

BASIC INTRODUCTION TO GEOSTATISTICS

by

Pierre DELFINER

Summer School

Septembre 1979

Cours C-78

CENTRE DE GÉOSTATISTIQUE

35, RUE SAINT-HONORÉ, 77305 FONTAINEBLEAU (France)



**ECOLE DES MINES
DE PARIS**

BASIC INTRODUCTION TO GEOSTATISTICS

=====

C O N T E N T S

<u>FOREWORD</u>	: Problems considered	1
<u>CHAPTER I</u>	: REGIONALIZED VARIABLES AND THEIR VARIOGRAMS	4
1	- A model for Regionalized Variables	4
2	- Some conceptual background	5
3	- The variogram	10
<u>CHAPTER II</u>	: THE PRACTICE OF VARIOGRAM INTERPRETATION, CALCULATION AND MODELING	23
1	- Examples of variogram interpretation	23
2	- How to compute a variogram	28
3	- Exercise	31
4	- Models for the variogram	33
5	- Models for the nugget effect	38
<u>CHAPTER III</u>	: DISPERSION AS A FUNCTION OF BLOCK SIZE	50
1	- The support of a Regionalized Variable	50
2	- Dispersion versus block size - An example	50
3	- Variances of dispersion within a volume V	53
4	- Change of support : regularization	59
5	- Coming back to our example	62
6	- Exercise : regularization by cores	65
<u>CHAPTER IV</u>	: VARIANCES OF EXTENSION AND THEIR APPLICATIONS TO ESTIMATION AND SAMPLING PATTERNS	78
1	- The concept of variance of extension and its relations relationship to the variance of dispersion	78
2	- Elementary variances of extension	81
3	- Approximation principle	84
4	- Sampling patterns	88
5	- Exercise	93

<u>CHAPTER V</u>	:	KRIGING	97
1	-	The purpose of kriging	97
2	-	Derivation of kriging equations	99
3	-	Properties and usage of kriging estimators. Applications to contour mapping.	105
4	-	How to take into account the uncertainty of the data	109
5	-	Applications of kriging to the estimation of hydrocarbon in place	112
6	-	How do kriging estimates compare with reality ? A case study	115
7	-	Variance reduction as a guide to locate a new observation.	123
<u>REFERENCES</u>			126
<u>APPENDIX</u>	-	A recollection of some elementary statistical results.	129

F O R E W O R D

This text is a revised version of an introductory course given to petroleum geologists at the National Iranian Oil Company, in Tehran, 1975. Hence the initial emphasis on petroleum applications. Some examples in other areas have been added afterwards, and it is expected that readers with different backgrounds and interests should have no difficulty finding equivalents in their own fields.

The appendix recollecting basic statistical results is also a reminder from the initial course. It has been left in for its possible usefulness as a formulary, especially concerning results on lognormal variables.

PROBLEMS CONSIDERED

Geostatistics is a methodology of resources evaluation that has been used for more than fifteen years by mining Companies. Its application to the Petroleum Industry is rather recent however, but it has proved to be a very powerful and flexible tool. Some Companies are now using Geostatistics on a routine basis and others who do not, are showing a strong interest in the subject.

Here are some possible applications of Geostatistics :

1 - Gridding and Contour Mapping of Seismic Variables.

The gridding operation consists of computing the value of a parameter at the nodes of a regular grid. From the grid a contour map is drawn. Seismic data yield :

- . time contour maps (isochrons)
- . depth contour maps (isobaths)
- . velocity contour map

2 - Determination of the Reservoir Top Depth using both Seismic and Well Data.

Depths measured at wells do not in general coincide with the information from seismic. Therefore it is necessary to make these two sets of data consistent and to use them simultaneously. Also, when the number of wells is small, we wish to utilize the gradients of the depth at the wells where a dip-meter logging has been performed.

3 - Interpolation of Reservoir Parameters between Wells.

Thicknesses, porosities, fluid saturations, permeabilities, etc... are measured at wells. We wish to interpolate these

values between the wells, getting isopach maps, isoporosity maps, etc.. It is also of interest to compute these maps using only some of the wells to see what additional information has been brought by the other wells.

4 - Error Maps.

No method of estimation can ever restore exactly the true unknown value. There is always an error. Moreover, an estimate is almost meaningless if we have no idea of the error involved. Is the error on the thickness 5 m, 20 m, 100 m? Is the porosity known with an accuracy of 1%, 5%, 10%? The technique used (kri-
ging) provides a measure of the error, which itself can be mapped. The two documents, map and error map, should always be consulted simultaneously.

5 - Estimation of Reserves in Place by Blocks.

The various grids : porosity, saturation, thickness, can be combined to provide an estimate of the hydrocarbon reserves block by block.

6 - Estimation of Reserves by Layers.

In reservoirs where the transition zone is important, it is necessary to divide the field into horizontal layers and carry out the estimation layer by layer.

The partitioning of the reservoir into layers according to lithologic types may also be recommended in order to work on homogeneous quantities, thereby reducing the risk of error. Geostatistics can provide guidance for the definition of layers.

7 - Estimation of Total Reserves in Place.

There the result consists mainly of one figure : the estimate of the total hydrocarbon content of the reservoir. Of course, if we can, an upper and lower bound to these reserves is most welcome.

8 - Simulation of Reservoir Limits and of Reserves.

Should we want more than just an estimate of the reserves, that is a probability distribution of these reserves, we have to simulate the reservoir boundaries. The technique called "conditional simulation" enables us to get several possible versions of the top surface of the reservoir. All these versions are constrained to go through the exact points known at the wells and to show the same spatial variability as the true top surface.

By simulating the water level depth between two limits we can get a histogram of the reservoir volume and from that of its reserves.

9 - Input Preparation for Reservoir Simulation Models.

The various grids computed in 5 are the basic input for a dynamic study of the reservoir during the production phase. As a future application it seems that the error maps can be used for the calibration of the model.

10 - Prediction and Guidance about Future Wells.

Depth, porosity, saturation, etc..can be estimated for any proposed location of a new well. It is also possible to study the influence of the number and the location of wells on the precision of reserves evaluation. This study can be done without actually drilling the wells.

11 - Assessment of the Variability of Reservoir Parameters.

This task put here at the end of the list, in fact comes first in any geostatistical treatment. There is a synthetic function, named the "variogram", which provides a quantitative understanding of the spatial behavior of the studied variables and is the basic mathematical tool of estimation procedures.

- CHAPTER I -

REGIONALIZED VARIABLES AND THEIR VARIOGRAMS

1 - A MODEL FOR REGIONALIZED VARIABLES.

The available information on a reservoir is fragmentary. If we want to draw conclusions on quantities that have not been measured, we need a model. Several can be considered.

The simplest one is the commonly used zonation procedure, in which the reservoir is divided (more or less arbitrarily) into a small number of zones in each of which the parameters are treated as uniform. While this scheme appeals for its simplicity, it may be grossly in error with respect to the actual behavior of reservoir properties.

Another approach is to fit an appropriate mathematical function to the surface of parameter values. Such techniques of exact interpolation are employed to produce seismic contour maps. The implicit assumption though, is that the charted surface is smooth enough to be modeled, at least locally, by a nice simple function. Again, this may be in contradiction with the real phenomenon.

Variables like porosity, water saturation and most of all permeability, tend to display considerable small scale scatter. Their spatial variations are so complex that they do not seem to be ever amenable to a representation in terms of ordinary mathematical functions. If we could draw them, the curves would look like Figure 1, with "jigsaw teeth" or unpredictable jumps. Contrary to the over-optimistic view of exact interpolation methods, it now seems that no interpolation is possible.

The truth lies in between. Most of the time, a closer inspection reveals that the variables are not completely random. Neighboring points seem to be related by a complex set of correlations : on the whole there are zones where the values tend to be high, or low ; there can be systematic patterns, clusters, orientation

effects, etc... The term "Regionalized Variable" (abbreviation Re.V.) was chosen by G. Matheron to emphasize the particular nature of these variables, which combine two apparently contradictory aspects :

- i) a random aspect, accounting for local irregularities
- ii) a structural aspect which reflects the overall features of the phenomenon.

The prevalent probabilistic model assumes that all the randomness is due to "errors" (or "noise"), and all the structure to a functional relationship called the "trend". In other words, if noise were absent the phenomenon would lend itself to exact interpolation. To identify this alleged underlying trend several least squares fits are used (trend surface analysis). However, there are cases where it is not suitable to put all the structures in the trend. For one thing, there can be no trend at all, or more precisely, the trend can assume a constant value throughout the field (e.g. porosity in a homogeneous layer). Also, even when a trend is indeed there and the fit is good, there can well remain structural information in the residuals.

A better model for reality seems to introduce randomness in terms of fluctuations around a fixed surface, which we shall call the drift - discarding the word "trend" to avoid any confusion with the previous model. Fluctuations are not errors but full features of the studied phenomenon, having a structure of their own. It will be the first task of the Geostatistician to identify these structural properties. Having done so, he will be able to solve various problems of estimation of a Re. V. from fragmentary sample data.

2 - SOME CONCEPTUAL BACKGROUND.

1 - Random Functions.

At each data point x_i the observed value of a certain property is considered as the outcome $z(x_i)$ of a random variable

$Z(x_i)$ - whose mean is the drift $m(x_i)$. At points x when no measurements are available the values $z(x)$ are unknown but well defined ; they can also be viewed as outcomes of corresponding random variables $Z(x)$. Mathematically, the family of all such random variables $Z(x)$ is called a random function. (Synonyms : stochastic process, random field). The function $z(x)$ is called a realization of the random function. A random function bears the same relationship with one of its realizations as a random variable with the numerical outcome (e.g. $x = 34$) of a single trial. Simply, instead of drawing from a sample space of numbers we draw here from a sample space of functions. A single drawing determines at once all the values assumed by the function at all points (Fig. 2).

A random function is characterized by its finite dimensional distributions, i.e. the joint distributions of any set of variables $Z(x_1), Z(x_2), \dots, Z(x_k)$, for all k and all points x_1, \dots, x_k . Of course our probabilistic model would remain pure conjecture if we were not able to reconstitute these distributions, or at least some of their moments. This is the problem of statistical inference. In general, when only a single realization is available, statistical inference requires further assumptions. (In the same fashion what could we say on the distribution of X on the basis of $x = 34$ only?). These additional hypotheses reduce the number of "parameters" on which the random function depends. The whole point is to introduce the really minimum amount of hypotheses so as to cover the broadest range of practical situations.

2 - Stationary and Intrinsic Hypotheses.

The usual hypothesis in random function theory is that of stationarity, i.e. the invariance of the law under translations. A stationary random function is homogeneous, self-repeating in space, and this gives an opportunity for statistical inference. Stationarity in "the weak sense" is when only the first two moments of the law are invariant under translations :

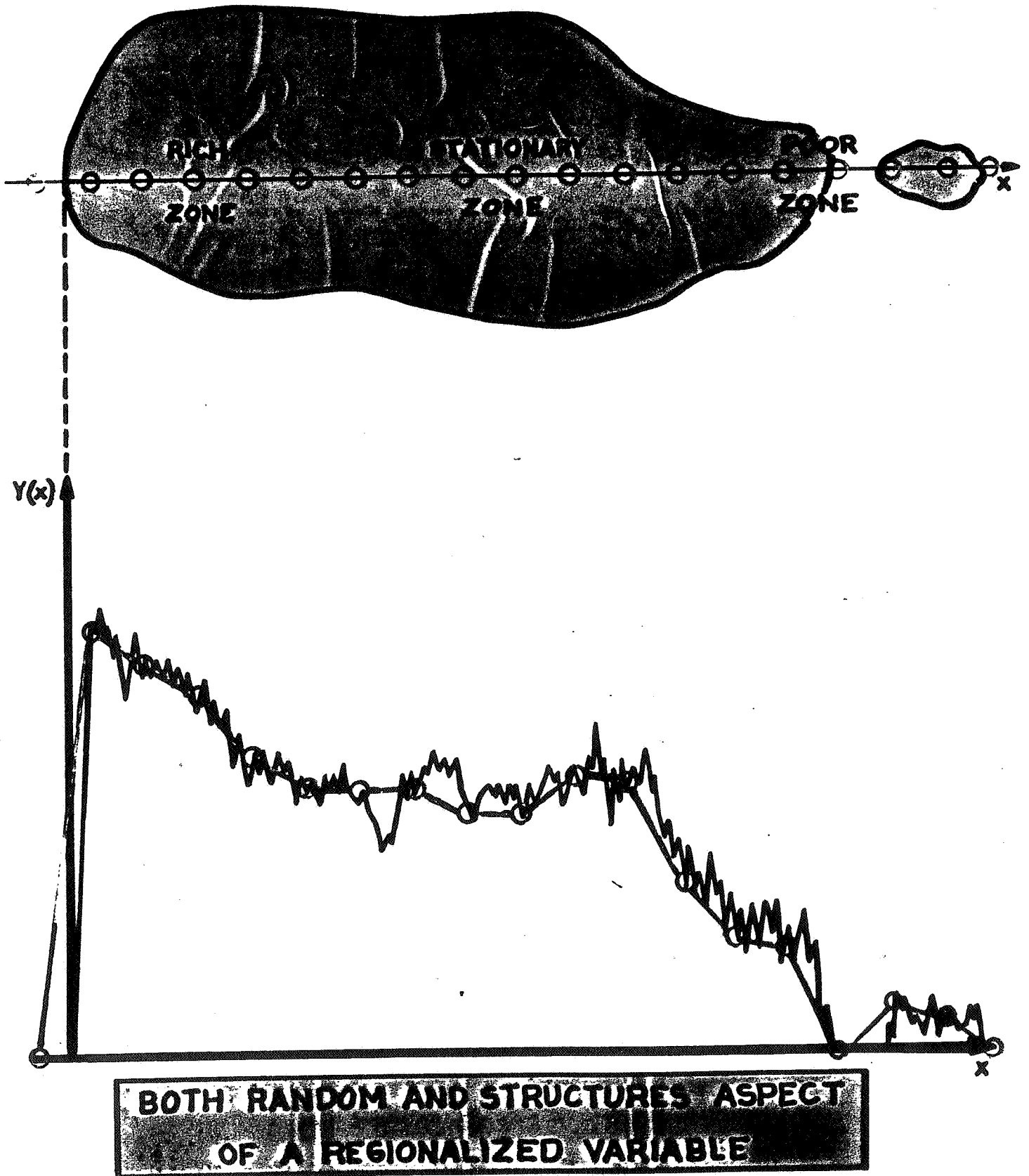
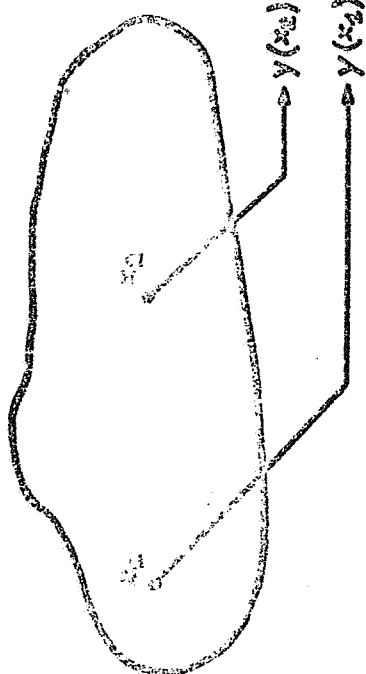


Fig. 1



RANDOM VARIABLE

RANDOM FUNCTION
 $Y(x) = [Y(x_1), Y(x_2), \dots]$

$y(x) = [y(x_1), y(x_2), \dots]$ REALIZATION OF

SPATIAL CORRELATION
 FUNCTION OF \vec{r}

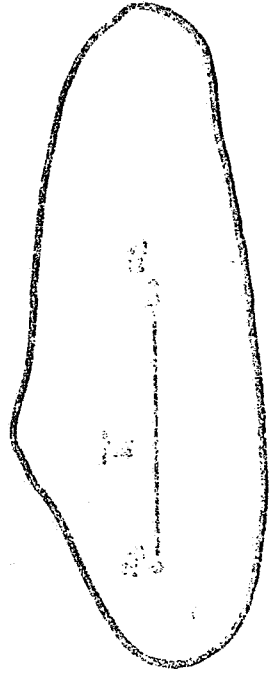
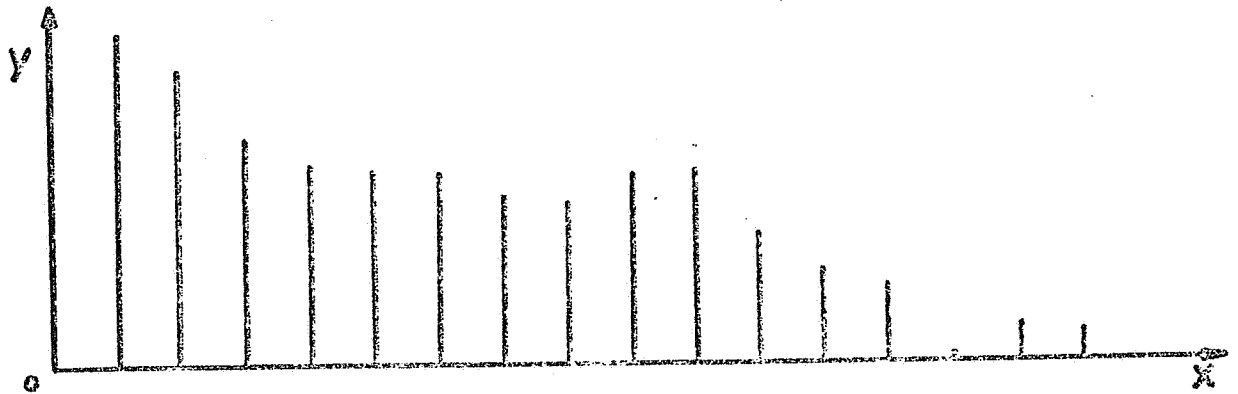


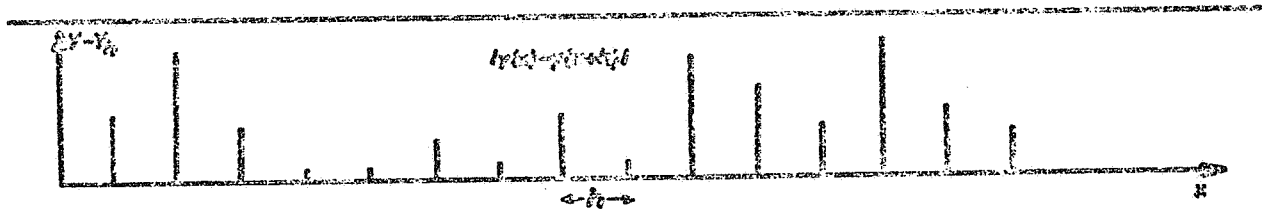
Fig. 2



$y(x_1), \dots, y(x_K) = K$ realizations of K different random variables

STATIONARITY HYPOTHESIS

$y(x_1), \dots, y(x_K) = K$ realizations of 1 unique random variable



INTRINSIC HYPOTHESIS

$(y(x_2) - y(x_1)), \dots, (y(x_K) - y(x_{K-1})) = K-1$ realizations of 1 unique random increment

Fig. 3

- i) the expectation of the function Z at any point x is a constant m : $E(Z(x)) = m(x) = m$ independent of x .
- ii) the covariance function between any pair of points x and $x+h$ is independent of the point x ; it depends on the vector h only : $E[(Z(x) - m)(Z(x+h) - m)] = C(h)$.

