variable, hence a realization of the random function. The parameters in the underlying model have no interest in themselves, and should only be estimated if absolutely necessary.
- regionalized variables from the field of earth sciences tend to have skewed and heavy tailed distribution of values. Hence the well established multi-normal theory has limited applications.
- the available set of observations represents only an extremely small part of the domain about which inference is to be made. The high expences for collecting data usually causes this.
- the observations are not independent, they are spatially correlated. This prohibits the use of most traditional statistical methods. The spatial correlation, however, is necessary if spatial interpolation is to be made.
- in several applications in the field of earth sciences there is a tendency to perform biased sampling. The preferential sampling in either high-value or low-value areas causes problems in the later analysis.
- normally the observations cannot be made exactly at the points, x_i; $i=1,N$, but have to be averages of volums centered at each of these points. The size of these volumes is denoted the support size. The variability of the observations is dependent on the phenomenum under study as well as the support size. The observations in this presentation are assumed to have approximately point support size.

Basic introductions to geostatistical theory and applications can be found in the papers, Huijbregts (1975) and Delfiner and Delhomme (1975) and the books, David (1977) and Clark (1979). Thorough theoretical discussions of the theory are presented in Matheron (1965), Journel and Huijbregts (1978) and Journel (1983).

This paper is divided in two main parts - Geostatistical Methods and - Geostatistical Applications, at the end some - Closing Remarks are added.

II. GEOSTATISTICAL METHODS

A geostatistical analysis may be divided into three parts: modeling, estimation and simulation. The modeling has to take place before either estimation or simulation can be performed.

II.1 Geostatistical Modeling

There exists only one realization, the regionalized variable $\{z(x);x\varepsilon V\}$, of the random function $\{Z(x);x\varepsilon V\}$. This obviously constrains the level of modeling possible. In order to get repeatability, assumptions of spatial stationarity in the random function have to be made. Usually intrinsic stationarity is assumed, this implies:

$$E\{Z(x)\} = m \; ; \text{ all } x\varepsilon V$$

$$E\{\left(Z(x)-Z(x+h)\right)^2\} = 2\cdot\gamma(h) \; ; \text{ all } x\varepsilon V$$

Hence both the expected value and the second moment of the increments are location independent. The $\gamma(h)$ is called the semi-variogram in the geostatistical terminology. In this paper it will be abbreviated to the variogram. In figure 1 the shape of a typical variogram function is presented. Under the somewhat more restrictive second-order stationarity assumption one has:

$$\gamma(h) = C(0) - C(h)$$

where

$$C(h) = \text{cov}\{Z(x),Z(x+h)\} \; ; \text{ all } x\varepsilon V$$

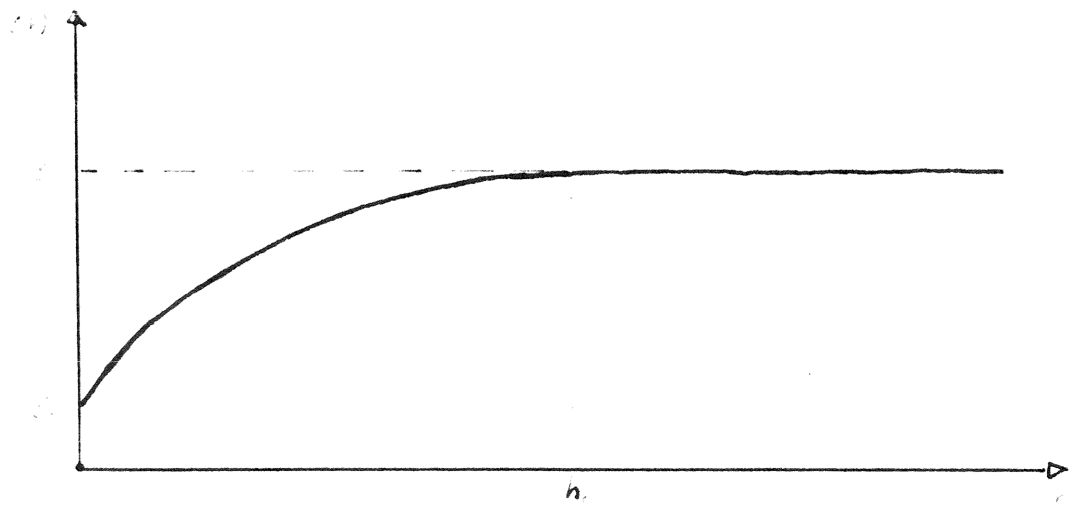


Figure 1. A schematic variogram function

Some properties of the variogram function in figure 1 have important physical interpretations:

- $\gamma(0) = 0$ implies that the regionalized variable is unique; once it is observed in a location the truth is known.

- the value of $\gamma_o = \lim \gamma(h);\ h \to 0$, is called the nugget effect and it reflects the degree of discontinuity in the regionalized variable.

- the distance h_o is called the range and determines the zone of spatial correlation in the regionalized variable. Note that $\mathrm{cov}\{Z(x),Z(x+h)\} = 0$ for all $h > h_o$.

- the value of $\gamma_\infty = \lim \gamma(h);\ h \to \infty$, is called the sill and reflects the variability in the regionalized variable. Note that $\mathrm{Var}\{Z(x)\} = \gamma_\infty$.

- the shape of the variogram function close to the orgin indicates the regularity of the regionalized variable. A parabolic shape characterizes a more smooth variable than a linear shape.

In figure 2 various examples of variogram functions and physical interpretation of them are presented. As can be seen, a large variety of physical characteristics can be modeled by the variogram function. Statistical theory requires the variogram functions to be conditional positive definite, which corresponds to having the covariance function positive definite in the case of second-order stationarity.