N. Wiener's well known optimal filtering method is based upon weak stationarity.

However in practice, it often happens that these assumptions are not satisfied. Clearly when the phenomenon shows a systematic trend it cannot be assumed that the mean is a constant. For example the depth of a reservoir top is systematically increasing towards the periphery of the reservoir. We shall see later how "Universal Kriging" takes trends into account. For the moment we consider cases for which the first hypothesis holds (a constant mean) but not necessarily the second one. On both theoretical and practical grounds it is convenient to assume then that only the increments of the function are stationary in the weak sense. This is called the intrinsic hypothesis (cf. Fig. 3). For any vector h the increment $Z(x+h) - Z(x)$ has an expectation and a variance which are independent of the point x :

$$\begin{cases} E[Z(x+h) - Z(x)] = 0 \\ \text{Var}[Z(x+h) - Z(x)] = 2 \gamma(h) \end{cases} \quad \begin{array}{l} \text{Intrinsic hypotheses} \\ \text{with a constant mean} \end{array}$$

The function $\gamma(h)$ is called the semi-variogram (we say variogram for short). The variogram is the basic tool for structural interpretation as well as for estimation and we shall dwell some time on this subject.

3 - THE VARIOGRAM.

The variogram of an intrinsic random function is by definition :

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x+h) - Z(x)]$$

As it has been assumed that $E(Z(x+h) - Z(x)) = 0$, $\gamma(h)$ is as well the mean square value of the difference $Z(x+h) - Z(x)$.

$$\gamma(h) = \frac{1}{2} E(Z(x+h) - Z(x))^2$$

In practice the following formula can be used to compute the experimental variogram from the available data :

$$\hat{\gamma}(h) = \frac{1}{2} \sum_{i=1}^N [Z(x_i+h) - Z(x_i)]^2$$

(We shall come back to the inference of the variogram later on). The x and $x+h$ refer to data points in an n -dimensional space. ($n = 1, 2$ or 3). For example when $n = 2$ (plane) x denotes the point with coordinates (x_1, x_2) and h is a vector with coordinates (h_1, h_2) . Therefore in two-dimensional space the variogram is a function of the two variables h_1 and h_2 ; considering polar coordinates, γ is a function of the modulus $|h|$ of h and of the polar angle α :

$$\gamma(h) = \gamma(h_1, h_2) = \gamma(|h|, \alpha)$$

For a fixed α the variogram indicates how different the values become as the distance between sample points increases. When computed for different α these variograms disclose the directional features of the phenomenon.

The graph of $\gamma(h)$ plotted versus h , generally presents the following behavior :

- it starts at 0 (for $h = 0$, $Z(x) = Z(x+h)$)
- it increases with h
- it goes on increasing or becomes stabilized about a certain level.

The properties of the variogram as a structure depictor are summarized in Fig. 4 and we now review them separately.

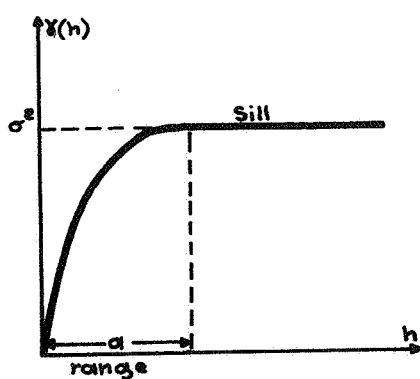
1 - Range and Zone of Influence.

The more or less rapid increase of the variogram reflects the rate of deterioration of the influence of the sample over the more and more distant points of the field. When the variogram reaches a limiting value (sometimes called a "sill") it means that there is a distance beyond which $Z(x)$ and $Z(x+h)$ are without correlation. This distance $a(\alpha)$ is called the range.

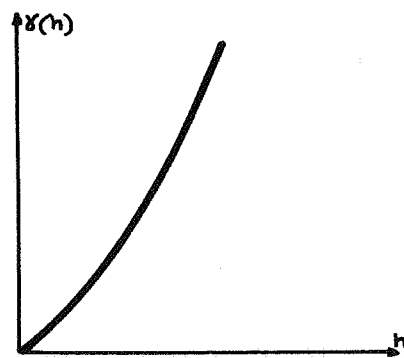
The range gives a precise significance to the traditional concept of "zone of influence" of a sample. The range of course depends on the direction and there can be several ranges reflecting different scales of structures (intermeshed structures). The limiting value itself is exactly the variance of the population of sample points. Indeed when the correlation between $Z(x)$ and $Z(x+h)$ vanishes we have :

$$\begin{aligned}\gamma(h) &= \frac{1}{2} \text{Var}(Z(x+h) - Z(x)) = \frac{1}{2} [\text{Var}(Z(x+h)) + \text{Var}(Z(x))] = \\ &= \frac{2 \sigma^2}{2} = \sigma^2\end{aligned}$$

Not all variograms reach a plateau. Unlike the usual covariance function variograms can well be unbounded (Fig. 5).



Bounded variogram



Unbounded variogram

Fig. 5

However, when the variogram is bounded it is easy to see that it relates to the usual covariance function by :

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

2 - Behavior near the Origin.

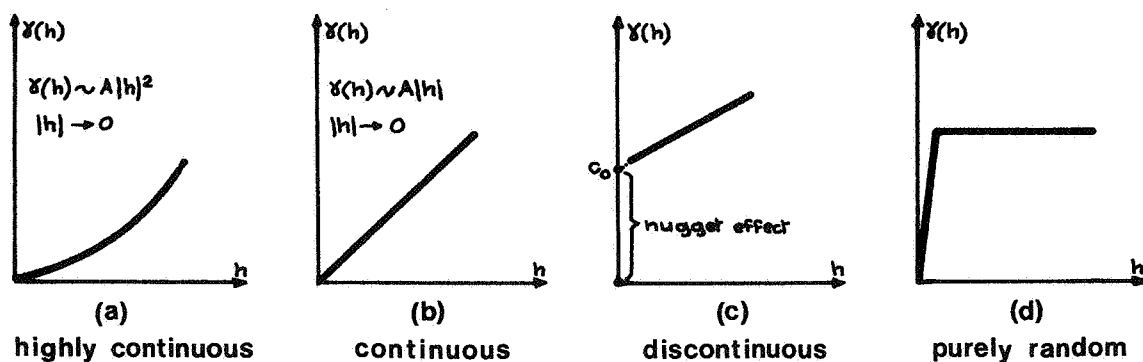


Fig. 6

We just examined the behavior of the variogram at long distances. But it is also most interesting to consider what happens near the origin because this is related to the continuity and spatial regularity of the Re.V. In Fig. 6 we can see four typical shapes :

- (a) a parabolic shape - this pertains to a highly continuous and even differentiable Re.V. (Ex. : reservoir depths). A parabolic behavior can also be associated with the presence of a drift.
- (b) a linear shape - The Re.V. is continuous (in the mean square sense) but not differentiable, and thus less regular than in (a) (Ex. : porosities).
- (c) a discontinuity at 0 : nugget effect - when h tends to 0 $\gamma(h)$ does not tend to 0. This means that the variable is not even continuous in the mean square sense, and thus highly irregular (Ex. : seismic velocities, permeabilities).

The origin of the term is the following : gold ore is often discovered in the form of nuggets, i.e. pebbles

of pure gold. Consequently the ore grade varies discontinuously from outside to inside the nugget. It is convenient to call this small scale variability "nugget effect" even when the effect has other causes. In general the nugget effect is due either :

- i) to a microstructure, i.e. a component of the phenomenon whose range is smaller than the smallest lag for which the variogram is computed.
- ii) to measurement errors
- iii) to location errors.

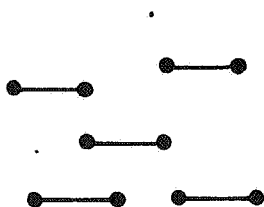
In the absence of short-distance sampling it is impossible to ascertain from the variogram alone which is the right interpretation. There may be several causes involved, especially i) and ii). Physical side information will be essential at the modeling stage. This aspect is developed further in Chapter II.

(d) a flat curve : pure randomness - $Z(x)$ and $Z(x+h)$ are uncorrelated for any two distinct points, regardless of how close they are. This is the limiting case of the total absence of a structure.

In the trend surface analysis model the error term is assumed to have such a flat variogram.

3 - Anisotropies.

As the argument h of the variogram is a vector, the variogram may be computed along several directions of space. It suffices to average only squared deviations of pairs of points which have a definite orientation, e.g. North-South, East-West.



Some pairs in E-W direction



Other pairs in N-S direction

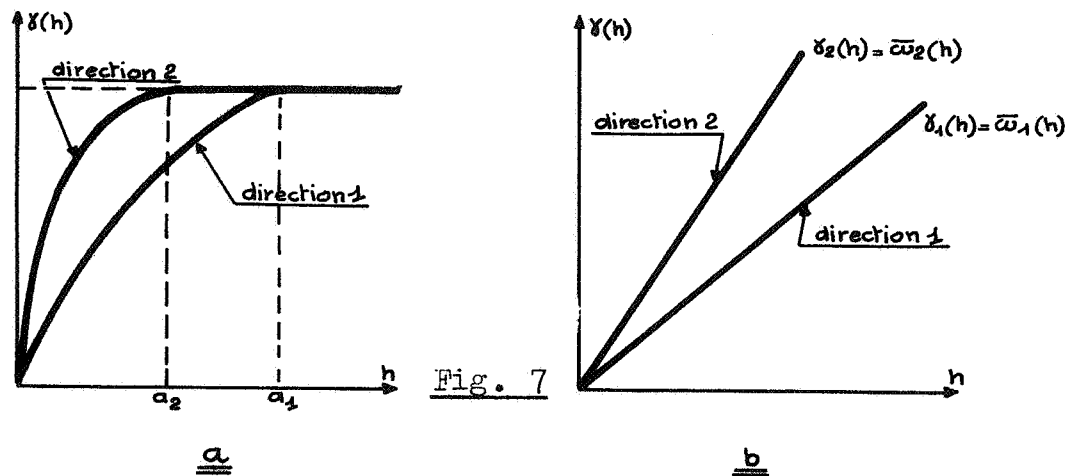
When $\gamma(h)$ does not vary with the direction of space it is said to be isotropic. Then $\gamma(h)$ is a function of the modulus of the vector h , i.e. the distance between sample points :

$$\gamma(h) = \gamma(r) \quad \text{with} \quad \begin{aligned} r &= \sqrt{h_1^2 + h_2^2 + h_3^2} && \text{in 3-D} \\ r &= \sqrt{h_1^2 + h_2^2} && \text{in 2-D} \end{aligned}$$

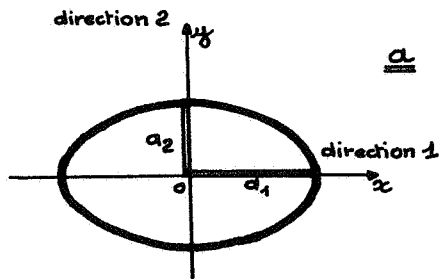
Departures from isotropy can be classified into two categories :

(a) Elliptic anisotropy (also called "geometrical" anisotropy)

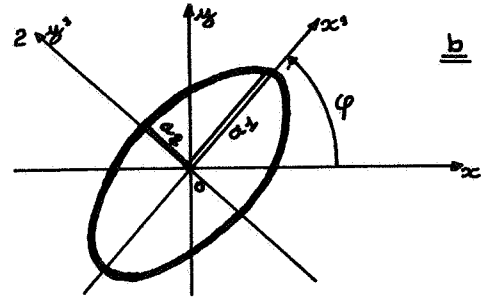
This is the case where **anisotropy may be corrected by an affine transformation of the coordinates**. Typical situations are shown in Fig. 7



In Fig. 7a the variograms have the same sill in all directions but different ranges ; in Fig. 7b variograms are linear but have different slopes. We can draw a diagram showing the variation of the range in case (a) or of the slope in case (b). If the curve is an ellipse (in 2-D) we are in a case of **elliptic anisotropy**. Indeed by a simple affinity on the coordinates it is possible to transform the ellipse into a circle and restore isotropy.



Main axes of anisotropy
coïncide with coordinate
axes



Along O_x and O_y the variograms
are apparently isotropic

Fig. 8

This is particularly simple if the main axes of the ellipse coïncide with the coordinate axes (Fig. 8(a)). Then if γ_1 is the 1-D equation of the variogram along the direction 1, the anisotropy corrected variogram is of the form :

$$\gamma(h) = \gamma_1 \left(\sqrt{(x_1 - x_2)^2 + k^2(y_1 - y_2)^2} \right)$$

where k is the anisotropy ratio, namely :

$$k = \frac{\text{range } a_1}{\text{range } a_2} \quad \text{or} \quad k = \frac{\text{slope } \omega_2}{\text{slope } \omega_1}$$

But we should be careful to compute the variogram in more than two perpendicular directions since we could miss the main axes of anisotropy and get an apparently isotropic variogram (Fig. 8b). In the case when the main axes of the ellipse do not coincide with the coordinate axes, the formula gets a little more complicated since it involves the angle φ between the x -axis and the main axis Ox' of the ellipse. We get

$$\begin{aligned} \gamma(h) &= \gamma_1 \left(\sqrt{h'Ah} \right) \\ h'Ah &= [(x_1 - x_2) \cos \varphi + (y_1 - y_2) \sin \varphi]^2 + \\ &\quad k^2 [(y_1 - y_2) \cos \varphi - (x_1 - x_2) \sin \varphi]^2 \end{aligned}$$

(b) Stratified anisotropy.

There are more complex types of anisotropies. For example in the 3-D space the vertical direction usually plays a particular role : there are more variations across the strata than within the strata. A common model is to decompose the variogram into 2 terms :

i) one being an isotropic variogram

$$\gamma_0(h) = \gamma_0(h_1, h_2, h_3) = \gamma_0(\sqrt{h_1^2 + h_2^2 + h_3^2})$$

ii) the other depending only on the vertical component

$$\gamma_1(h_3)$$

$$\gamma(h) = \gamma_0(h) + \gamma_1(h_3)$$

4 - Presence of a Drift.

Theory shows that for large lags h variograms have to grow more slowly than a parabola. More specifically :

$$\frac{\gamma(h)}{h^2} \rightarrow 0 \quad \text{as } h \rightarrow \infty$$

In practice however it happens that variograms seem to grow as fast or even faster than h^2 . This indicates the presence of a drift.

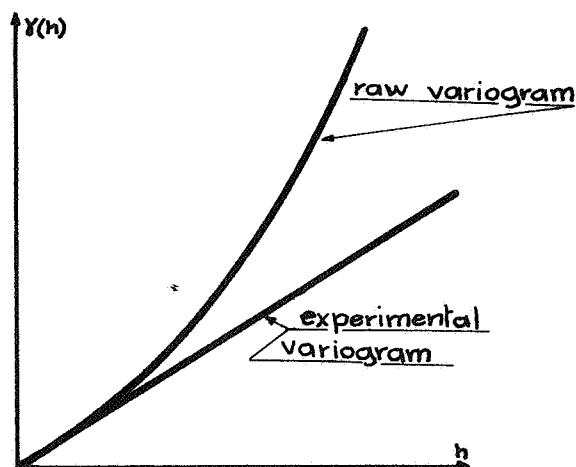


Fig. 9 - Effect of linear drift on variogram.

Indeed the experimental variogram we compute by taking

$$\hat{\gamma}(h) = \frac{1}{2N} \sum_{i=1}^N (Z(x_i+h) - Z(x_i))^2$$

is an estimate of $\frac{1}{2} E[Z(x+h)-Z(x)]^2$, which we call the raw variogram $\gamma_B(h)$, as opposed to the true variogram $\gamma(h)$, also called the underlying variogram. The two coincide only if the increments have zero mean. Otherwise from

$$\underbrace{E[Z(x+h)-Z(x)]^2}_{\text{mean square}} = \underbrace{\text{Var}[Z(x+h)-Z(x)]}_{\text{variance}} + \underbrace{\{E[Z(x+h)-Z(x)]\}^2}_{(\text{bias})^2}$$

it follows that

$$\gamma_B(h) = \frac{1}{2} E[Z(x+h)-Z(x)]^2 = \gamma(h) + \frac{1}{2} [m(x+h)-m(x)]^2$$

So, in case of a drift the empirical variogram $\hat{\gamma}(h)$, which estimates $\gamma_B(h)$, is always an upward biased estimator of the variogram.

To fix the ideas, consider first a linear drift in 1-D.

$$m(x+h)-m(x) = [a(x+h)+b] - [a x+b] = a h$$

and

$$\gamma_B(h) = \gamma(h) + \frac{1}{2} a^2 h^2$$

A term $\frac{1}{2} a^2 h^2$ (a parabola) is added to the variogram. For short distances this term is small but for long ones it is practically the leading term, and we observe that wild growth of Fig. 9. The same conclusions (the added parabola) holds for a linear drift in 2-D or 3-D except that $a h$ becomes an inner product $a_1 h_1 + a_2 h_2 + \dots$

The effect of a drift other than linear is more complex because $m(x+h)-m(x)$ then depends on x , and so does the raw variogram (increments are no longer stationary). The bias becomes a spatial average of $\frac{1}{2} [m(x+h)-m(x)]^2$ over the domain of x . It may distort

the variogram in nasty ways. For example a dome-shaped drift will result in a dome-shaped variogram with a maximum at about half the span of the dome. Geometrically it is clear that the maximum variation occurs when one point is on top of the dome and the other on the periphery, and that the difference goes down at longer distance since the two points overstep the dome.

Unless stated otherwise, it will be considered in the rest of this text that the mean $m(x)$ is a constant.

5 - Other features

Between the origin and infinity the behavior of the variogram displays various features of the Re.V. Some typical cases are :

- nested structures : superposition of different scales of variation. Such structures tend to appear on variograms computed along wells.
- periodicities : variograms, like covariances, can depict periods. But one has to make sure these are real, especially with spatial variables. Time phenomena actually do have periods. Variables of nature, and by consequence human activities, are influence at least by the fundamental daily and yearly cycles. No such clear periods generally exist in space. Let us quote Matérn (1960, pp. 67-68) :

Milne also studied 20 complete enumerations (yields of agricultural crops, horticultural and orchard crops; number of larvae, eggs or adults of several kinds of beetles) and discovered no sign of periodicity. Nor could he find "reasonable grounds for expecting spatial periodicity anywhere on this earth except where man himself, either directly or indirectly, has imposed periodic conditions sufficiently accurate to override the natural environmental irregularity". Milne also was of the opinion that "man-made spatial periodicity will nearly always be suspected either from external signs or past history". The surface drainage of irrigation channels and the equal spacing of planted trees were mentioned as examples.

Milne's formulation must probably be considered as an overstatement. It is not difficult to find spatial "quasi-periodicity": the reflection of the yearly cycle in sedimentary rocks and soils, the regular patterns in organic tissues (cf. Ladell 1959), not to speak of sea-waves and the effect of their action on the shore. However, regarding the spatial variation that is encountered in forest surveys, it may be concluded that no clear case of periodicity has been reported.