Unfortunately the variogram functions of phenomena are not apriori known, but have to be estimated. Both observations of the regionalized variable and prior knowledge of its general characteristics are used in this process. The traditional estimator of the variogram function at distance h', based on the observations, is:

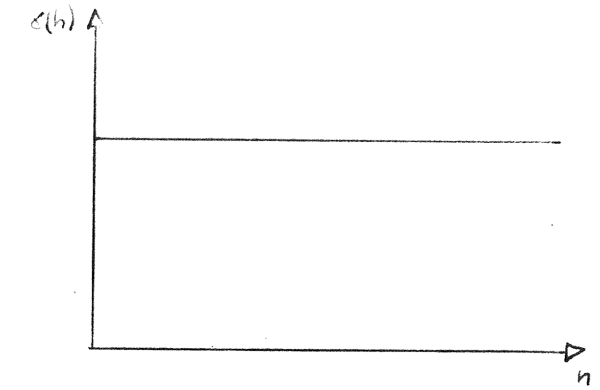
$$\hat{\gamma}(h') = \frac{1}{2 \cdot N_{h'}} \sum_{(i,j)\varepsilon P_{h'}} \left(Z(x_i)-Z(x_j)\right)^2$$

where

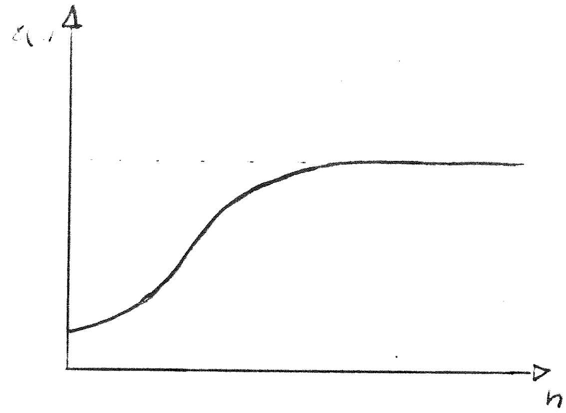
$$P_{h'}:\ \{(i,j) \mid Z(x_i),Z(x_j)\varepsilon S \text{ and } x_i - x_j = h'\}$$

$N_{h'}$ is the number of elements in $P_{h'}$

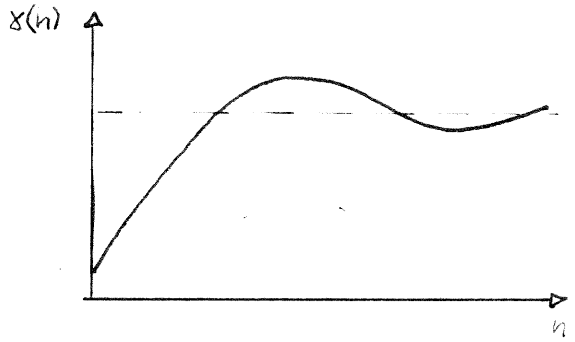
The prior knowledge is introduced through the choice of shape of the variogram function. The final estimate of the variogram function should be verified by confirming the physical interpretations of the properties previously discussed.



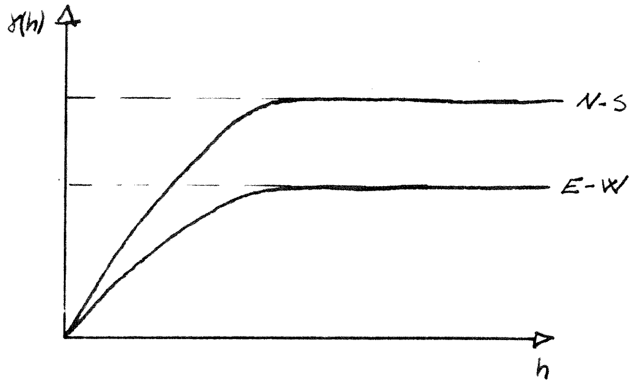
A. Pure nugget effect.
There is no spatial correlation, observations from all locations will be uncorrelated. In this case the geostatistical theory will be identical to ordinary statistics on iid observations, and the consept of spatial interpolation hardly has any meaning.



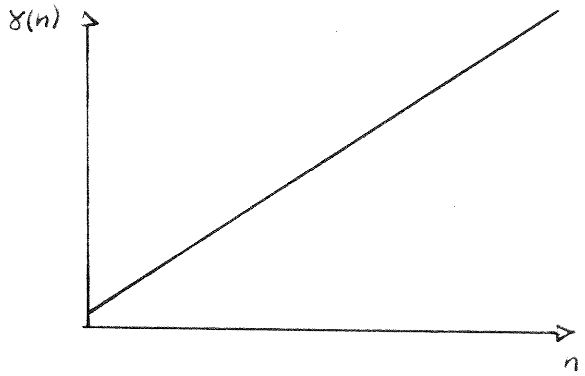
B. Parabolic shape close to the orgin. The phenomenum is relatively smooth. Examples are depth to groundwater, thickness of coal layer.



C. Hole effect. The phenomenum has a periodic component. Examples are sediments on a coastline.



D. Anisotrophy. The correlation structure and the variability are direction dependent. Examples are mineral deposits in areas with tectonic activity.



E. Linear model. The phenomenum is intrinsic stationary, but not second order stationary. Actually, the variability in the variable is infinity.

Figure 2.

Characteristic variogram functions.

So far the point variable $\{z(x);x\varepsilon V\}$ has been discussed. Consider the spatial integral:

$$z_v(x) = \frac{1}{v(x)} \int_{v(x)} z(u)\ du$$

where

$v(x)$ is a volume centered at x .

Ignore the bordereffects and define s_v: $\{z_v(x);x\varepsilon V\}$ and correspondingly S_v: $\{Z_v(x);\ x\varepsilon V\}$. Define the dispersion variance as

$$D^2(v|V) = \frac{1}{V} \int_V E\{(Z_v(u)-Z_V)^2\}du = \bar{\gamma}(V,V)-\bar{\gamma}(v,v)$$

where

$$Z_V = \frac{1}{V} \int_V Z(u)du$$

$$\bar{\gamma}(w,w) = \frac{1}{w^2} \iint_{ww} \gamma(u-u')du\ du' .$$

The dispersion variance defines the variability of the average values of volumes of size v within the domain V. It can be shown, and it makes intuitively sense, that normally $D^2(v_1|V) < D^2(v_2|V)$ if $v_1 > v_2$.

After having justified the stationarity assumptions, the only modeling taking place is the determination of the variogram function. No parametric assumptions are made. The modeling, however, is quite extensive since the variogram function carries information about the continuity, the zone of correlation, eventual periodicity and eventual anisotrophy in the regionalized variable.

II.2 Geostatistical Estimation

Various methods for geostatistical estimation are discussed in the literature, see Journel and Huijbregts (1978), Matheron (1976), Verly (1982) and Journel (1981). In this section only linear estimators will be presented. The ordinary kriging estimator is discussed in some detail, while an outline of the associated techniques, universal kriging, cokriging and block kriging, is given.

The general problem is to estimate $z(x_o)$, where x_o is an arbitrary location in V, from the observations $\{z(x_i);i=1,N\}$. The estimated regionalized variable $\{z^*(x);x\varepsilon V\}$ will get the following properties:

- it is the best linear unbiased estimate in the least square sense, when the variogram of the regionalized variable is known.

- it coinsides with the true regionalized variable in the observation locations, i.e. $z^*(x_i)=z(x_i)$; $i=1,N$.

- its variogram function will not be identical to the variogram function of the true regionalized variable. The former shows less variability and more smoothness than the latter.

In figure 3 an example of these properties is presented.

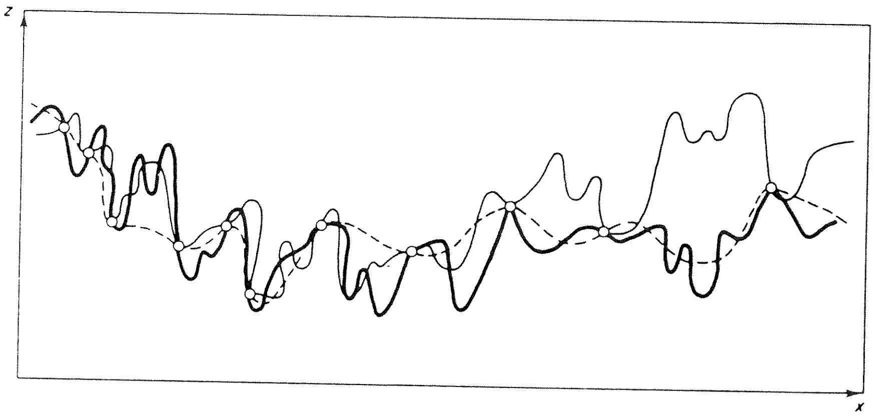


Figure 3. True, simulated and estimated profiles.
———, Reality; ———, Geostatistical simulation; -----, Kriging; o, Observations.
(From Journel and Huijbregts (1978))

Ordinary Kriging.

This estimator is constructed for regionalized variables which fluctuates around an approximately constant level. The hypothesis of intrinsic stationarity reflects this. The form of the ordinary kriging estimator is:

$$Z^*(x_o) = \sum_{i=1}^{N} \alpha_i \cdot Z(x_i)$$

where

α_i; i=1,N are unknown weights to be determined.

Under the specified stationarity assumption, the expected value of the estimator is:

$$E\{Z^*(x_o)\} = \sum_{i=1}^{N} \alpha_i \cdot E\{Z(x_i)\} = m \cdot \sum_{i=1}^{N} \alpha_i$$

Since $E\{Z(x)\} = m$; all $x \varepsilon V$, unbiasedness obviously requires

$$\sum_{i=1}^{N} \alpha_i = 1$$

The optimal set of weights is determined by minimizing the estimation variance under the unbiasedness constraints:

$$\underset{\alpha_i;i=1,N}{\text{Min}} \left\{E\{\left(Z(x_o)-Z^*(x_o)\right)^2\}\right\}$$

$$= \underset{\alpha_i;i=1,N}{\text{Min}} \{2\cdot\sum_{i=1}^{N} \alpha_i \cdot \gamma(x_o-x_i) - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \cdot \alpha_j \cdot \gamma(x_i-x_j)\}$$

$$\sum_{i=1}^{N} \alpha_i = 1$$

This is a system of constrained minimization with a quadratic object function and linear equational constraints. The standard Lagrangean technique will provide a solution to this. The solution will be dependent on the variogram function, $\gamma(h)$. By inserting the estimated variogram function from the modeling stage, an optimal set of weights, $\hat{\alpha}_i$; $i=1,N$, will be obtained. From this the estimate and the corresponding estimation variance can be obtained:

$$z^*(x_o) = \sum_{i=1}^{N} \hat{\alpha}_i \cdot z(x_i)$$

$$\sigma_E^2(x_o) = 2\cdot\sum_{i=1}^{N} \hat{\alpha}_i \cdot \gamma(x_o-x_i) - \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\alpha}_i \cdot \hat{\alpha}_j \cdot \gamma(x_i-x_j)$$

The estimator and estimation variance are dependent on:

- the spatial characteristics of the regionalized variable through its variogram function
- the distance between the location x_o and the location of the observations
- the location of the observations relative to each other

Note that the set of weights will change if the variable in another location shall be estimated.