It may be added that to the author's knowledge Matern's statement on forests also holds for all case studies encountered so far in the application of Geostatistics. Problems clearly of an oscillatory nature seem more adequately approached in the frequency domain by Fourier methods. The analysis of propagating waves by means of an array of sensors is an example of such problems (Capon et al.; 1967 ; Capon, 1969). Note however, that time is involved and not only space.

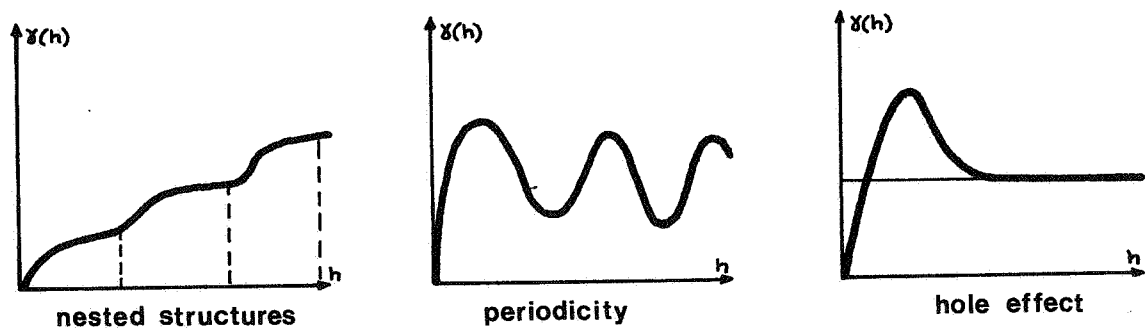


Fig. 10

- hole effect : presence of one or several bumps on the variograms (these would be holes on the covariance). The hole effect denotes a tendency for areas of high values to be surrounded by areas of low values, or "holes".

As with periodicities one must be careful not to confuse a genuine hole effect with mere sample variogram fluctuations. A safe recommendation is to disregard any bump unless it can be supported by a reasonable physical interpretation. For example J. Serra (1968), in a study of a thin section of Lorraine oolitic iron ore, found that calcite crystals tend to be separated by intervals roughly proportional to their sizes, and explained it by a process of calcite concentration around random germs. Another example of hole effect would be that induced by competition between plants, although as

noted by Matérn (1960, p. 62), such effect is usually hidden (i) by strong correlations in soil properties, (ii) by the averaging effect of sampling areas rather than points (effect of support).

- CHAPTER II -

THE PRACTICE OF VARIOGRAM INTERPRETATION,

CALCULATION AND MODELING

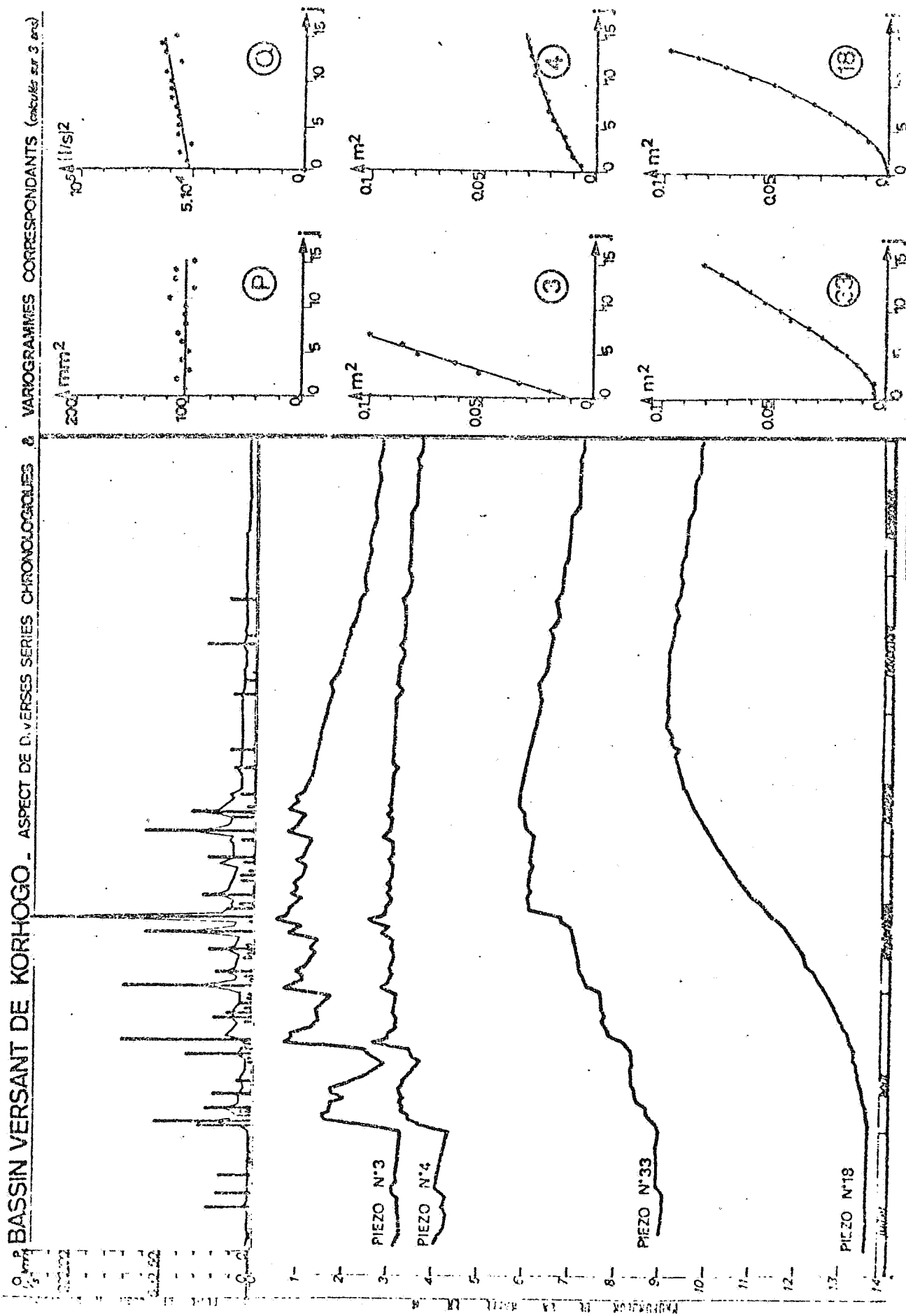
1 - EXAMPLES OF VARIOGRAM INTERPRETATION.

Let us first start with a very illustrative example of the relationship between the variability of the Re.V. itself and the slope of its variogram at the origin. The Re. V. is a piezometric level in an aquifer, measured through time at four different sites. For comparison, the rainfall and run-off in the catchment area have also been recorded. All these variables are regionalized in time and the curves on the left part of Fig. 1 show their evolution over a six month period from July to December. The variograms themselves were calculated over a period of three years and are drawn for short lags not exceeding 15 days.

The curve "Piezo n° 18" has a very regular evolution and its variogram is of parabolic type. Piezo n° 33 is also regular on the whole, except for some jumps which are accounted for by a nugget effect on the variogram. Piezo n° 4 has a greater variability and also jumps : its variogram is of linear type with nugget effect. As for Piezo n° 3 the fluctuations are of a larger amplitude : the variogram is the same as for n° 4, but with a greater slope. On the top of the figure we can see the rainfall (vertical lines) and the run-off. Rainfall is purely random from one day to another and run-off is scarcely structured.

Now here as in other situations the man who knows his data can go further into the interpretation. An hydrogeologist would notice that the regularity of the water level is related to the depth of the aquifer. Piezometers that are close to the surface (n° 3 and 4) are strongly influenced by erratic rainfall, whereas those which are deeper (n° 33 and 18) are regularized by the integrating effect of the superincumbent geological layers.

BASSIN VERSANT DE KORHOGO - ASPECT DE DIVERSES SERIES CHRONOLOGIQUES & VARIOGRAMMES CORRESPONDANTS (continue sur 3 ans)



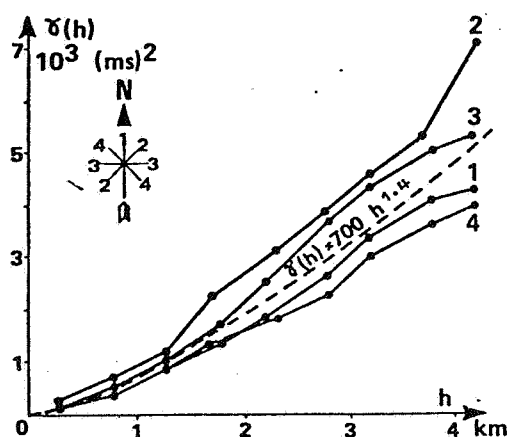


Fig. 2 - Time Variogram of the horizon H1

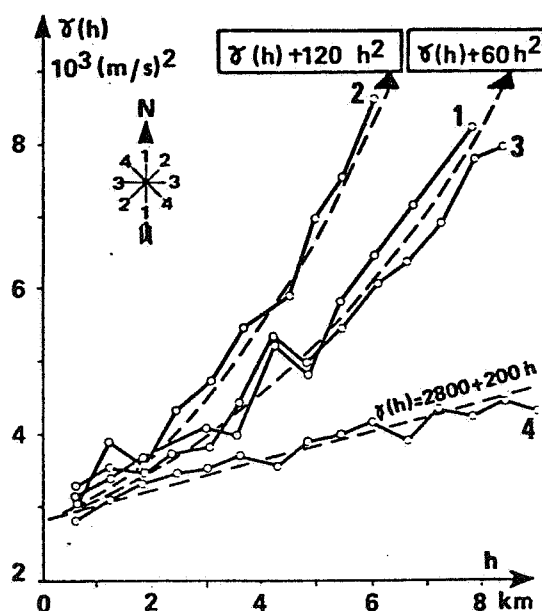


Fig. 3 - Velocity variogram of horizon H2

(From A. HAAS and C. JOUSSELIN, 1975)

In Fig. 2 we see variograms of seismic times computed in 4 directions. The variability is maximum in the SW-NE direction ($n^\circ 2$) and minimum in the orthogonal direction ($n^\circ 4$). Roughly we are in a case of **elliptic** anisotropy. However it is not strongly marked and for estimation purposes we can as well neglect it, provided we do not use the variogram for large distances. At the origin the variograms are of the parabolic type, which denotes a great spatial continuity of seismic times here.

On the contrary, the variogram of velocities (Fig. 3) shows a nugget effect estimated at $C_0 = 2800 \text{ (m/s)}^2$. It must be due to measurement errors whose standard deviation is therefore

$$\sqrt{2800 \text{ (m/s)}^2} \approx 53 \text{ m/s, i.e. an error range of about } \pm 100 \text{ m/s.}$$

This figure gives an idea of the magnitude of the discrepancy that can be expected between two close velocity analyses. At long distances we see a strong anisotropy: in the SE-NW direction ($n^\circ 4$) the

variogram increases gently while it grows steeply in the perpendicular direction. This growth being parabolic or faster indicates the presence of a drift along direction 2. Fortunately for the inference of the underlying variogram we are in a favorable situation since variogram n° 4 seems unaffected by this drift and can therefore be taken as the true underlying variogram.

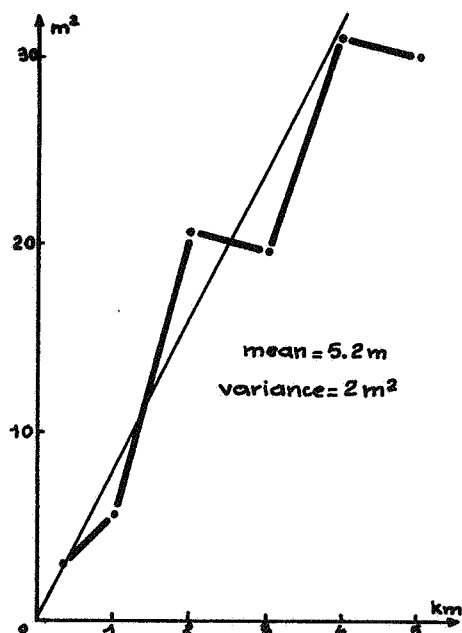


Fig. 4 - Thickness of a layer

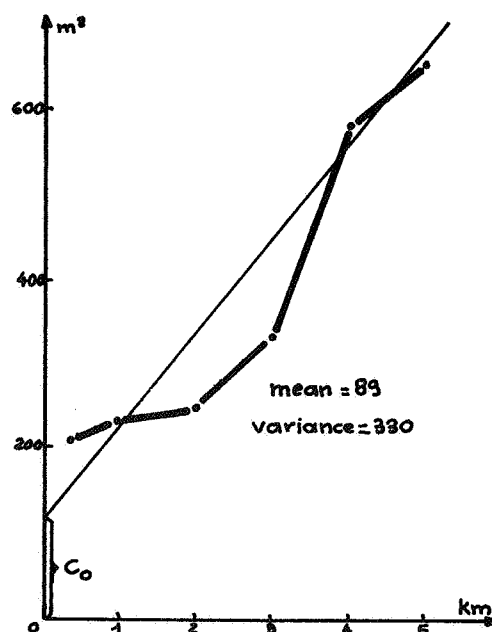
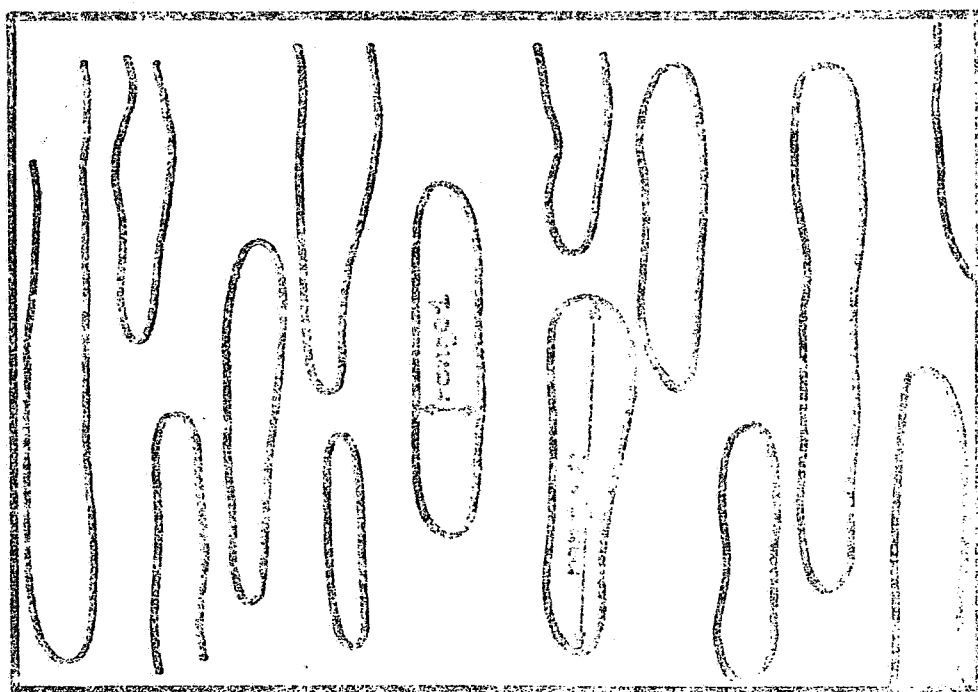


Fig. 5 - Cumulated log - Permeability

In Fig. 4 and 5 we have examples of linear variograms. For thickness there is no nugget effect : the thickness varies continuously, but not as continuously as the depth of a reservoir for example. On the variogram of Fig. 5 the nugget effect has been extrapolated down to $C_0 = 107$ (thickness is in meters and permeabilities in 10^{-4} Darcys) while the slope is estimated at $\omega = 0.095$. This implies that for a distance of 1 km the nugget effect accounts for half of the variance. Indeed we know that permeabilities have a highly irregular spatial behavior.

By means of a geometrical example, Fig. 6 illustrates the significance of the range and of anisotropies. The example is drawn from a lenticular mineral deposit, but such a pattern could be found in other fields. If grades in each lenticular body are

ZONE OF INFLUENCE



PERIODIC REGULARIZATION

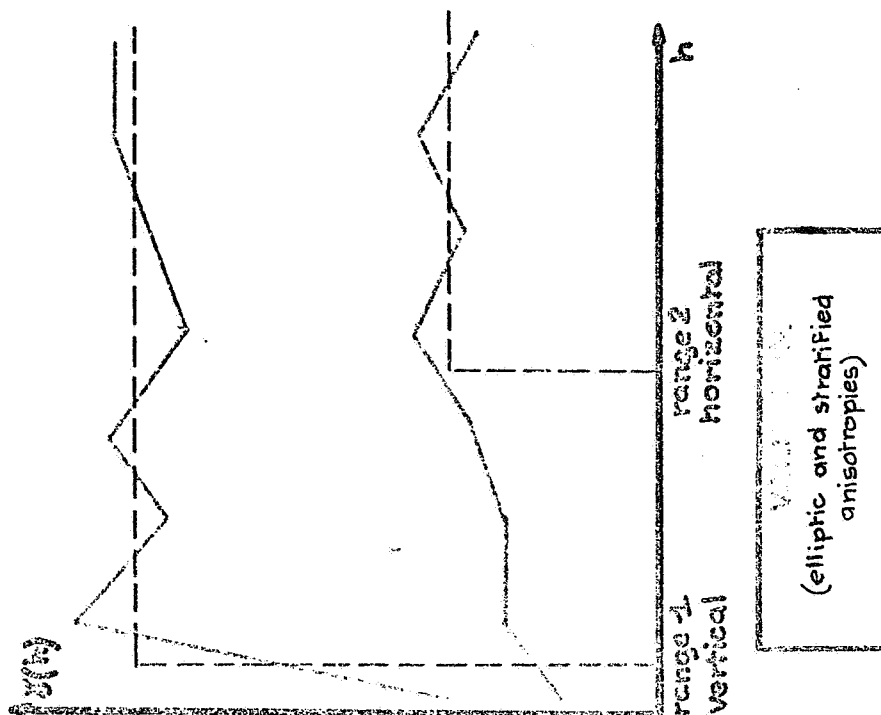


Fig. 6

independent, correlations extend only within a single body. Thus the range in one direction is the average length of the maximum cross-section in that direction. The ranges 1 and 2 show evidence of the geometrical anisotropy of the bodies. Moreover, the sills of the variogram are a lot different : the variation is greater along a vertical than in horizontal planes (stratified anisotropy).

2 - HOW TO COMPUTE A VARIOGRAM.

1 - Computation along a line.

Sometimes the problem can be reduced to computation along a line. Ex. : variogram of cores along a drill hole, variogram of seismic times along a profile, etc.. If the points are at a regular spacing a on the line, the variogram can be computed for lags h multiple of a by the formula :

$$\hat{\gamma}(h) = \frac{1}{2 N(h)} \sum_{i=1}^{N(h)} (Z(x_i+h) - Z(x_i))^2$$

where

- . $Z(x_i)$ are the data
- . x_i are the locations such that data are available both at x_i and x_i+h .
- . $N(h)$ is the number of such x_i 's, i.e. the number of pairs of points actually taken into the sum (if data are missing the pair is simply ignored, ex. Fig. 12).