Universal Kriging.

This estimator is constructed for the case with an obvious trend in the regionalized variable. The underlying assumptions are:

$$E\{Z(x)\} = \sum_{k=1}^{L} \xi_k \cdot f_k(x) \ ; \ \text{all } x\varepsilon V$$

$$E\{(Z(x) - E\{Z(x)\} - Z(x+h) + E\{Z(x+h)\})^2\} = \gamma_R(h) \ ; \ \text{all } x\varepsilon V$$

where

L is a fixed number of terms

$f_k(x)$; $k=1,L$ are known functions of location x

ξ_k; $k=1,L$ are unknown coefficients.

Note that only the form of the trend has to be known, since the coefficients are unknown. An example is $E\{Z(x)\} = \xi_o+\xi_1 x+\xi_2 x^2$.

The form of the universal kriging estimator is

$$Z^*(x_o) = \sum_{i=1}^{N} \alpha_i \cdot Z(x_i)$$

where

α_i, $i=1,N$ are unknown weights to be determined.

The expected value of the estimator is

$$E\{Z^*(x_o)\} = \sum_{i=1}^{N} \alpha_i \cdot E\{Z(x_i)\} = \sum_{i=1}^{N} \alpha_i \sum_{k=1}^{L} \xi_k \cdot f_k(x_i) = \sum_{k=1}^{L} \xi_k \sum_{i=1}^{N} \alpha_i \cdot f_k(x_i)$$

Hence unbiasedness requires:

$$\sum_{i=1}^{N} \alpha_i \cdot f_k(x_i) = f_k(x_o) \ ; \quad k = 1,L$$

Minimizing the estimation variance under the unbiasedness constraints gives:

$$\min_{\alpha_i;i=1,N} \{E\{(Z(x_o)-Z^*(x_o))^2\}\}$$

$$= \min_{\alpha_i;i=1,N} \{2\cdot\sum_{i=1}^{N} \alpha_i\cdot\gamma_R(x_o-x_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\cdot\alpha_j\cdot\gamma_R(x_i-x_j)\}$$

$$\sum_{i=1}^{N} \alpha_i\cdot f_k(x_i) = f_k(x_o) \; ; \quad k = 1,L$$

The standard Lagrangean technique will provide a solution to this minimization system. The set of weights, can be determined when an estimate of the variogram function of the residuals, $\gamma_R(h)$, is estimated. It should be added that this is not trivial because the variogram function and the trend has to be estimated simultaneously.

When the weights are obtained, the estimate of $z(x_o)$ and the corresponding estimation variance can be determined.

Cokriging.

This estimator is used in the multivariate case, i.e. more regionalized variables are defined over the domain V. The bivariate case will be discussed. Define in addition to $\{z(x); x\varepsilon V\}$ the regionalized variable $\{y(x); x\varepsilon V\}$. Let the notational convention for the latter correspond to the one of the former. Introduce an index y if ambiguities occur. Assume that the corresponding random functions are both intrinsic stationary. Define the cross-variogram between the two regionalized variables as:

$$E\{(Z(x)-Z(x+h))(Y(x)-Y(x+h))\} = \gamma_{zy}(h) \; ; \quad \text{all } x\varepsilon V$$

The estimation of the cross-variograms are done equivalently to the estimation of the variograms. Several constraints have to be put on the system of variogram functions to ensure conditional positive definitness.

The cokriging estimator for $z(x_o)$ has the form:

$$Z^*(x_o) = \sum_{i=1}^{N} \alpha_i\cdot Z(x_i) + \sum_{i=1}^{N} \beta_i\cdot Y(x_i)$$

where

α_i, β_i; i=1,N are unknown weights to be determined.

An equivalent estimator for $y(x_o)$ can of course be defined.

The expected value of the estimator is:

$$E\{Z^*(x_o)\} = \sum_{i=1}^{N} \alpha_i \cdot E\{Z(x_i)\} + \sum_{i=1}^{N} \beta_i \cdot E\{Y(x_i)\} = m \sum_{i=1}^{N} \alpha_i + m_y \cdot \sum_{i=1}^{N} \beta_i$$

Unbiasedness requires, since $E\{Z(x)\} = m$; all $x \varepsilon V$:

$$\sum_{i=1}^{N} \alpha_i = 1$$

$$\sum_{i=1}^{N} \beta_i = 0$$

Minimizing the estimation variance under the unbiasedness constraints gives:

$$\min_{\substack{\alpha_i;i=1,N \\ \beta_i;i=1,N}} \{E\{(Z(x_o)-Z^*(x_o))^2\}\}$$

$$= \min_{\substack{\alpha_i;i=1,N \\ \beta_i;i=1,N}} \{2\cdot\sum_{i=1}^{N} \alpha_i \cdot \gamma(x_o-x_i) + 2\cdot\sum_{i=1}^{N} \beta_i \cdot \gamma_{zy}(x_o-x_i)$$

$$- \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \cdot \alpha_j \cdot \gamma(x_i-x_j) - 2\cdot\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \cdot \beta_j \cdot \gamma_{zy}(x_i-x_j)$$

$$- \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_i \cdot \beta_j \cdot \gamma_y(x_i-x_j)\}$$

$$\sum_{i=1}^{N} \alpha_i = 1$$

$$\sum_{j=1}^{N} \beta_i = 0$$

The standard Lagrangean technique will provide a solution to this minimization system. The sets of weights, can be determined when estimates of the variogram functions are available. The theory can easily be extended to the multivariate case.

When the weights are obtained, the estimate of $z(x_o)$ and the corresponding estimation variance can be determined.

Block Kriging.

This estimator is constructed for estimating spatial averages of the regionalized variable. The variable under study is:

$$z_v(x_o) = \frac{1}{v(x_o)} \int_{v(x_o)} z(u)du$$

where $v(x_o)$ is a volume centered in location x_o.

The corresponding random variable is $Z_v(x_o)$.

The estimator may be defined for any of the kriging variants previously discussed. For notational convenience the presentation is made for the ordinary kriging case. Assume intrinsic stationarity and define the estimator form as:

$$Z^*_v(x_o) = \sum_{i=1}^{N} \alpha_i \cdot Z(x_i)$$

where α_i; $i=1,N$ are unknown weights to be determined.

The expected value of the estimator is

$$E\{Z^*_v(x_o)\} = \sum_{i=1}^{N} \alpha_i \cdot E\{Z(x_i)\} = m \cdot \sum_{i=1}^{N} \alpha_i$$

Obviously, unbiasedness requires:

$$\sum_{i=1}^{N} \alpha_i = 1$$

Minimizing the estimation variance under the unbiasedness constraints gives:

$$\underset{\alpha_i;i=1,N}{\text{Min}} \{E\{(Z_v(x_o)-Z_v^*(x_o))^2\}\}$$

$$= \underset{\alpha_i;i=1,N}{\text{Min}} \{2\cdot\sum_{i=1}^{N} \alpha_i\cdot\bar{\gamma}(v(x_o),x_i) - \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\cdot\alpha_j\cdot\gamma(x_i-x_j) - \bar{\gamma}(v(x_o),v(x_o))\}$$

$$\sum_{i=1}^{N} \alpha_i = 1$$

where $\bar{\gamma}(v(x_o),x) = \frac{1}{v(x_o)} \int_{v(x_o)} \gamma(u-x)du$

$$\bar{\gamma}(v(x_o),v(x_o)) = \frac{1}{v(x_o)^2} \int_{v(x_o)} \int_{v(x_o)} \gamma(u-u')dudu'$$

The standard Lagrangean lechnique will provide a solution to this minimization system. The set of weights can be obtained when an estimate of the variogram function is available.

When the weights are obtained, the estimate of $z_v(x_o)$ and the corresponding estimation variance can be determined.

II.3 Geostatistical Simulation

The procedure for geostatistical simulation or conditional simulation, is presented in Journel and Huijbregts (1978). In this section only an outline of the procedure will be given.

Geostatistical simulation was introduced as an alternative to estimation. The objective was to avoid the smoothing tendency in geostatistical estimation. The simulated regionalized variable $\{z_S(x);\ x\epsilon V\}$ resulting from the geostatistical simulation gets the following properties:

- it has the same distribution of values and the same variogram function as the true regionalized variable.
- it coinsides with the true regionalized variable in the observation locations, i.e. $z_s(x_i) = z(x_i);\ i=1,N$.
- it is not unique, since infinitely many regionalized variables have the properties listed above.

In figure 3 an example is presented.

Assume that the random function is strictly stationary to the first order. This implies that the univariate distribution of $Z(x)$ is equal for all $x \varepsilon V$. The procedure of geostatistical simulation can be outlined as follows:

- transform the observations $\{z(x_i); i=1,N\}$ univariately so that the resulting set $\{z_T(x_i); i=1,N\}$ reflects a standard normal distribution.
- estimate the variogram function, $\gamma_T(h)$, of the transformed regionalized variable.
- simulate a regionalized variable $\{z_{SN}(x); x\varepsilon V\}$ with normal distribution of values and variogram function $\gamma_T(h)$.
- to ensure that the simulated regionalized variable passes through the observations define the transformed simulated regionalized variable as

  $$\{z_{ST}(x) = z_T^*(x) + \left[z_{SN}(x)-z_{SN}^*(x)\right] \; ; \; x\varepsilon V\}$$

  where

  $z_T^*(x)$ is the kriging estimate in location x based on $\{z_T(x_i); i=1,N\}$

  $z_{SN}^*(x)$ is the kriging estimate in location x based on $\{z_{SN}(x_i); i=1,N\}$
- obtain $\{z_S(x); x\varepsilon V\}$ by an univariate backtransform of $\{z_{ST}(x); x\varepsilon V\}$ according to step one.

A theory for geostatistical simulation for the multivariate case is also available.

In geostatistical simulation emphasis is put on reproducing the spatial characteristics of the phenomenum, no effort is made to minimize the estimation variance. It can be shown that the estimation variance in an arbitrary location is twice the one obtained by ordinary kriging. The expected value of the simulations is identical to the ordinary kriging estimate.

III. GEOSTATISTICAL APPLICATIONS

The geostatistical methods can be applied in analysis of all kinds of spatial variables. Examples of fields of applications are: mining, petroleum exploration, forestry, soil science, meteorology, sivil engineering, hydrographic mapping and pollution studies. The main characteristics of the application to mining and petroleum exploration will be discussed.

III.1 Application to Mining.

In mining the in situ resource in the deposit and the recoverable mineral reserve are of great interest. The ore-grade is usually directly observed through laboratory analysis of small rock samples from drillholes. The number of observations is often large and the observations are reasonably homogeneously located over the deposit.

The geostatistical modeling is sometimes complicated by skewed and heavytailed distributions of grades in the deposit. Extreme examples are uranium and gold deposits. When an estimate of the variogram is established geostatistical techniques for estimating the in situ resource are available. An estimate of the corresponding estimation variance can also be provided.