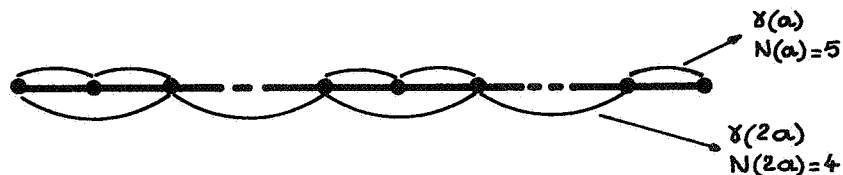


Fig. 7

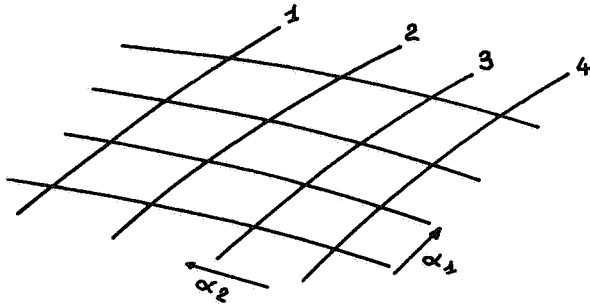


Fig. 8.

In case several profiles are available, one can compute the average variogram for each group of parallel lines along one direction.

$$\hat{\gamma}(h) = \frac{\sum_j N_j(h) \hat{\gamma}_j(h)}{\sum_j N_j(h)}$$

where $\hat{\gamma}_j(h)$ is the variogram for the j^{th} profile.

If the spacing along a line is not constant, a grouping by classes of distance should be made.

2 - Computation in the plane.

(a) Regular grid

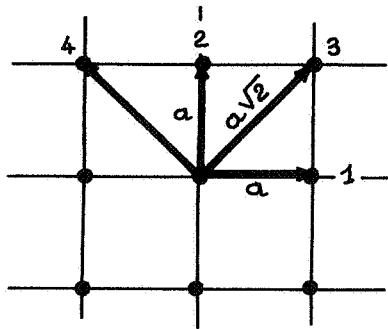


Fig. 9

This case is essentially similar to the case of parallel profiles just mentioned. The variogram will be computed in the 4 main directions (Attention! the lags are different in length along the diagonals of the grid).

(b) Profile sampling

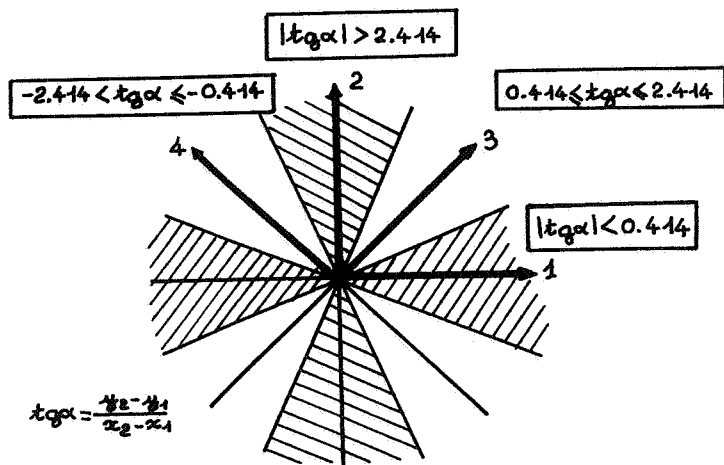


Fig. 10

Sometimes data are aligned along profiles but these are not necessarily parallel. Then we must group the variograms by classes of angles. The algorithm consists essentially of adding up the squared deviations corresponding to pairs of points which have the desired orientation. The selection can be done using for example the tangent of the angle.

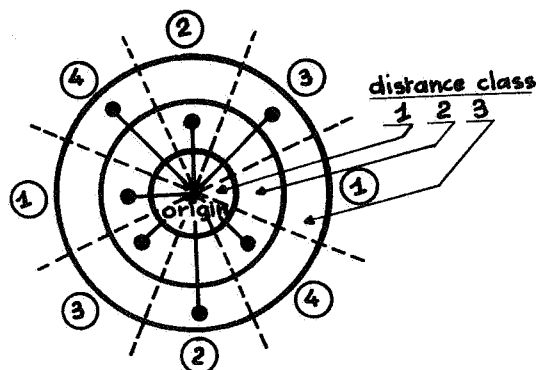


Fig. 11

(c) General case.

If the points do not show any systematic pattern, the variogram is computed by grouping both by classes of angles and classes of distances. This results in average values of the variogram within each of the given classes.

Practically, the procedure goes as follows: take each point in turn as origin, say x_i . Then for each other point, say x_j ($j \neq i$) compute the difference $Z(x_i) - Z(x_j)$, square this difference, look in what direction the vector (x_i, x_j) is, in what class of distance its modulus falls, and add the squared difference $(Z(x_i) - Z(x_j))^2$ to the relevant subtotal. Also, add 1 to the count of elements in that subtotal. Then delete the point taken as an origin and start again with another point. At the end of the process, divide all totals by twice the number of terms they have been computed with. Note that the procedure is independent of the order in which the successive origins are chosen.

3 - Computation in three dimensions.

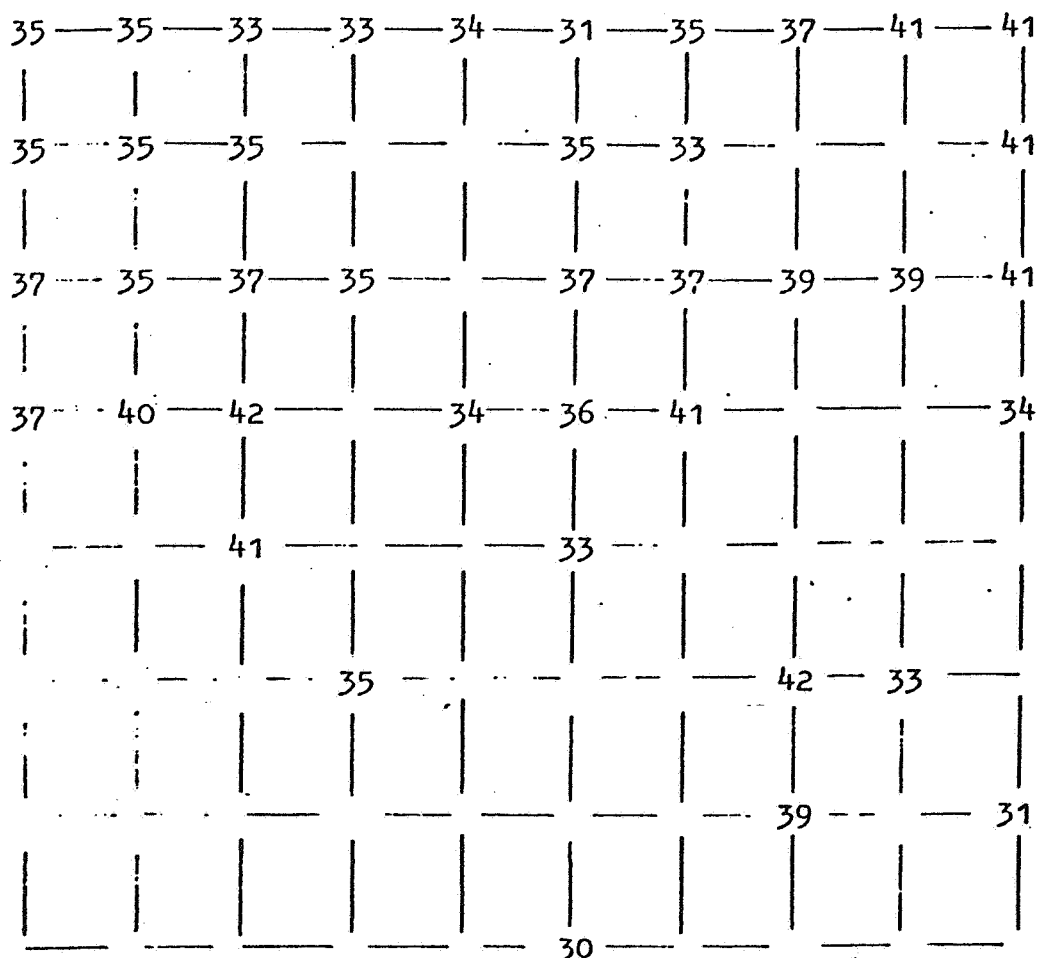
Theoretically, one could generalize the technique of distance and angle classes, the latter being solid angles. But in practice the third dimension usually plays a special role due to the vertical stratification of natural phenomena (geologic or atmospheric layers). One usually computes variograms layer by layer, and also vertical variograms along each bore-hole (or radio-sounding), possibly pooling these in the end. This information then serves as a basis for elaborating a 3-D variogram model.

3 - EXERCISE : Construction of a Variogram.

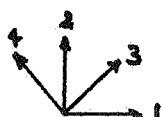
These data are taken from a computer simulation of a realization of a random function with a linear variogram $\gamma(h) = \omega|h|$. The data are set on a square grid of mesh a . The semi-variograms will be computed with 4 principal directions $\alpha = 1, 2, 3, 4$ and for the values of h ranging from lag 1 to lag 3. It will be noted that in directions 3 and 4 the lag will be equal to $a\sqrt{2} = 1.4 a$.

The number of calculations is small enough to be performed by hand, or with the help of a small calculator.

The 4 semi-variograms obtained as well as the mean semi-variogram (of the 4 directions) will be plotted on a graph and interpreted.



The points where no values exist are considered as unknown data. They must not play any role in the computation of the variograms.



Main directions

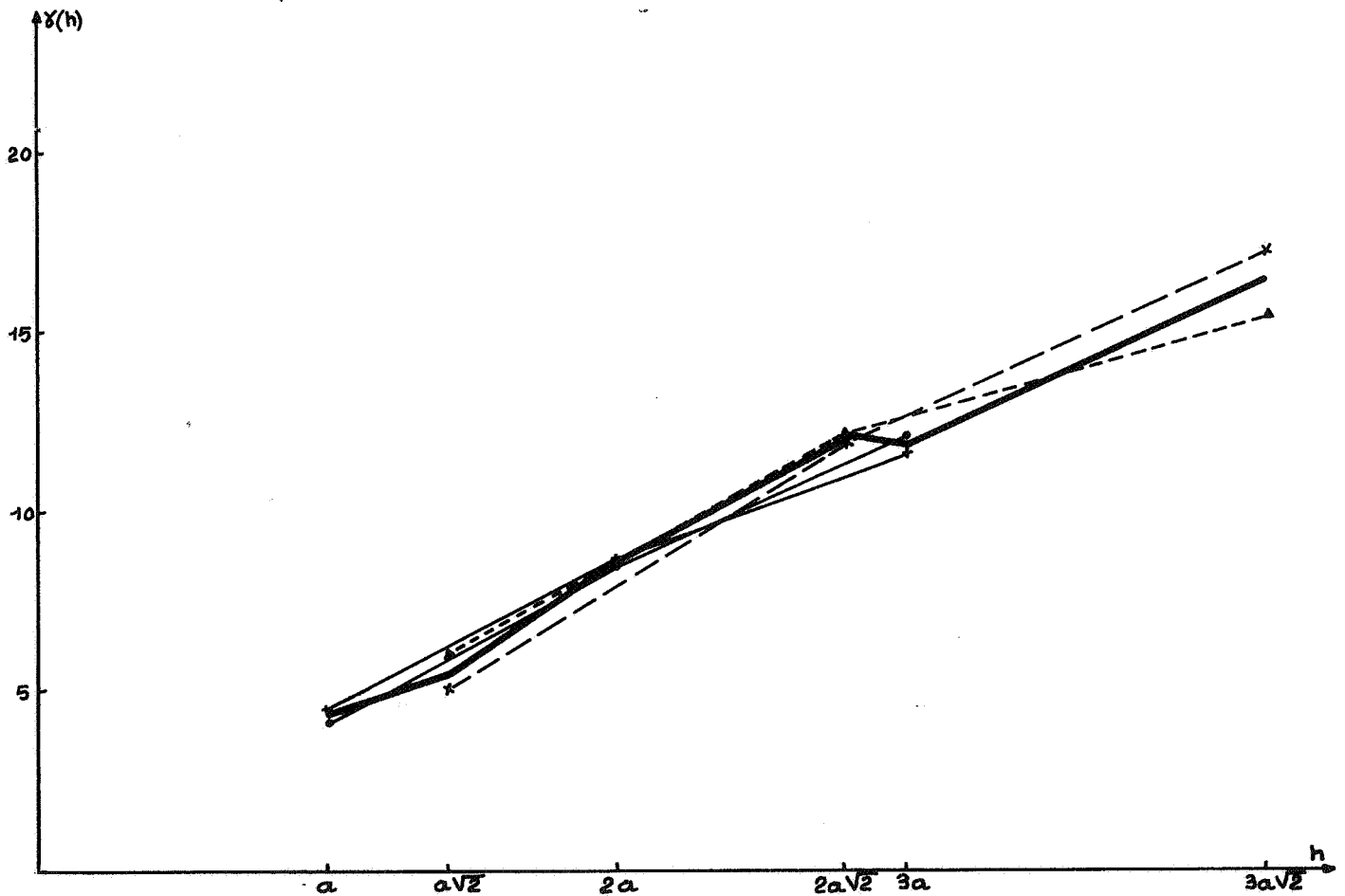
RESULTS

Variograms

		<u>h = 1 lag</u>		<u>h = 2 lags</u>		<u>h = 3 lags</u>	
		N(1)	$\gamma(1)$	N(2)	$\gamma(2)$	N(3)	$\gamma(3)$
Direction	1	24	4.10	19	8.42	18	12.08
	2	21	4.45	17	8.7	14	11.53
	3	17	6.58	13	12.11	8	15.44
	4	19	5.02	16	11.9	10	17.25

Mean Variogram

<u>h</u>	<u>a</u>	<u>$a\sqrt{2}$</u>	<u>2a</u>	<u>$2a\sqrt{2}$</u>	<u>3a</u>	<u>$3a\sqrt{2}$</u>
N.	45	36	36	29	32	18
$\gamma(h)$	4.26	5.76	8.55	11.99	11.84	16.44



The variograms are isotropic. They can be fitted to a linear model, with a slope $\omega = \frac{4.2}{a}$ and without nugget effect. This example is an "ideal case". In practice, for such a small number of data we we would have much more fluctuations in the variograms.

4 - MODELS FOR THE VARIOGRAM.

For estimation purposes, the experimental variogram may not be used directly. A theoretical model must be fitted first. The reason is that variograms have to satisfy certain conditions of mathematical consistency without which one would possibly come up with negative variances.

1. Admissible linear combinations.

Let us first consider a stationary variable $Z(x)$ with a covariance $C(h)$. The variance of a linear combination

$$Z^* = \sum_i \lambda_i Z(x_i)$$

is given by

$$(2.1) \quad \text{Var}[Z^*] = \sum_{i,j} \lambda_i \lambda_j C(x_i - x_j)$$

and must be non-negative whatever the system of points x_i and of weights λ_i . This condition expresses that $C(h)$ is necessarily a positive definite function. Conversely, any positive definite function may be considered as the covariance of some stationary random function $Z(x)$.

Now, if $Z(x)$ is only assumed to satisfy the intrinsic hypotheses it is not possible to calculate the variance of any linear combination but only that of linear combinations of increments. These will be called "admissible linear combinations". They are characterized by the restriction :

$$(2.2) \quad \sum_i \lambda_i = 0$$

Indeed, it is clear that any linear combination of increments satisfies (2.2) since any single increment involves the weights +1 and -1. Conversely, if (2.1) holds then :

$$\sum_i \lambda_i Z(x_i) = \sum_i \lambda_i [Z(x_i) - Z(o)]$$

is a linear combination of increments. As such it has a variance given by

$$(2.3) \quad \text{Var}\left[\sum_i \lambda_i Z(x_i)\right] = \sum_{i,j} \lambda_i \lambda_j \text{Cov}[Z(x_i)-Z(o), Z(x_j)-Z(o)]$$

To calculate the covariances of increments we use the following identity :

$$\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2 \text{Cov}(X,Y)$$

and write :

$$\begin{aligned} \text{Var}[Z(x)-Z(y)] &= \text{Var}[Z(x)-Z(o)] + \text{Var}[Z(y)-Z(o)] \\ &\quad - 2 \text{Cov}[Z(x)-Z(o), Z(y)-Z(o)] \end{aligned}$$

i.e.

$$2 \gamma(x-y) = 2 \gamma(x) + 2 \gamma(y) - 2 \text{Cov}[Z(x)-Z(o), Z(y)-Z(o)]$$

Solving for Cov(,) and substituting into (2.3) gives

$$\begin{aligned} \text{Var}\left[\sum_i \lambda_i Z(x_i)\right] &= \sum_{i,j} \lambda_i \lambda_j [\gamma(x_i) + \gamma(x_j) - \gamma(x_i-x_j)] \\ &= \left[\sum_j \lambda_j\right] \left[\sum_i \lambda_i \gamma(x_i)\right] + \left[\sum_i \lambda_i\right] \left[\sum_j \lambda_j \gamma(x_j)\right] \\ &\quad - \sum_{i,j} \lambda_i \lambda_j \gamma(x_i-x_j) \end{aligned}$$

Because of (2.2) the first two terms cancel and finally :

$$(2.4) \quad \boxed{\text{Var}\left[\sum_i \lambda_i Z(x_i)\right] = - \sum_{i,j} \lambda_i \lambda_j \gamma(x_i-x_j)} \quad \left(\sum_i \lambda_i = 0\right)$$

Thus the very useful result : the variance of a linear combination whose sum of weights is 0 may be calculated as with a covariance (formula 2.1), except that this covariance should be replaced by $-\gamma$.

As variances must be non-negative (2.4) indicates the condition that a function $\gamma(h)$ must satisfy to be a variogram :

for any system of points x_1, \dots, x_N and of weights $\lambda_1, \dots, \lambda_N$ subject to $\sum_i \lambda_i = 0$ one has

$$(2.5) \quad - \sum_{i,j} \lambda_i \lambda_j \gamma(x_i - x_j) \geq 0$$

$-\gamma$ is said to be conditionally positive definite.

Note that the condition on variograms is weaker than the condition on covariances since (2.5) has to hold only for systems of weights λ_i such that $\sum_i \lambda_i = 0$, rather than for all λ_i . Therefore the class of variograms is richer than that of covariances. It contains all bounded variograms associated with a covariance by $\gamma(h) = C(0) - C(h)$, but also unbounded variograms which have no covariance counterpart.

2. Common isotropic variogram models.

(a) Power Functions

$$\gamma(h) = \omega |h|^\lambda \text{ with } 0 < \lambda < 2$$

As a particular case we have the linear model $\gamma(h) = \omega |h|$ where the variogram is simply proportional to the distance.

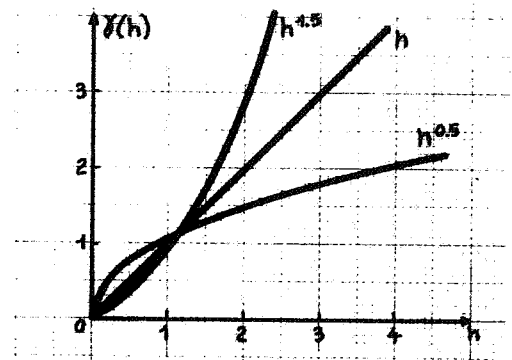


Fig.12

(b) Spherical model

$$\begin{cases} \gamma(h) = C \left[\frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \frac{|h|^3}{a^3} \right] & |h| \leq a \\ \gamma(h) = C & |h| > a \end{cases}$$

$a = \text{range}$ $C = \text{sill}$

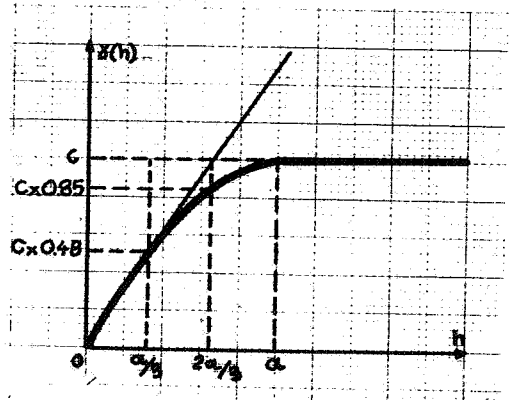


Fig.13

The spherical model is perhaps the most used one. It has a simple polynomial expression and its shape matches well what is often observed : a quasi-linear growth and then a stabilization. The tangent at the origin intersects the sill at a point of abscissa $2/3 a$, which is quite near the range.

The associated covariance $C(h) = C - \gamma(h)$ is proportional to the volume of the intersection of a sphere of diameter a with its translate by h , hence the name "spherical" model.

(c) Exponential model

$$\gamma(h) = C[1 - e^{-\frac{|h|}{a}}]$$

practical range $\approx 3a$ (95% of sill)
 $C = \text{sill}$

The tangent at the origin intersects the sill at a point of abscissa a .

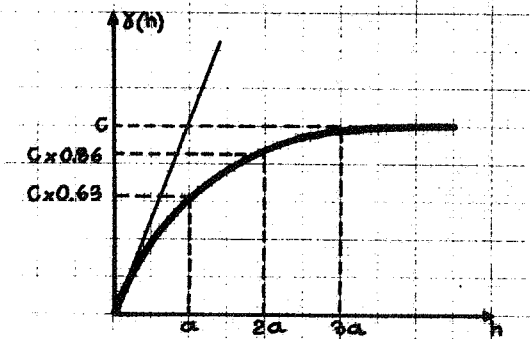


Fig.14

(d) Gaussian model

$$\gamma(h) = C[1 - e^{-\frac{|h|^2}{a^2}}]$$

practical range $\approx 1.73 a$
 $C = \text{sill}$ (95% of sill)

The Gaussian model represents an extremely continuous phenomenon. Practice however shows that numerical instabilities can occur when this model is used without a nugget effect.

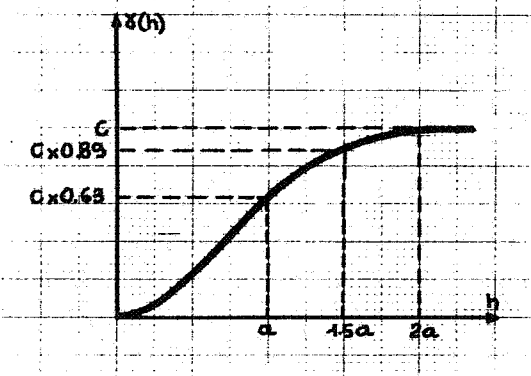


Fig. 15

(e) Cubic model

$$\gamma(h) = c \left(7 \frac{h^2}{a^2} - \frac{35}{4} \frac{h^3}{a^3} + \frac{7}{2} \frac{h^5}{a^5} - \frac{3}{4} \frac{h^7}{a^7} \right) \quad |h| \leq a$$

$$\gamma(h) = c \quad |h| \geq a$$

We name this model "cubic" because its irregular term of lowest degree is $|h|^3$ (Even powers being infinitely differentiable do not matter to define the regularity of the random function). The cubic model has the advantage of being smooth at the origin, though much less than the infinitely differentiable Gaussian model. Its overall shape resembles that of the spherical model.

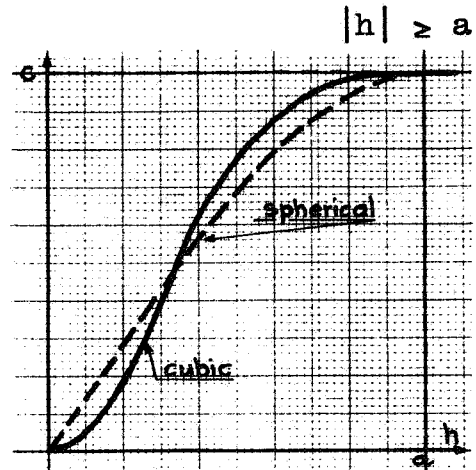


Fig. 16

3. Fitting the variogram.

This aspect is best seen through practical exercises. By and large, practice shows that the analytic form of the variogram does not matter too much as long as its major features are taken into account. These are :

- the slope at the origin
- the nugget effect
- the range
- the sill
- the anisotropies

The slope at the origin is usually assessed by means of the first three variogram points. Extrapolation to the origin gives the nugget effect C_0 (see next section).

The range is usually assessed visually. The sill is placed where the variogram stabilizes and is checked against the overall

variance of the data. They should be roughly equal if the range is smaller than the dimensions of the studied domain.

Anisotropies are the greatest source of difficulty. The main two methods (elliptic correction, stratification) have been discussed earlier. It may be hard to distinguish between drift and anisotropy.

In general a reasonably good fit can be obtained with the sum of two or at most three basic variogram models. The fitting is done by trial and error, preferably using an interactive visual display.

5. MODELS FOR THE NUGGET EFFECT.

The nugget effect was defined as a discontinuity of the variogram at the origin and received a simple physical interpretation: if measurements are taken at arbitrarily close points x and $x+h$, the difference $Z(x+h) - Z(x)$ does not tend to zero (in the mean square sense) but continues to fluctuate with an irreducible dispersion. Things are more complicated when it comes to modeling the nugget effect.

1. Apparent versus real nugget effect.

The primary difficulty is that in general there is a minimum data interdistance below which the variogram is simply not known. To make up for this lack of short distance information, usual practice is to extrapolate to the origin the first known variogram points. If the curve intersects the y-axis at a positive ordinate C_0 , then this is taken as an indication of the presence of a nugget effect of magnitude C_0 . On the other hand if the curve goes through the origin, it is surmised that there is no nugget effect. Naturally there is a risk of being wrong in both cases. An analysis at shorter distances may reveal that the variogram in fact dips continuously to zero and that the apparent discontinuity is merely due to the spacing of the data, which is too wide to detect structures at a smaller scale (Fig. 17-a). Alternatively, the first

observed variogram points may be aligned with the origin though the variogram has a bend and a nugget effect (Fig. 17-b) ; such behavior is typical of location errors (cf. II-5-5).

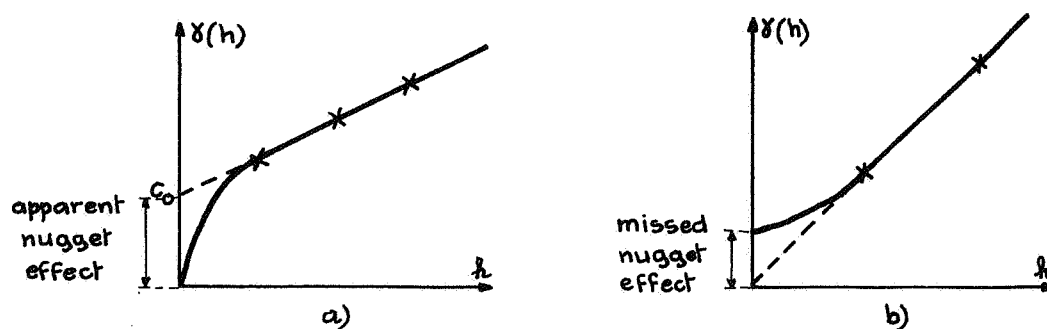


Fig. 17 : Examples of wrong diagnosis on nugget effect using extrapolation.

The risk of wrong diagnosis is inherent to discrete sampling. Readers familiar with time series will recall that if the sampling interval is Δt , it is impossible to observe frequencies beyond the Nyquist folding frequency $1/2 \Delta t$. Either it is assumed that higher frequencies are absent or else Δt must be reduced. Equivalently, one can either hypothesize the short distance behavior of the variogram or, which is always safer, measure it with an ad hoc short distance survey.

Hypotheses about the variogram may derive from physical information available about the studied variable. Often, the variance of measurement errors is known from separate experiments and may account for part of the nugget effect. In other cases we know that the variable is continuous - e.g. thickness of a layer - and model the apparent nugget effect by a spherical variogram with a sill C_0 and a range smaller than the shortest observed lag (Fig. 17-a).

2. Discrete noise and white noise.

A simple way of generating a nugget effect is to add "noise" to the data, i.e., uncorrelated, zero-mean random variables ε_i , with common variance C_0 . Indeed if $Z(x)$ has a continuous variogram $\gamma(h)$ and if ε_i is a noise uncorrelated with $Z(x)$, the variogram of

$$Z_1(x_i) = Z(x_i) + \varepsilon_i$$

is

$$\begin{aligned} \frac{1}{2} E[(Z(x_i) + \varepsilon_i) - (Z(x_j) + \varepsilon_j)]^2 &= \frac{1}{2} E[Z(x_i) - Z(x_j)]^2 + \frac{1}{2} E[\varepsilon_i - \varepsilon_j]^2 \\ &= \gamma(x_i - x_j) + C_0 - \text{Cov}(\varepsilon_i, \varepsilon_j) \end{aligned}$$

or equivalently

$$\begin{aligned} \gamma_1(h) &= 0 & h &= 0 \\ &= \gamma(h) + C_0 & h &\neq 0 \end{aligned}$$

and shows a discontinuity at the origin. The nugget effect is simply modeled here as a variogram with a zero range.

The above approach however, is essentially discrete. The ε_i are added to the data, as measurement errors, and not to the phenomenon $Z(x)$ itself. In the continuous case, it is extremely difficult to imagine a process $\varepsilon(x)$ with zero range, that is, whose values are uncorrelated at any two distinct points, arbitrarily close to one another. Realizations of such a "purely random process" are so erratic that we could not draw them. In addition $Z(x)$ has puzzling properties. Its average over any interval is zero exactly : no matter how small, the interval contains infinitely many uncorrelated terms that compensate. More, the Fourier transform of the covariance of $\varepsilon(x)$

$$\begin{aligned} C(h) &= C_0 & h &= 0 \\ &= 0 & h &\neq 0 \end{aligned}$$

is zero identically : the process has no power ! In fact Wiener-

Khintchine's spectral theory is applicable only to processes that are continuous in the mean square and therefore $\varepsilon(x)$ is excluded (cf. Doob, 1952, p. 523). Mathematical theory precisely manages to avoid processes with too unpleasant sample paths, such as $\varepsilon(x)$, which cannot realistically model any observable phenomenon.

The interesting generalization of discrete noise to the continuous parameter case is the so-called "white noise". It is defined as a zero-mean process with a constant spectral density over all frequencies, and is named "white" by analogy with the flat spectrum of white light (as opposed to a preponderance of low frequencies for red light and of high frequencies for blue light). Strictly speaking, such a process cannot exist in the ordinary sense for its total power, the variance, is infinite. However it receives a clear interpretation in the scope of "generalized random processes" (cf. Yaglom, 1962, p. 208 ff.). The idea is that physically a process is never observed at "points" since the measuring device always introduces averaging. Likewise, a generalized random process does not necessarily have point values but convolution by a regular enough function turns it into an ordinary process. Such a convolution which makes the process more regular is called a regularization. In that theory, the covariance of white noise is $A \delta$ ($A > 0$) where δ is the generalized function (Dirac δ -function) defined by :

$$\delta(\varphi) = \int \delta(x) \varphi(x) dx = \varphi(0)$$

(it is assumed that $\varphi \in \mathcal{C}$, the class of continuous functions vanishing outside a bounded domain). δ is often described as a "function" which is zero everywhere except at the origin where it is infinite, and has an integral equal to one (Fig. 18 b).

This infinite variance makes all the difference from $C(h)$ of the zero-range model and intuitively explains why the compensation effect over intervals does not take place anymore. A telling heuristic derivation is given by Yaglom (1962, p. 64) : in 1-D consider the covariance :

$$C(h) = C e^{-h/a}$$

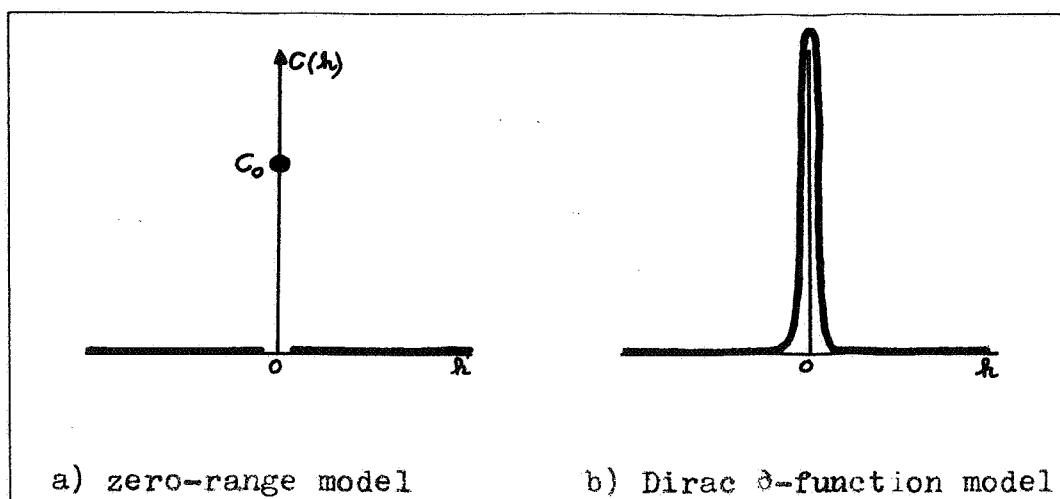


Fig. 18

and let $a \rightarrow 0$ and $C \rightarrow \infty$ in such a way that $Ca = \text{constant}$. Then $C(h)$ tends to zero for any $h \neq 0$, while $C(0) \rightarrow \infty$ in such a way that

$$\int_{-\infty}^{+\infty} C(h) dh = Ca$$

remains constant. The limit of $C(h)$ is indeed $A\delta$ with $A = Ca$; a similar result holds in \mathbb{R}^n with $A = Ca^n$. Note that the zero-range covariance corresponds to $a \rightarrow 0$ with C fixed.

Convolution of white noise through a function φ produces an ordinary process with covariance :

$$C_{\varphi}(h) = A \int \varphi(x) \varphi(x+h) dx = A(\varphi * \check{\varphi})$$

where $\check{\varphi}(x) = \varphi(-x)$.

In particular, data with support v are of the form :

$$Z_v(x) = \frac{1}{v} \int_v Z(x+u) du$$

i.e. a convolution of Z by $\check{\phi}(u)$ where

$$\phi(u) = I(u)/v$$

$I(u)$ being the indicator function of v (equal to 1 inside v , to 0 outside). Let us now introduce the function

$$P(h) = \int I(u) I(u+h) du = I * \check{I}$$

called the "geometric covariogram of v " (Matheron, 1965, p. 20). The product $I(u) I(u+h)$ is one inside the intersection of v with its translate by $-h$, denoted v_{-h} , and is zero outside. Therefore $P(h)$ is the measure (length, area or volume) of $v \cap v_{-h}$:

$$P(h) = \text{Meas} (v \cap v_{-h})$$

For $h = 0$, $P(0) = \text{Meas} (v \cap v) = \text{Meas} (v) = v$

so that the variance of the white noise affecting Z_v is

$$\frac{A}{v^2} P(0) = \frac{A}{v}$$

inversely proportional to the block volume. $P(h)$ vanishes as soon as $|h| > D$, the maximum diameter of v (Fig. 19). On the variogram of Z_v the nugget effect is reflected by the addition of a variogram with a sill $C_0 = A/v$ and a transition zone of the order of the dimensions of v .

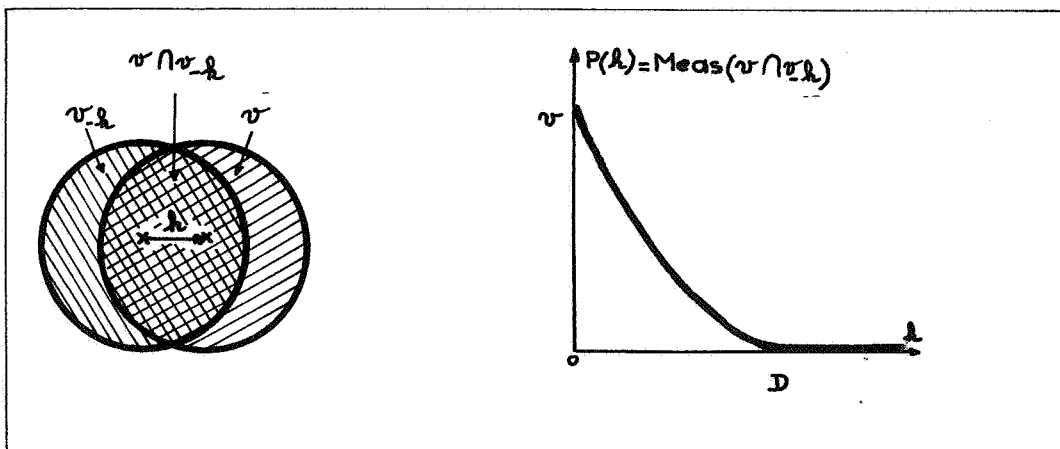


Fig. 19 : An example of geometric covariogram

In conclusion, we have seen two types of noise that can generate a nugget effect :

- i) one is a discrete noise which is defined only at the data points and reflects measurement errors.
- ii) another is a component of the studied variable itself. A strictly punctual model is that of white noise with covariance $A\delta$. But as soon as data have a finite support v , as is always the case, the nugget effect is seen as a variogram reaching a sill $C_0 = A/v$ after a transition zone reflecting the geometry of the support v .

3. Pure and compound Poisson models : Exercise.

Suppose that the circle in Fig. 19 represents a pail collecting rain water. Two pails side by side will not collect exactly the same amount of water and rainfall data will display a nugget effect (experiments have been made with rain gauges). A simple model is to consider that rain drops are independent points falling at random with a probability proportional to the area offered, i.e. that they follow a 2-D Poisson process. If all drops have the same weight w_0 the amount of water falling perpendicularly to a pail of upper cross-sectional area S is $w_0 N(S)$, where $N(S)$ is a Poisson random variable with parameter θS .

1° - Let $Z(x) = w_0 N(S_x)$ be the weight of water collected when the pail is centred at point x . Show that the covariance of $Z(x)$ and $Z(x+h)$ is w_0^2 times the variance of $N(S_x \cap S_{x+h})$, i.e. $w_0^2 \theta P(h)$ where $P(h)$ is the geometric covariogram of S .

2° - Assume now that the rain drops have independent random weights (mean w_0 , variance σ_w^2). The weight $Z(x)$ of water collected is the sum $w_1 + \dots + w_N$ of the $N = N(S_x)$ drops that fell in S . Calculate the mean and variance of $Z(x)$ and hence its covariance.