The recoverable reserve is dependent on both the mining method and the economical conditions. Selective mining is often used either for financial or technical reasons. Selective mining implies that only ore above a certain cut-off is taken to the mill while the remaining is taken to waste. There are obvious problems connected to predicting the amount of ore that will be recovered and the location of the areas that should be recovered. An additional problem is that the selection is not made on the grade of units of the size of the observations, but on the average grade of larger panels. The panel size may correspond to one truck load, which either is taken to the mill or to waste. The dispersion variance concept explains why the dispersion variance of grades in the observations will be larger than the dispersion variance of average grades in the panels. An example is presented in figure 4A and B. If perfect selection was made on panelsize $(20 \times 20 \times 13)m^3$, at cut-off equal 3% for example, the true average recovered grade can be computed from the truncated distribution to the right of the 3%-mark in figure 4B. If the unit size on which selection is going to be made is ignored, and the average recovered grade is computed from figure 4A, overestimation will take place. The mean of the truncated distribution at the 3%-mark obviously is larger in figure 4A than in figure 4B.

Unfortunately perfect selection cannot be made since knowledge of the true panelaverages, z_v, are not available when the selection takes place. The selection has to be made on estimated averages. Block kriging can provide these, z_v^*. The bivariate properties of (Z_v, Z_v^*) will determine the efficiency in this selection process. In figure 5 this property is presented. If selection is made at cut-off z_o of the estimated panel averages, the proportion under the heavily hatched area in figure 5 will be correctly selected. The proportion under the hatched area below this will be wrongly selected while the proportion under the hatched area to the left will wrongly not be selected. To estimate the expected amount of recovered ore unbiasedly for all cut-off values, the conditional expectation has to be $E\{Z_v | Z_v^*=z\} = z$ for all z.

Geostatistical analysis is the only approach which adresses these prediction problems. The practical consequences of thoroughly understanding this are great, both technically and economically. Geostatistical methods are extensively applied by the mining industry. Journel and Huijbregts (1978) discussed this application in much greater detail.

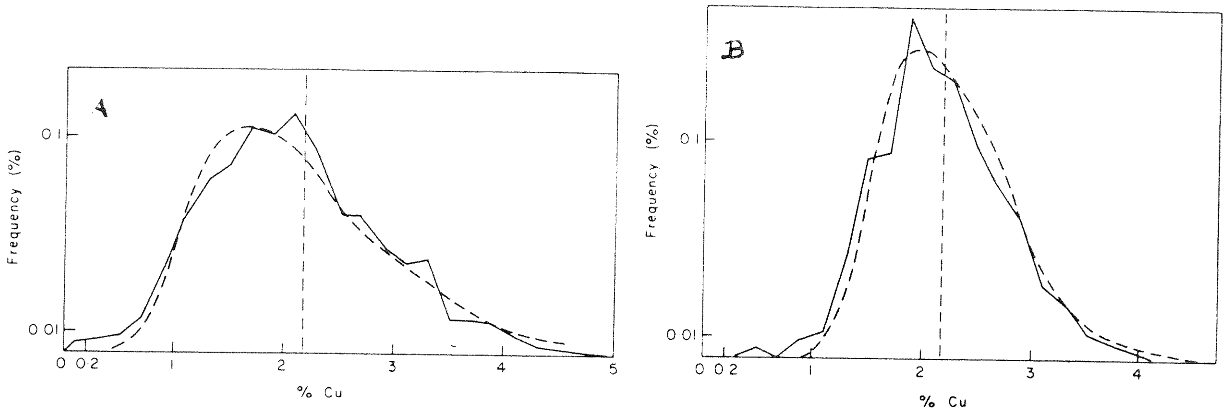


Figure 4. The distribution of grades in a copper deposit.

A. The dispersion of observations with approximately point support; m = 2.12% Cu; s^2 = .94 $(\%Cu)^2$. A log-normal distribution is fitted to the observations.

B. The dispersion of volumes of size $(20x20x13)m^3$; m = 2.16% Cu; s^2 = .35 $(\%Cu)^2$. A log normal distribution is fitted to the data.

The deviation in the estimates of the mean is due to border effects.
(From Journel and Huijbregts (1978))

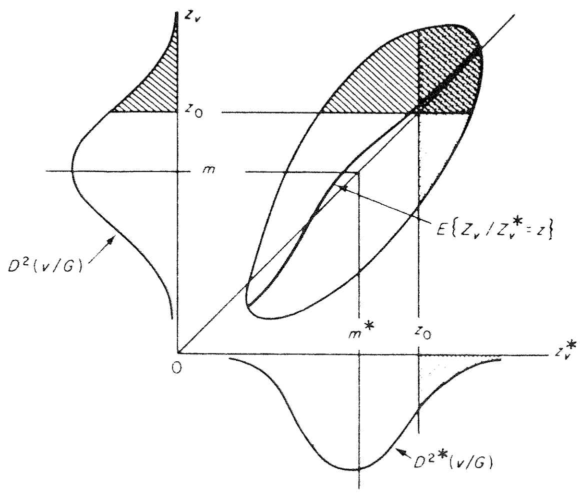


Figure 5. Bivariate distribution of the true and estimated average grades of volume v, Z_v and Z_v^*.
(From Journel and Huijbregts (1978))