(Hints : condition on $N = n$ and randomize N).

(Answers : $E(Z) = w_0 \theta S$, $\text{Var}(Z) = (w_0^2 + \sigma_w^2) \theta S$, $C(h) = (w_0^2 + \sigma_w^2) \theta P(h)$).

These results are identical to those found with the regularized white noise model ; this becomes intuitive when considering that the covariance measure of the Poisson process is $\theta \delta$.

The same model applies in 3-D to represent the distribution of mineralized grains (nuggets) in a sample of ore.

4. Integration of microstructures.

Modeling mineralized grains as points makes sense only if the support of the sample is large enough with respect to these grains. If mineral grade were analyzed on a point basis, for example by taking thin sections, the variogram would reflect the structures at the scale of the grains. It has been seen above (§ II-5-1) that when the sampling interval is large, microstructures are not perceptible and appear as a nugget effect whose magnitude is equal to (or greater than) the variance of the microstructures.

Just as for white noise or Poisson points, integration of microstructures over a volume v results in a nugget effect of magnitude A/v , provided that v is large enough with respect to the range of the microstructures. To fix the ideas let $C(h)$ denote their covariance and a their (maximum) range. The variance of

$$Z_v = \frac{1}{v} \int_v Z(x) dx$$

is

$$\begin{aligned} \text{Var}[Z_v] &= \frac{1}{v^2} \int_v dx \int_v C(x-y) dy \\ &= \frac{1}{v^2} \int C(h) P(h) dh \end{aligned}$$

where $P(h)$ is the geometric covariogram of v . Now, $C(h) P(h)$ is zero as soon as $|h| > a$ because $C(h)$ vanishes, and on the other hand, if v is large with respect to a :

$$P(h) \simeq P(0) = v \quad \text{for } |h| < a$$

So
$$\text{Var}[Z_v] \simeq \frac{1}{v} \int C(h) \, dh$$

which is indeed of the form A/v with

$$A = \int C(h) \, dh$$

Note that letting $a \rightarrow 0$ while $\int C(h) \, dh$ remains constant leads to the $A\delta$ covariance encountered above.

5. Location errors.

A location error occurs when the value associated to a point x has actually been measured at another point $x+u$. Instead of studying a variable $Z(x)$ we are in fact studying :

$$Z_1(x) = Z(x+u)$$

where u is a random vector with probability distribution $p(du)$ whose form is assumed to be known. Likewise we assume that the form of the joint probability $p(du, du')$ is known for any pair of vectors (u, u') .

For fixed u the mean of $Z_1(x)$ is

$$E[Z_1(x) | u] = E[Z(x+u) | u] = m(x+u)$$

so that randomizing u entails

$$E Z_1(x) = \int m(x+u) \, p(du)$$

If $E Z(x) = m = \text{constant}$ then also $E Z_1(x) = m$ and the variogram of $Z_1(x)$ is

$$\begin{aligned} \gamma_1(h) &= \iint \frac{1}{2} E[Z(x+h+u) - Z(x+u')]^2 \, p(du, du') \\ &= \iint \gamma(h+u-u') \, p(du, du') \end{aligned}$$

The formula is valid for $h \neq 0$. Even though γ may be continuous, γ_1 will in general appear as having a nugget effect of magnitude

$$C_0 = \iint \gamma(u-u') p(du, du')$$

The only exception is for the distribution $p(du, du')$ concentrated on the line $u = u'$, corresponding to equal location errors at x and $x+h$.

An interesting example is given by Chilès (1976). The data are depths of sea floor measured along lines by non-specialized ships. Measurement errors affect both the depth values and the coordinates. The uncertainty on location is modeled as an isotropic Gaussian vector with a circular standard deviation of 5 km - which is quite appreciable. As a first approximation it may be assumed that all points along a single survey line are subject to the same displacement error ; the variograms calculated along lines are therefore not touched. On the other hand, displacement errors on two different lines may be assumed independent and the variogram calculated with pairs of points belonging to two different lines is of the form :

$$\tilde{\gamma}(h) = \iint \gamma(h+u-u') p(u) p(u') du du'$$

Fig. 20 shows an example of the experimental curves and of the fits obtained. This has permitted the determination of the standard deviation of the location error.

Note that the variogram is not only shifted as in the case of measurement errors : its shape is also altered towards a greater regularity (parabolic behavior at the origin instead of a linear one).

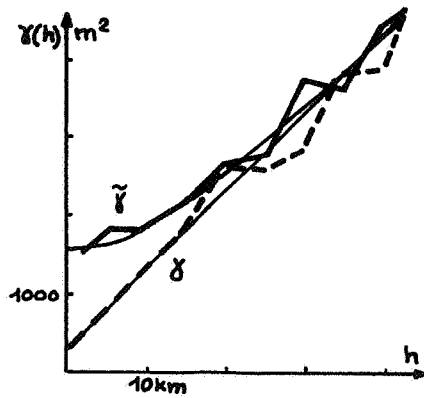


Fig. 20 : Experimental vario-grams of sea-floor depths

γ : along survey lines

$\tilde{\gamma}$: across survey lines

EXERCISE (from Chilès, 1977).

Let $\gamma(h) = \omega r$ be an isotropic linear variogram and assume that location errors are independent with a Gaussian distribution :

$$p(du) = \frac{1}{2\pi \sigma^2} e^{-\frac{r^2}{2\sigma^2}} r dr d\theta \quad \text{where } r^2 = |u|^2$$

Show that the observed variogram $\tilde{\gamma}(h)$ admits the following limited expansions :

$$\tilde{\gamma}(h) \simeq \omega \sqrt{\frac{\pi}{2}} \sigma \left(1 + \frac{h^2}{4\sigma^2}\right) \quad h \text{ small}$$

$$\tilde{\gamma}(h) \simeq \omega h \left(1 + \frac{\sigma^2}{2h^2}\right) \quad h \text{ large}$$

Fig. 21 shows the two curves. Note that the apparent nugget effect is :

$$\tilde{\gamma}(0) = \omega \sqrt{\frac{\pi}{2}} \sigma$$

which provides a means of evaluating σ .

Hint : Use the second order expansion of $\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8}$

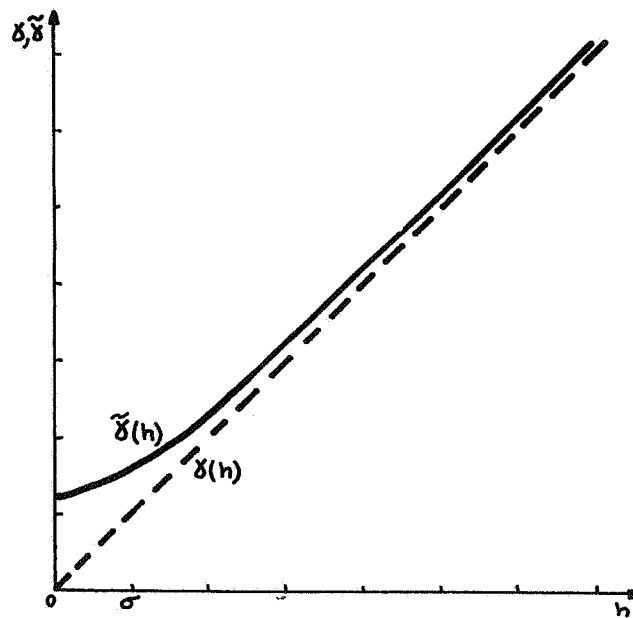


Fig. 21 : Linear variogram $\gamma(h)$ altered by location errors

- CHAPTER III -

DISPERSION AS A FUNCTION OF BLOCK SIZE

1 - THE SUPPORT OF A REGIONALIZED VARIABLE.

Often a Re.V. is not defined at a point but rather on a surface or a volume. A porosity or a permeability are examples of such a Re.V. The basic volume on which a Re.V. is measured is called the support of the Re.V. If we change the support we obtain a new Re.V., related to the initial one, but with different structural characteristics. It is well known that porosities measured on a thin section of a plug, on the plug itself, on a full size core, along the whole well, do not have the same significance. This is even more so for permeabilities. Hence the important question : how can we relate one support to the other? In other words : knowing the porosities of cores, what can we say about the porosities of blocks?

The answer will be given in two stages. First we shall consider only the dispersion as a function of the support ; then we shall see how the variogram of larger volumes can be deduced from the variogram of the samples.

2 - DISPERSION VERSUS BLOCK SIZE - AN EXAMPLE.

Table I shows porosity data as measured on a thin section of a very porous and permeable sandstone (Vosges sandstone). The section was divided into 324 contiguous square areas with a side of length 800 microns ($1 \text{ micron} = 10^{-6} \text{ meter}$). (It may seem strange to work at such a small scale but where else could one get such a complete survey?). Porosity values were then averaged by groups of 4 (blocks 2×2), by groups of 9 (blocks 3×3), by groups of 36 (blocks 6×6). The results are shown in Table II.

20.81	20.42	24.43	25.67	26.05	21.22	18.17	21.14	27.43	19.46	20.57	8.77	7.53	15.19	24.11	28.58	29.83	23.88
19.37	13.94	21.62	20.02	11.93	6.81	18.46	19.38	23.27	29.30	25.06	22.40	29.84	25.10	21.57	29.11	26.57	17.72
20.92	23.60	18.81	16.29	25.20	20.13	22.33	20.91	24.68	26.30	20.75	22.14	19.20	19.54	20.80	13.94	20.41	19.26
28.45	22.61	24.70	15.96	23.34	21.50	25.61	29.23	23.91	35.63	33.76	21.58	21.27	24.37	23.55	16.43	25.33	20.10
22.82	22.29	16.97	26.87	27.28	19.51	25.37	28.08	15.49	17.23	24.70	29.04	22.93	31.76	18.63	22.29	27.55	29.51
22.32	25.64	21.35	24.68	21.39	21.75	21.59	31.30	33.57	21.99	22.78	25.95	26.10	26.39	37.22	27.03	15.09	18.41
20.96	19.89	24.44	29.59	25.34	32.10	22.48	28.12	23.34	24.15	27.42	18.49	28.17	21.38	21.46	29.95	26.31	33.14
21.93	23.48	22.76	24.46	22.16	30.37	26.43	28.07	28.11	30.80	25.72	28.99	25.85	26.76	18.87	25.18	22.15	26.74
16.82	16.83	21.03	23.89	26.17	21.93	15.83	19.66	27.46	24.77	36.22	24.12	23.55	25.51	32.85	24.38	32.79	26.88
27.89	28.24	25.09	25.48	22.84	24.30	28.29	20.51	24.72	19.63	24.34	24.34	26.21	23.33	16.53	21.50	16.46	22.02
13.60	21.14	17.65	23.84	21.69	23.70	17.89	24.50	18.42	16.51	23.18	30.37	22.86	19.47	24.93	17.45	25.35	25.95
23.68	23.73	15.96	29.98	9.34	26.86	29.14	30.63	26.94	22.04	22.30	25.44	21.48	26.35	13.96	26.38	17.60	23.71
14.96	20.84	20.50	22.79	22.88	20.51	25.65	24.79	24.84	23.54	21.98	23.22	25.66	21.05	21.63	23.77	25.04	23.28
20.75	26.58	21.19	18.45	20.37	23.68	27.81	23.39	21.47	19.91	26.44	19.10	22.02	12.16	15.31	23.14	16.10	23.56
15.98	20.66	19.98	17.78	20.43	24.15	23.35	27.11	29.41	24.72	18.91	26.53	24.48	21.95	23.15	25.51	24.52	21.41
21.30	27.13	25.13	19.37	19.48	24.01	29.95	21.98	21.70	20.58	26.63	18.37	16.28	23.87	21.37	14.45	19.19	20.32
19.36	22.50	22.22	6.63	19.12	18.72	27.77	22.45	26.15	26.20	21.63	27.85	21.44	19.46	19.08	16.86	26.83	21.33
20.45	24.61	22.43	26.00	23.88	25.59	24.24	25.89	23.99	23.40	21.30	17.80	19.77	15.52	9.75	20.99	16.36	15.82

TABLE I

18.63	22.93	16.50	19.29	24.86	19.20	19.41	25.84	24.48
23.89	18.94	23.04	24.52	27.63	24.56	21.09	18.68	21.27
23.27	22.47	22.48	26.58	22.07	25.62	26.79	25.84	22.64
21.56	25.31	27.49	26.27	26.60	25.15	25.54	23.86	27.08
22.44	23.87	23.81	21.07	21.14	27.25	24.65	23.81	24.54
20.54	21.86	20.40	25.54	20.98	25.32	22.54	20.68	23.15
20.78	20.73	21.86	25.41	22.44	22.68	20.22	20.96	21.99
21.27	20.56	22.02	25.60	24.10	22.61	21.64	21.12	21.36
21.73	19.32	21.83	25.09	25.43	22.14	19.05	16.67	20.08

POROSITIES - BLOCKS 2 x 2

20.44	19.48	21.75	21.64	20.32	23.25
23.02	22.70	26.02	25.85	25.80	22.41
20.90	26.22	24.39	26.74	24.93	27.50
21.89	23.11	24.56	23.13	21.68	21.82
20.16	21.23	25.31	22.70	20.82	22.92
22.79	20.31	25.12	22.64	18.50	19.13

POROSITIES - BLOCKS 3 x 3

21.35	23.81	22.95
23.03	24.70	23.98
21.12	23.95	20.34

POROSITIES - BLOCKS 6 x 6

TABLE II

The histograms of the original data and of the aggregated ones are shown in Fig. 1, 2, 3. We see that of course the mean values are the same, but the dispersions decrease sharply with the size of the averaging blocks. While in the original batch of data values were ranging from 6,63% to 37,22%, with 2×2 blocks they range only from 16,50% to 27,63%, with the 3×3 blocks from 18,50% to 27,50% and with the 6×6 blocks from 21,12% to 24,70%. As for the variances they are respectively (in 10^{-4} units) :

$$\sigma_{1 \times 1}^2 = 22.31 \quad \sigma_{2 \times 2}^2 = 6.42 \quad \sigma_{3 \times 3}^2 = 5.11 \quad \sigma_{6 \times 6}^2 = 1.99$$

Now suppose we did not have all the 324 values but only some of them. Then we could not compute the variances we have just obtained.

But Geostatistics makes it feasible. The only things we need are the variogram and adequate graphs (or a computer program).

3 - VARIANCES OF DISPERSION WITHIN A VOLUME V.

1. Variance of a point within V.

For convenience we shall name the supports V or v but if we are working in 2-D it should be understood that V or v are surfaces.

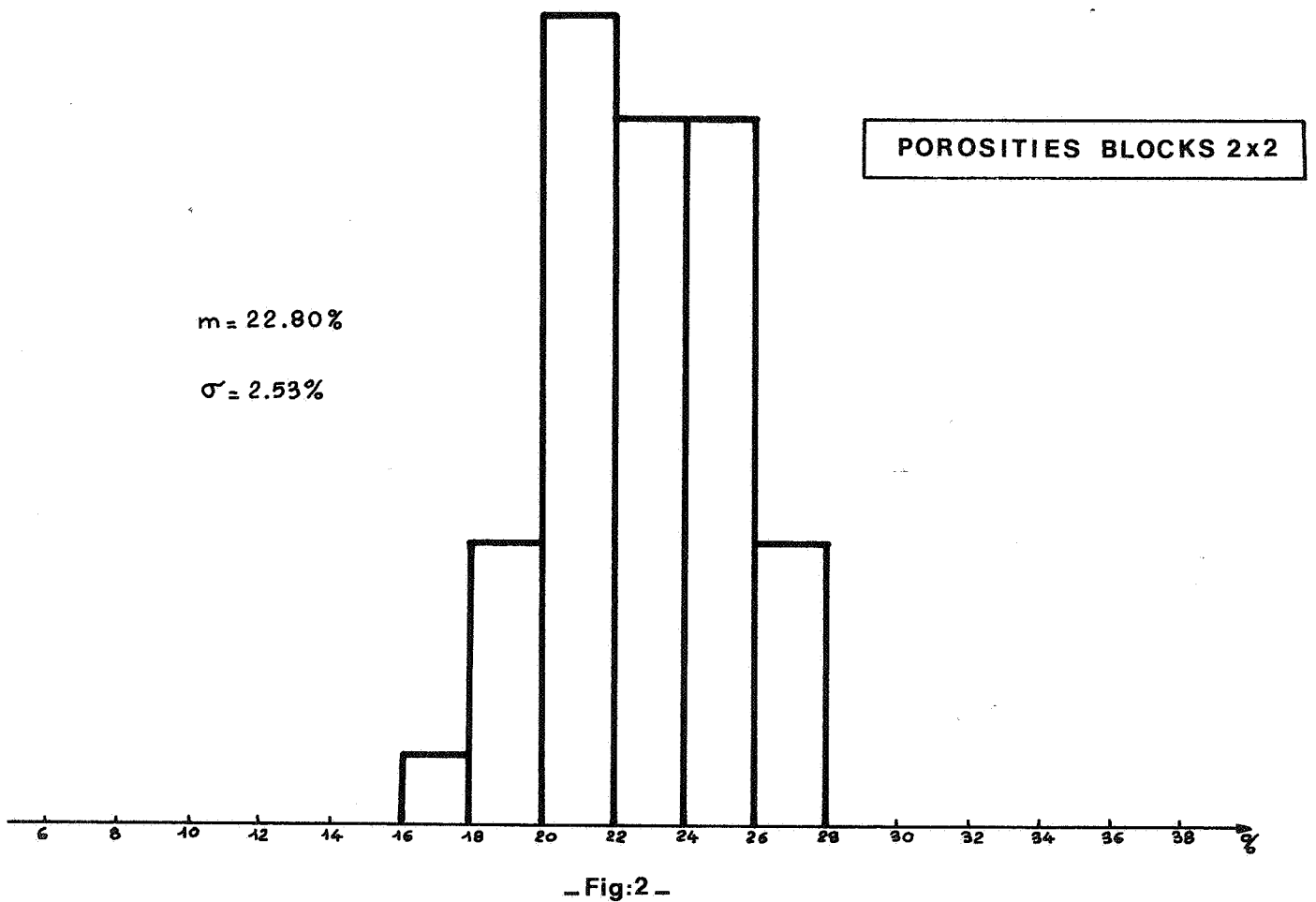
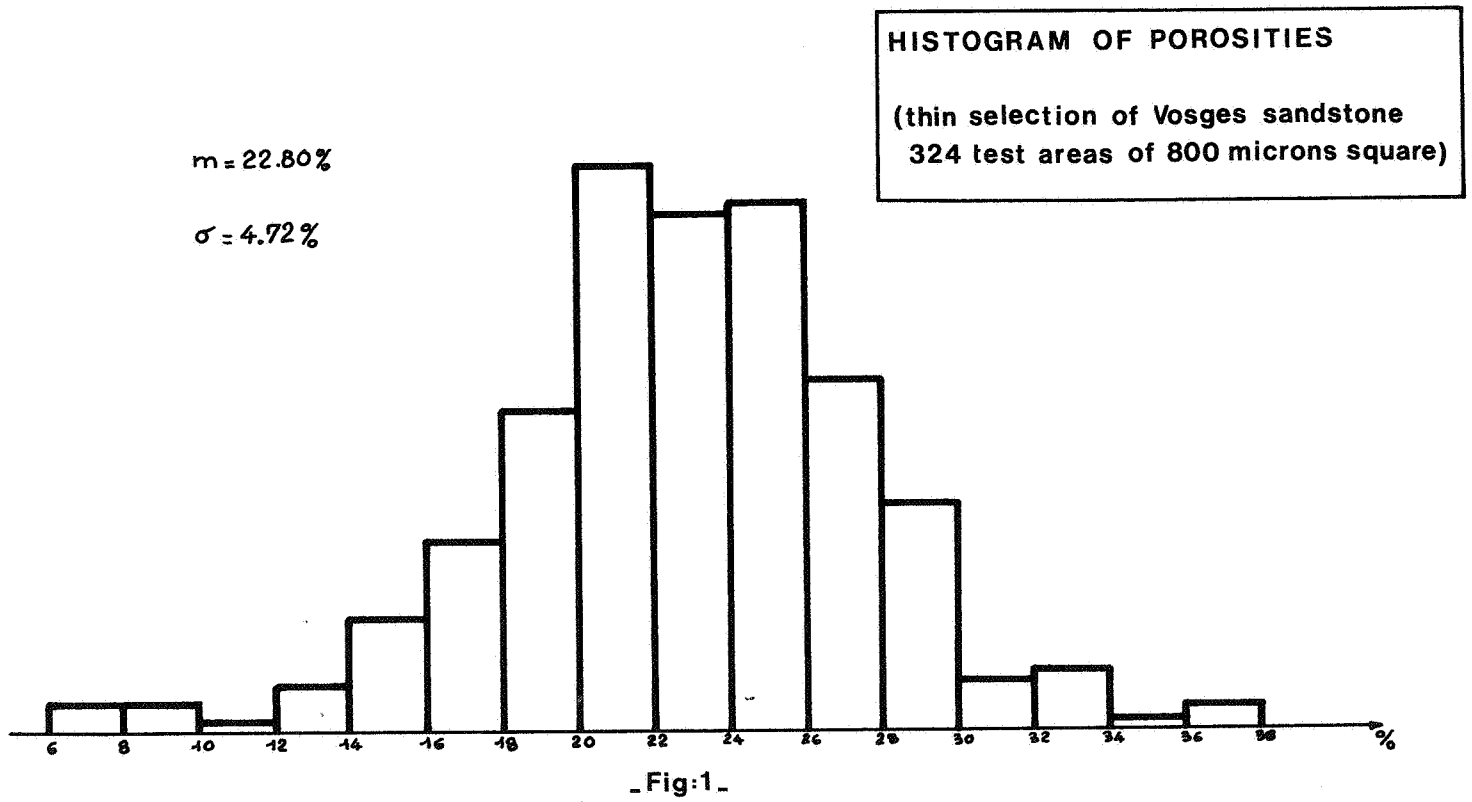
Under our model the studied variable is viewed as a realization $z(x)$ of a random function $Z(x)$. If we had all the values of $z(x)$ over a domain V , we could compute the mean and variance of $z(x)$ within V as :

$$m_V = \frac{1}{V} \int_V z(x) dx$$

$$s^2(0|V) = \frac{1}{V} \int_V (z(x) - m_V)^2 dx$$

Letting now the realization vary, the variance of $Z(x)$ within V , denoted $\sigma^2(0|V)$ is defined as the expected value of $s^2(0|V)$ over all possible realizations :

$$\sigma^2(0|V) = E(s^2(0|V))$$



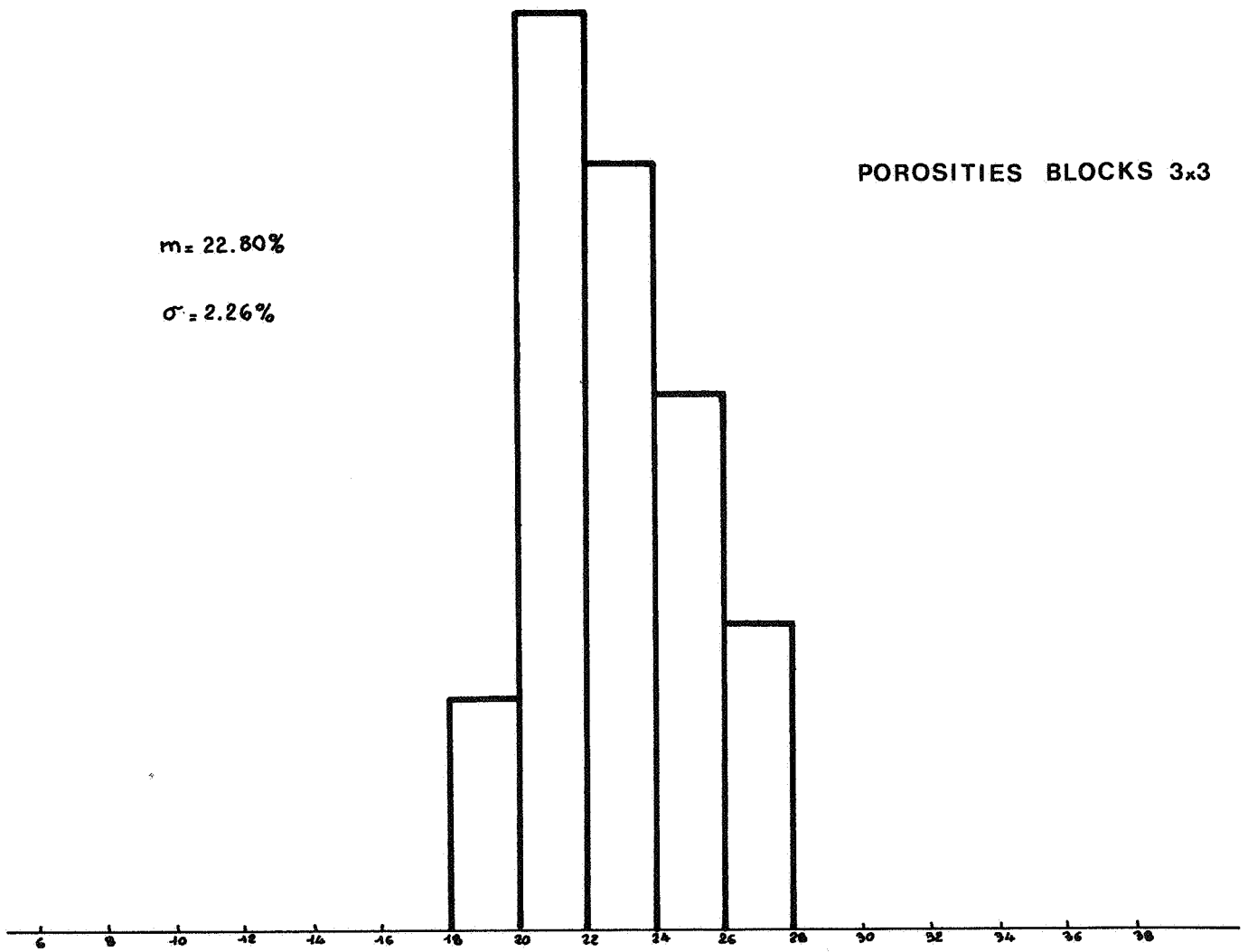


Fig:3

It can be shown that this variance is related to the variogram of $Z(x)$ by :

$$\sigma^2(0|V) = \frac{1}{V^2} \int_V dx \int_V \gamma(x-y) dy$$

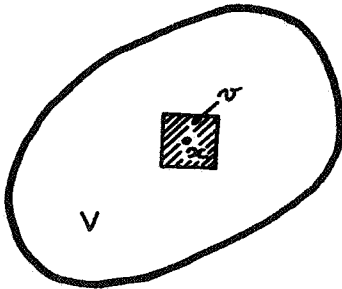
The integral is the average value of the variogram when x and y move independently within the volume V . For this reason we write it as $\bar{\gamma}(V,V)$. This mean value is also called the F function and has been graphed for simple geometrical shapes of V and several models of variograms (cf. Graphs 1 to 4). So we have :

$$\sigma^2(0|V) = \bar{\gamma}(V,V) = F(V)$$

2 - Variance of v within V .

Suppose now that we are considering a volume v and the Re.V. defined on that support :

$$Z_v(x) = \frac{1}{v} \int_v Z(x+u) du$$



We are interested in the dispersion of $Z_v(x)$ when it is moved within a larger domain V . For example v can be a section of a core and V the whole length of the core or a block, v can be a block and V the whole reservoir, etc...

The variance of v within V is denoted $\sigma^2(v|V)$ and is equal to

$$\sigma^2(v|V) = E \left[\frac{1}{v} \int_v (Z_v(x) - m_V)^2 dx \right]$$

In this case we obtain

$$\sigma^2(v|V) = \frac{1}{V^2} \int_V \int_V \gamma(x-y) dx dy - \frac{1}{v^2} \int_v \int_v \gamma(x-y) dx dy$$

i.e.

$$\sigma^2(v|V) = \bar{\gamma}(V,V) - \bar{\gamma}(v,v) = F(V) - F(v)$$

3 - Additivity Relationship.

Comparing the results of 1 and 2 we get :

$$\sigma^2(v|V) = \sigma^2(0|V) - \sigma^2(0|v)$$

that is :

$$\sigma^2(0|V) = \sigma^2(0|v) + \sigma^2(v|V)$$

This formula can be generalized to any of the three volumes v, V and V'

$$v \subset V \subset V' \Rightarrow \sigma^2(v|V') = \sigma^2(v|V) + \sigma^2(V|V')$$

Let for example
 v = core section
 V = block
 V' = reservoir

then the formula reads : "the variance of a core section within the reservoir is equal to the variance of a core section within a block plus the variance of the blocks within the reservoir". This formula shows that :

$$\sigma^2(\text{cores}|\text{reservoir}) > \sigma^2(\text{blocks}|\text{reservoir})$$

4 - Calculation of dispersion variances using the F function.

(a) Graph 1 gives the variance of a point within a rectangle of sides ℓ and h for a spherical model with a sill C = 1 and for distances given in range units $\frac{h}{a}$, $\frac{\ell}{a}$. So, to enter the graph, first convert your distances into range units, look for the value on the graph and then multiply by the value C of the sill.

$$\text{Ex. : if } \frac{\ell}{a} = 1, \frac{h}{a} = 0.3 \quad F(1, 0.3) = C \times 0.49$$

(b) Graph 2 gives the variance of a point within a parallelepiped with sides ℓ , h, h. Same reading as in (a).

(c) Graph 3 : analogous to (a) for the exponential model

$$\gamma(h) = C \left[1 - e^{-\frac{|h|}{a}} \right]$$

(d) Graph 4 : same as for (b).

REMARKS

- . A linear variogram $\gamma(h) = \omega|h|$ can be considered as a spherical one with an arbitrarily large range a and a sill C such that $\omega = \frac{3}{2} \frac{C}{a}$. To use the graphs, choose the range a to be large enough with respect to the dimensions ℓ and h ($\frac{\ell}{a}$ and $\frac{h}{a} < 0.4$). For square blocks in 2-D one can use

$$F(\ell, \ell) = 0.5214 \omega \ell$$

- . In case of **affine** anisotropy divide h and ℓ by the ranges in the corresponding directions.
Ex.: Suppose we are in 3-D with a spherical variogram. The horizontal ranges are 800 and 400 and the vertical range is 120. Suppose the block is oriented the same way with dimensions $200 \times 100 \times 50$. Then

$$\frac{h}{a} = \frac{200}{800} = \frac{100}{400} = 0.25 \quad \frac{\ell}{a} = \frac{50}{120} = 0.42$$

On graph 2 we read

$$\frac{1}{C} F(0.42, 0.25, 0.25) = 0.3$$

- . In the case of model in the set of several variograms we just add up the contributions of each of them.
- . If there is a nugget effect its contribution must be evaluated separately. To do so the easiest way is to refer to the white noise point model with covariance $A\delta$ defined in section II-5-2. The variance of averages over a support v , which represents $\sigma^2(v|\infty)$ is of the form :

$$\sigma^2(v|\infty) = \frac{A}{v}$$

($\sigma^2(o|v)$ is not defined in this model). Hence :

$$\sigma^2(v|V) = \sigma^2(v|\infty) - \sigma^2(V|\infty) = A \left[\frac{1}{v} - \frac{1}{V} \right]$$

The constant A is given by $C_o = A/v_o$ where v_o is the volume of the support of the data used to compute the variogram on which C_o was fitted. So, finally

$$\sigma^2(v|V) = C_o \left[\frac{v_o}{v} - \frac{v_o}{V} \right]$$

This formula also works in the discrete case where v_o and v are points and V is made up of N points : then $v_o/v = 1$ and $v_o/V = 1/N$ (case of measurement errors).

4 - CHANGE OF SUPPORT : REGULARIZATION.

Let $Z(x)$ be a point variable and consider a moving average :

$$Z_v(x) = \frac{1}{v} \int_v Z(x+u) du$$

$Z_v(x)$ is a new Re.V. with a support v , while the support of $Z(x)$ was a point. It may be interesting to express the variogram of $Z_v(x)$ as a function of the variogram of $Z(x)$. In order to do this the trick is to incorporate the domain of integration v inside the integral writing $Z_v(x)$ as

$$Z_v(x) = \frac{1}{v} \int Z(x+u) I(u) du$$

$I(u)$ being the indicator function of v . Then

$$Z_v(x+h) - Z_v(x) = \frac{1}{v} \int [Z(x+h+u) - Z(x+u)] I(u) du$$

and the variogram γ_v of Z_v is equal to :

$$\gamma_v(h) = \frac{1}{2} \frac{1}{v^2} \int du \int E[Z(x+h+u)-Z(x+u)][Z(x+h+u')-Z(x+u')] I(u) I(u') du'$$

Using the rule of II-4-1 one finds the covariance of the increments under the integral, namely :

$$- [2 \gamma(u'-u) - \gamma(h+u'-u) - \gamma(h+u-u')]$$

Setting $u'-u = t$ we get

$$\gamma_v(h) = \frac{1}{2} \frac{1}{v^2} \int du \int [\gamma(h+t) + \gamma(h-t) - 2 \gamma(t)] I(u) I(u+t) dt$$

Integrating in u first, defining the geometric covariogram of v :

$$P(t) = \int I(u) I(u+t) du$$

and taking the symmetry into account, one finds the final result :

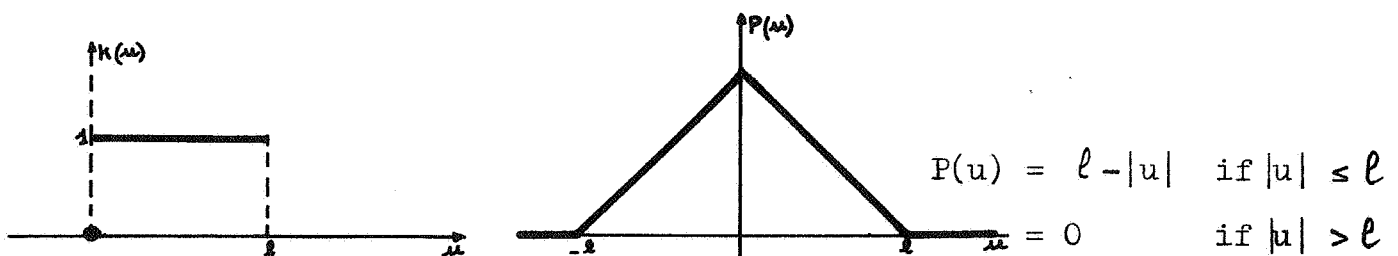
$$\gamma_v(h) = \frac{1}{v^2} \left[\int \gamma(h+t) P(t) dt - \int \gamma(t) P(t) dt \right]$$

It has the form of a convolution product

$$\gamma_v = \frac{1}{v^2} [\gamma * P - \text{constant}]$$

the constant equal to the value of $\gamma * P$ at $h = 0$ being there to ensure that $\gamma_v(0) = 0$.

As an example suppose we have measured porosities along a well on core sections small enough to be considered as points. We want to know the variogram $\gamma_\ell(h)$ of porosities averaged over a length ℓ . We have :



So

$$\gamma_\ell(h) = \frac{1}{\ell^2} \int_{-\ell}^{+\ell} (\ell - |u|) \gamma(h+u) du - \frac{1}{\ell^2} \int_{-\ell}^{+\ell} (\ell - |u|) \gamma(u) du$$

(we are working here in one dimension).

These integrals can be calculated directly but they can also be related to the F function defined above. For a linear variogram

$$\gamma(h) = \omega|h|$$

the results are :

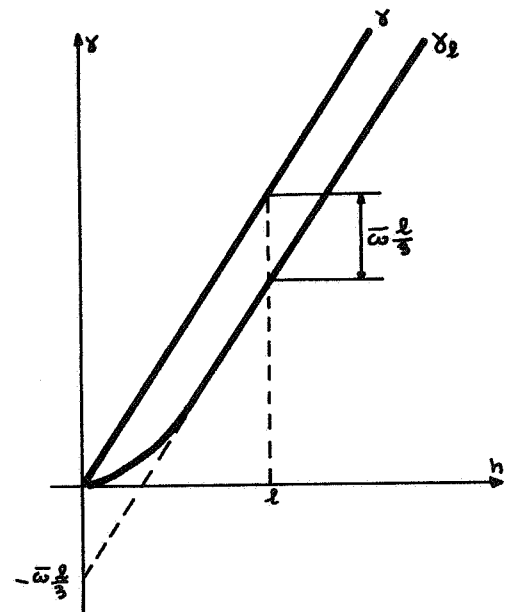
$$\underline{0 \leq h \leq \ell}$$

$$\gamma_\ell(h) = \omega \frac{h^2}{\ell^2} \left(\ell - \frac{h}{3} \right)$$

$$\underline{\ell < h}$$

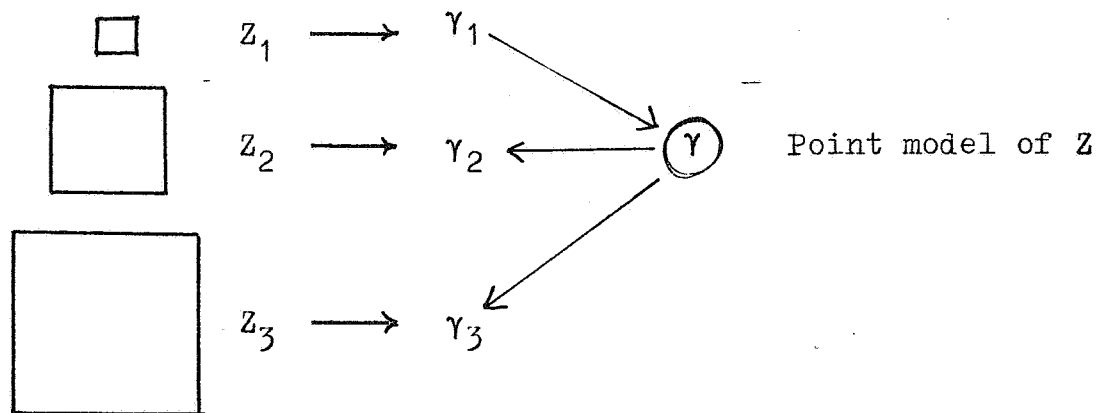
$$\gamma_\ell(h) = \omega \left(h - \frac{\ell}{3} \right)$$

We notice that γ_ℓ has a parabolic behavior at the origin : it is more regular than the initial variogram γ .