III.2 Application to Petroleum Exploration

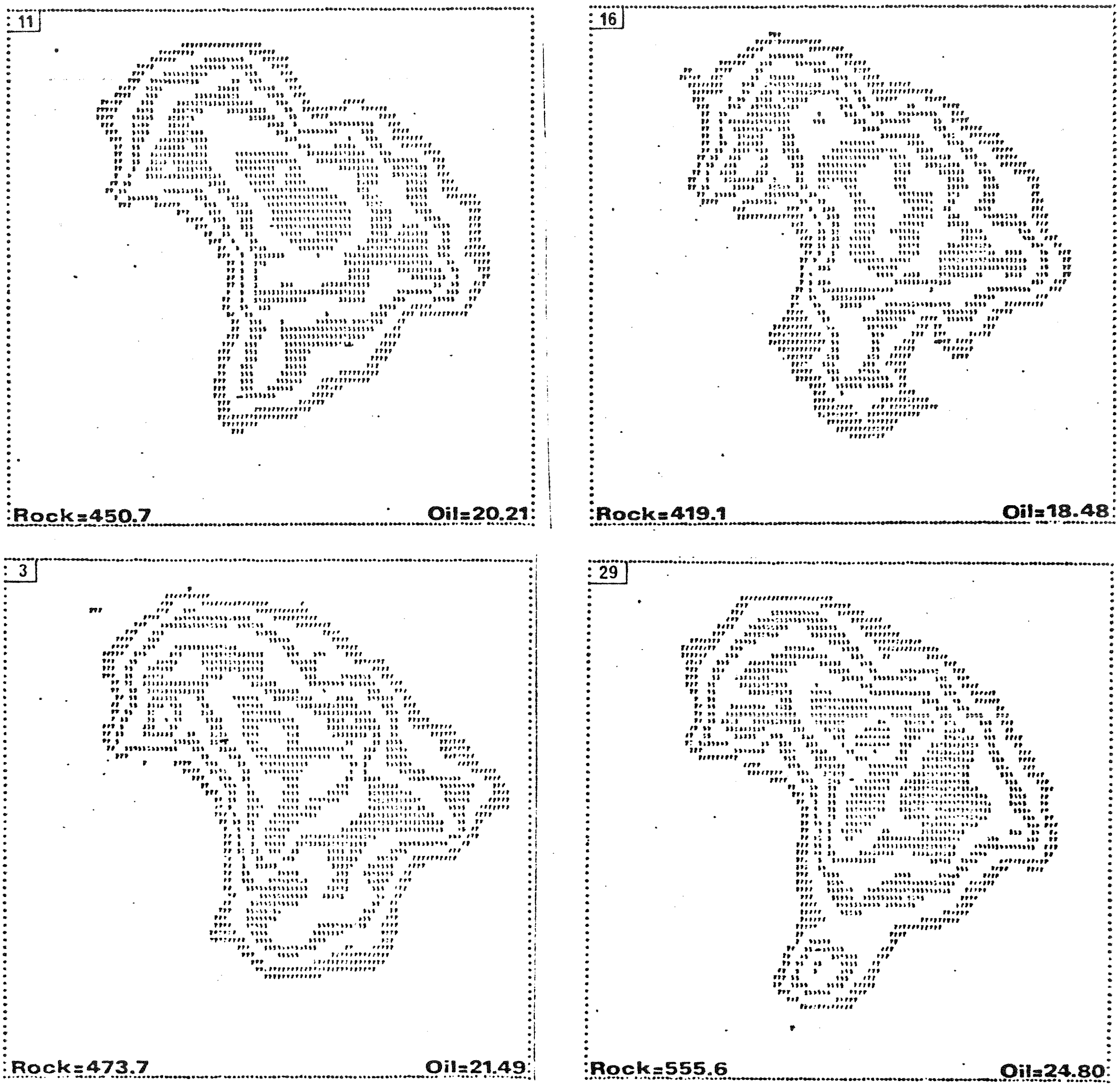
In petroleum exploration it is of great importance to establish a reliable spatial model of the reservoir and its relevant characteristics, in order to estimate the oil and gas in place and the recoverable reserves. This includes estimation of the depth to actual geologic horizons in all positions within the area of interest, and the value of the petrophysical variables in all points in the reservoir. Relevant petrophysical variables are porosity, watersaturation, permeability.

Several of these geophysical and petrophysical variables can only be indirectly observed. Seismic methods and well-logging are important indirect measurement methods. Often these indirect observations actually reflect a composite of the petrophysical variables. Other variables may be directly observed in the wells itself or through core samples. The observations are obtained from either wells or seismic surveys, which implies that there are fairly precise observations in point locations and unprecise geophysical data on a relatively dense net of profiles. The qualitative geological knowledge also is of great importance, particularly in the determination of the geophysical characteristics of the reservoir.

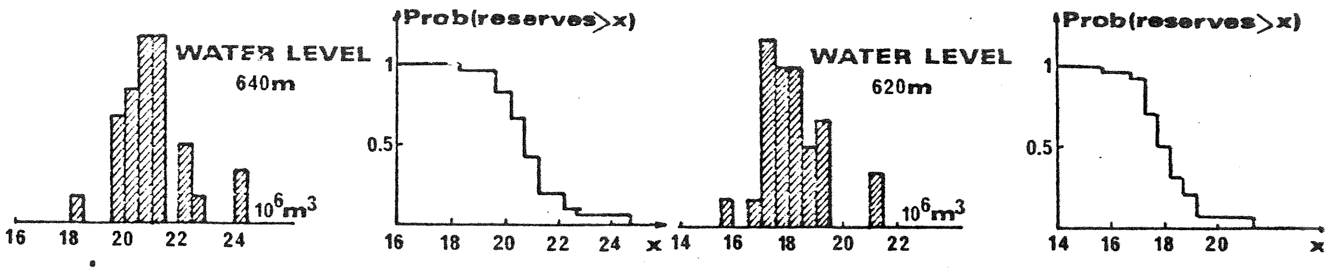
The construction of a spatial model of the reservoir and its characteristics, obviously implies multivariate spatial interpolation from observations on various levels of precision. This fits into the framework of cokriging. Aspects of the qualitative geological knowledge can be incorporated in the geostatistical modeling process. The fact that observations of some of the petrophysical variables can be sparse may complicate the estimation of the variograms. When the spatial model, with corresponding estimation variances, is established, oil and gas in place can be estimated by simple integration. The spatial model also can be used as input to reservoir simulation programs in order to determine the recoverable reserves.

The kriging procedure provides a spatial model of the reservoir and its relevant characteristics which is optimal in the least square sence. Hence the spatial characteristics are not kept. An alternative approach could include geostatistical simulation. Several sets of spatial models of the reservoir and its characteristics can be simulated. For each simulation, oil and gas in place can be computed and a probability distribution of these two variables can be obtained. It is also possible to take each of these simulations through the reservoir simulation to obtain the distribution of the estimates of recoverable reserves. This, however, would require large computer resources.

Delfiner and Chiles (1979) presents a study of this type. They simulated the depth to the top geologic horizon of a reservoir thirty times. Four of the resulting outcomes are presented in figure 6A. Given the porosity and the watersaturation in the reservoir the oil in place was computed for various levels of oil/water contact for each simulation. The distribution of the estimates of the oil in place for two contact levels are presented in figure 6B.



A. Four representative outcomes of the simulation of the top geologic horizon.
Baselevel is 640 m, sections of depth 20 m are interchangably hatched and non-hatched. Rock and oil volume in 10^6 m^3.



B. The distribution of estimates of oil in place. Two possible oil/water contactlevels are considered.

Figure 6. Results from geostatistical simulation.
(From Delfiner and Chiles (1979))

Geostatistical methods have to some extent been used by the petroleum industry. Since geostatistics is the only alternative if spatial multivariate data shall be analysed the methods should have a large potensial in this area of application. The petroleum industry could probably also obtain more reliable estimates for their variables of interest if more effort was put into analysing their data more cautiously. The fact that the expences associated with collecting additional information is huge, should further encourage the use of geostatistical methods. References in this field of application are Haas and Jousselin (1976) and Delfiner and Chiles (1979).

IV. CLOSING REMARKS

There exist alternatives to geostatistics. First, deterministic procedures for spatial interpolation. These provide an estimate of the regionalized variable. They stop short of presenting any precision measure of the estimate; any treatment of multivariate aspects or any thorough discussion of the phenomenum, because no statistical modeling takes place. Second, ordinary statistical methods can handle multivariate aspects. Normally, the spatial correlation in the observations is ignored, hence the most characteristic property of the phenomenum is left out. This leaves geostatistics as the most powerful method when analysing spatial variables.

Geostatistics is an applied branch of statistics. Emphasis is put on solving problems occuring in practical work. In recent years, practical problems have increased in complexity and marginal solutions to them are required. These facts have merged theory and practice. In the field of earth sciences, geostatistics offers a set of theoretically justified procedures for solving problems of large practical importance.

V. REFERENCES

Box; G.E.P. and G.M. Jenkins, 1970: Time Series Analysis - forecasting and control. Holden Day, 353 p.

Clark; I., 1979: Practical Geostatistics. Applied Science Publisher Ltd., London.

David; M., 1977: Geostatistical Ore Reserve Estimation. Elsevier Scientific Publ. Co., 364 p.

Delfiner; P. and J.P. Chiles, 1979: Conditional Simulations: A new Monte Carlo Approach to Probabilistic Evaluation of Hydrocarbon in Place. Specialized Techniques of Geostatistics, Notes from Short Course by Batelle Pacific; NW Lab and Centre de Geostat; École de Mines de Paris; 8.-12.10.79.

Delfiner; P. and J.P. Delhomme, 1975: Optimum Interpolation by Kriging. Proc. NATO ASI for Display and Analysis of Spatial Data; Ed. J.C. Davis and M.J. McCullagh; Wiley. p 96-114.

Haas; A. and C. Jousselin, 1976: Geostatistics in Petroleum Industry. Advanced Geostatistics in the Mining Industry; Ed. M. Guarascio et al., D. Reidel Publ. Co., p 333-347.

Huijbregts; C.J., 1975: Regionalized Variables and Quantitative Analysis of Spatial Data. Proc. NATO ASI for Display and Analysis of Spatial Data; Ed. J.C. Davis and M.J. McCullagh, Wiley, p 38-53.

Journel; A.G., 1981: The Indicator Approach to Estimation of Spatial Distributions. Unpublished Notes; Stanford University, 14 p.

Journel; A.G., 1983: Geostatistics, Simple Tools Applied to difficult Problems. Paper presented at the 50th Iowa Statistical Lab. Anniversary, June 1983.

Journel; A.G. and C.J. Huijbregts, 1978: Mining Geostatistics. Academic Press, 600 p.

Matern; B, 1960: Spatial Variation. Meddelanden från Statens Skogforskningsinstitut, Sweden, Band 49, No 5, pp 144.

Matheron; G., 1965: Les Variables Regionalisées et leur Estimation. Masson & Cie, Paris.

Matheron; G., 1976: A Simple Substitute for Conditional Expectation: The Disjuctive Kriging. Advanced Geostatistics in the Mining Industry; Ed. M. Guarascio et al; D. Reidel Publ. Co, p. 221-236.

Verly; G., 1982: The Multigaussian Approach and its Applications to the Estimation of Local Reserves. Paper at First MGUS Conf. on the Management, Display and Analysis of Geoscience Data; Golden, CO, 27.-29.1.82.

Yaglom; A.M., 1962: An Introduction to the Theory of Stationary Random Functions. Prentice-Hall, Inc.