In 2 or 3 dimensions the formulae are more complicated and we may omit them in an introductory course.

So far we have seen how to go from a point support to a block support. Now often our basic data are already measured on a block support. The standard procedure to follow then is epitomized by the following sketch :



5 - COMING BACK TO OUR EXAMPLE.

The variograms of the 324 data have been computed along rows and columns and plotted on Fig. 4. The variograms of 2×2 blocks and 3×3 blocks have also been computed from Table II. We see how variograms are modified as the size of the support of the Re.V. increases and we understand why this is called a "regularization".

Now suppose our information is the variogram γ_{v_1} of the elementary blocks and with that we wish to predict the variances of the other blocks. Let us first give them names :

$$\begin{aligned} v_1 &= \text{elementary block} \\ v_2 &= \text{block } 2 \times 2 = 4 v_1 \\ v_3 &= \text{" } 3 \times 3 = 9 v_1 \\ v_6 &= \text{" } 6 \times 6 = 36 v_1 \\ V &= \text{" } 18 \times 18 = 324 v_1 = \text{whole domain} \end{aligned}$$

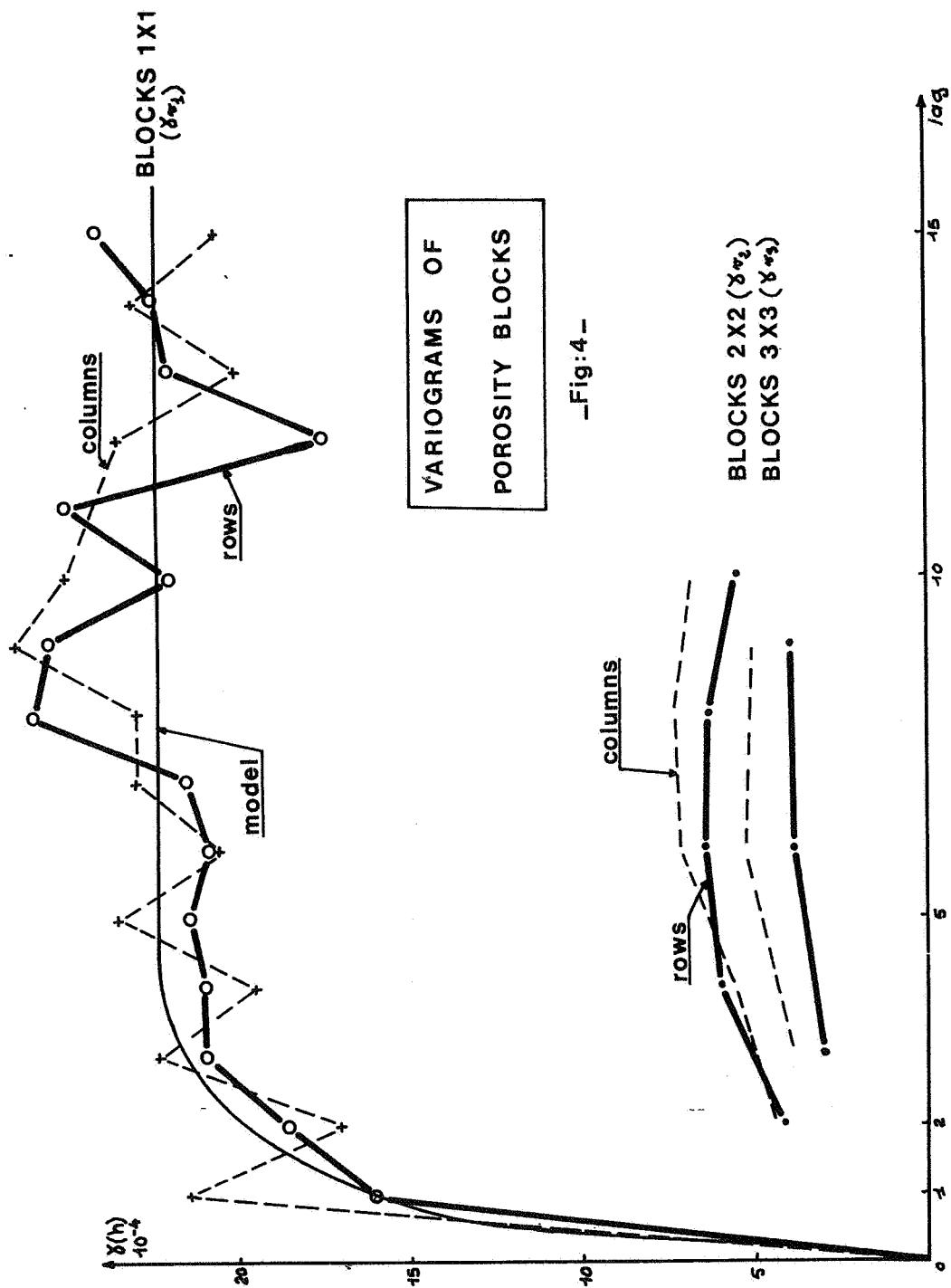
If we had a model for the point variogram γ we could use its F function and get simply :

$$\sigma^2(v_i|V) = F(V) - F(v_i)$$

In most practical situations the dimensions of the support on which our data are measured are negligible with respect to the dimensions of the volume we want to compute the dispersion in (plug, core, or log influence zone versus block or reservoir). As a consequence our data can be treated as point data and their variograms can be used without deconvolution.

Admittedly in our example this approximation is not valid, at least to compute $\sigma^2(v_2|V)$ since v_1 is only one fourth of v_2 . For v_3 the approximation seems more acceptable and we shall content ourselves with it to obtain numerical results without too much effort. The variogram can be roughly fitted to the sum of two spherical models :

$$\gamma(h) = \gamma_1(h) + \gamma_2(h)$$



-Fig:4 -

with

$$\begin{aligned} & \left\{ \begin{aligned} \gamma_1(h) &= 8 \left[\frac{3}{2} \frac{|h|}{4} - \frac{1}{2} \left(\frac{|h|}{4} \right)^2 \right] & |h| \leq 4 \\ &= 8 & |h| > 4 \end{aligned} \right. \\ & \text{(in } 10^{-4} \text{ units)} \end{aligned}$$

and

$$\begin{aligned} & \left\{ \begin{aligned} \gamma_2(h) &= 14.31 \left[\frac{3}{2} \frac{|h|}{1} - \frac{1}{2} \left(\frac{|h|}{1} \right)^2 \right] & |h| \leq 1 \\ &= 14.31 & |h| > 1 \end{aligned} \right. \\ & \text{(in } 10^{-4} \text{ units)} \end{aligned}$$

(The fit is only rough because it seems that along columns there is a hole effect not accounted for by our model).

We get (cf. Graph 1) :

$$\begin{aligned} \sigma^2(v_3|V) &= 8 \left[1 - \frac{1}{6} F\left(\frac{3}{4}, \frac{3}{4}\right) \right] + 14.31 \left[1 - \frac{1}{6} F\left(\frac{3}{1}, \frac{3}{1}\right) \right] \\ &= 8 [1 - 0.530] + 14.31 [1 - 0.940] \\ &= 4.62 \text{ (in } 10^{-4} \text{ units)} \end{aligned}$$

$$\begin{aligned} \sigma^2(v_6|V) &= 8 \left[1 - \frac{1}{6} F\left(\frac{6}{4}, \frac{6}{4}\right) \right] + 14.31 \left[1 - \frac{1}{6} F\left(\frac{6}{1}, \frac{6}{1}\right) \right] \\ &= 8 [1 - 0.810] + 14.31 [1 - 0.983] \\ &= 1.76 \text{ (in } 10^{-4} \text{ units)} \end{aligned}$$

It is interesting to compare these values with what a blunt application of a well known statistical rule would have given, namely dividing the variance by the number of elementary blocks as if they were independent (22.31 divided by 9 for v_3 and divided by 36 for v_6).

Variance in 10^{-4} units	Observed	Predicted by Geostatistics	Predicted by σ^2/n law
Blocks 3×3	5.11	4.62	2.48
Blocks 6×6	1.99	1.76	0.62

These results clearly show that it would be wrong to neglect spatial correlations.

6 - EXERCISE : REGULARIZATION BY CORES.

A core of length $L = 50$ has been cut up into sections of constant length ℓ (for convenience $\ell = 1$). The experimental half-variogram $\hat{\gamma}_1$ of the porosities of section $\ell = 1$ has been computed up to a distance $\frac{L}{2} = 25$. Now, we have seen that for variance calculations a model of the point variogram γ is needed and the purpose of this exercise is to show how it can be found and checked.

A simple procedure consists in choosing a priori a point model γ among the classical ones, compute its regularized version γ_ℓ and adjust the parameters of γ so as to obtain a good fit of γ to the experimental $\hat{\gamma}_\ell$. If possible, one can also group the core sections by lengths $2\ell, 3\ell, \dots$ compute the experimental variograms $\hat{\gamma}_{2\ell}, \hat{\gamma}_{3\ell}, \dots$ and compare them with the theoretical $\gamma_{2\ell}, \gamma_{3\ell}, \dots$ deduced by convolution of the point model γ . If the model is correct, the fit should be good also for these variograms.

The following table gives the values of the experimental half-variograms for core sections of length $\ell = 1, \ell = 3, \ell = 5$ and $\ell = 10$.

QUESTIONS

- 1°) Among the models you know, which one would you select as a model for the point variogram?
- 2°) Fit the parameters of this model using the experimental $\hat{\gamma}_1(h)$.
- 3°) Using graphs 5 and 6 given hereafter, find the theoretical values $\gamma_\ell(h)$ deduced by convolution of the point model and compare them to the experimental values. Discuss the discrepancies.

HINTS

- 1°) Look at the behavior of γ_1 at the origin ; see if there is a sill and how fast it is reached.

$$\hat{\gamma}_\ell(h) \text{ (in } 10^{-4} \text{ units)}$$

h	$\ell = 1$	$\ell = 3$	$\ell = 5$	$\ell = 10$
1	0.9			
2	2.3			
3	3.8	2.6		
4	4.9			
5	6.3		4.2	
6	7.2	6.5		
7	8.1			
8	9.6			
9	10.7	10.0		
10	12.0		9.3	6.0
11	12.0			
12	12.2	11.5		
13	14.0			
14	13.5			
15	13.9	12.7	12.0	
16	14.0			
17	13.2			
18	12.0	11.8		
19	12.9			
20	13.8		11.5	11.0
21	13.4	13.0		
22	13.4			
23	13.6			
24	13.2	12.7		
25	12.5		10.5	

- 2°) Is there a nugget effect? If a_1 is the range of γ_1 then $a = a_1 - \ell$ is the range of the point variogram γ . Why? To compute the sill of γ use the additivity relationship and the function $F(\ell) = \frac{\ell}{3}$ of a linear variogram.
- 3°) Graphs 5 and 5b give for $C = 1$ the values of $\gamma_\ell(\frac{h}{\ell})$ for $\frac{a}{\ell}$ ranging from 0.1 to 20 and for lags $\frac{h}{\ell}$ ranging from 0 to 25 or 13.

SOLUTION

1°) The inspection of the experimental half-variogram $\hat{\gamma}_1(h)$ will give us clues for the point model γ .

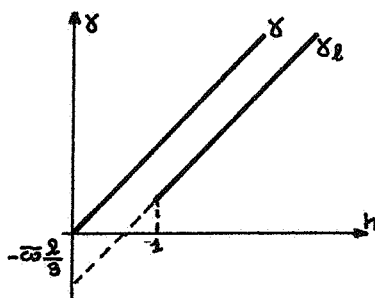
We first note that $\hat{\gamma}_1(h)$ has a linear shape from lag 1 to lag 3. Yet if we extrapolate this down to the y-axis we find a "negative nugget effect". This means that $\hat{\gamma}_1(h)$ has to bend towards the origin and reach it with a parabolic behavior. Recalling that the regularization of a linear variogram by a segment of length ℓ is parabolic for $|h| \leq \ell$ and linear with the same slope for $|h| > \ell$, we see that the point variogram we are after is linear at the origin.

Since $\hat{\gamma}_1(h)$ has a sill, so has $\gamma(h)$, and we are led to choose between a spherical and an exponential model. The fact that $\hat{\gamma}_1(h)$ reaches its sill rather rapidly argues in favor of the spherical model.

2°) If $\gamma(h)$ is linear, $\gamma_\ell(h)$ is also linear with the same slope for $|h| > \ell$. Thus we can estimate the slope of $\gamma(h)$ at the origin from the slope of $\hat{\gamma}_1(h)$ between $h = 1$ and $h = 3$. We find :

$$\omega_1 \approx 1.4$$

Now $\gamma(h)$ lies above $\gamma_\ell(h)$ by a translation of $\omega \frac{\ell}{3} \approx 0.47$.



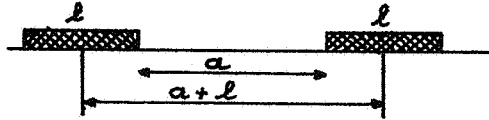
Thus we see on the plot of $\hat{\gamma}_1(h)$ that the variogram $\gamma(h)$ deduced from $\hat{\gamma}_1(h)$ by this translation would pass almost exactly through the origin. Hence there is no nugget effect.

Rough estimates of the range and the sill of $\hat{\gamma}_1(h)$ can be :

$$a_1 = 16 \quad \text{and} \quad C_1 = 13.5$$

This implies that the range of $\gamma(h)$ is $a = 16 - 1 = 15$. Indeed it is clear from the next figure that if a is the minimum distance above which there is no correlation between

two points, then two segments of length ℓ will be uncorrelated only if their closest end points are at least a apart, i.e., the two segments are $a + \frac{\ell}{2} + \frac{\ell}{2} = a + \ell$ apart.



To find the sill C we notice that

$$C = \text{variance of a core of length } \ell = \sigma^2(\ell/\infty)$$

$$\text{but } \sigma^2(\ell/\infty) = F(\infty) - F(\ell) = C - F(\ell)$$

$$\text{thus } C_1 = C - F(1)$$

where F is the F -function of the spherical variogram $\gamma(h)$. But for distances like $h = 1$ - much smaller than the range $a = 15$ - this spherical variogram is very well approximated by a linear variogram with a slope $\omega = \frac{3}{2} \frac{C}{a}$. Thus we have :

$$F(1) \simeq \omega \frac{\ell}{3} = \frac{3}{2} \frac{C}{a} \frac{1}{3} = \frac{C}{2a}$$

$$\text{so } C_1 = C - \frac{C}{2a} = C \left[1 - \frac{1}{2a} \right] \Rightarrow C = \frac{C_1}{1 - \frac{1}{2a}} = \frac{13.5}{1 - \frac{1}{30}} \simeq 14$$

(This result can be compared with the one obtained using the exact expression for the F -function of the spherical model, i.e.

$$F(\ell) = C \left[\frac{\ell}{2a} - \frac{\ell^3}{20 a^3} \right] \text{ for } |\ell| \leq a$$

Finally our point model is

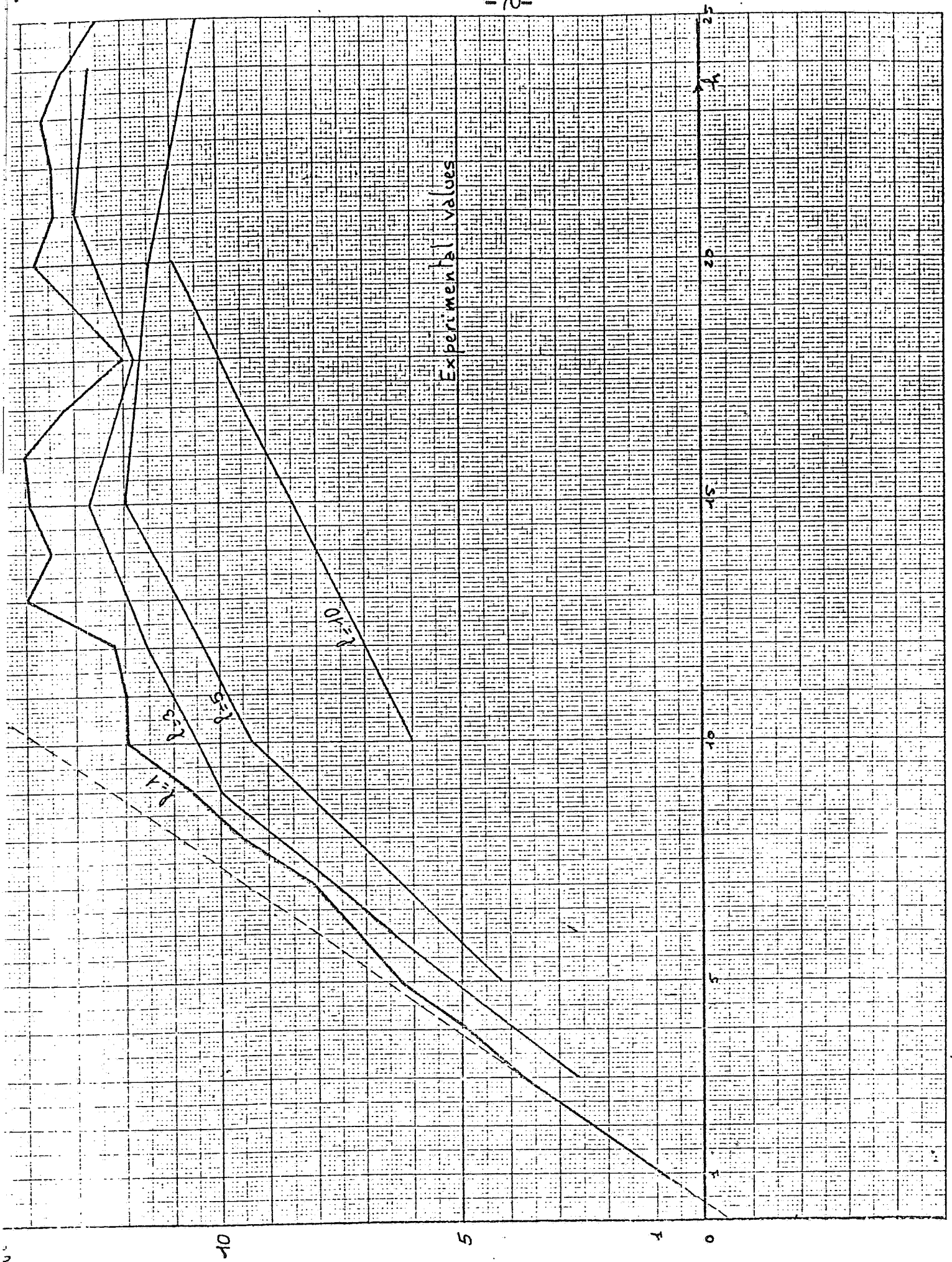
$$\gamma(h) = \begin{cases} 14 \left[\frac{3}{2} \frac{|h|}{15} - \frac{1}{2} \left(\frac{|h|}{15} \right)^3 \right] & |h| \leq 15 \\ 14 & |h| > 15 \end{cases}$$

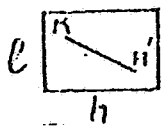
3°) Using graph 5 bis we find the following theoretical values of γ_1 , γ_3 , γ_5 and γ_{10} under the hypothesis that

our previously determined point model γ is correct.

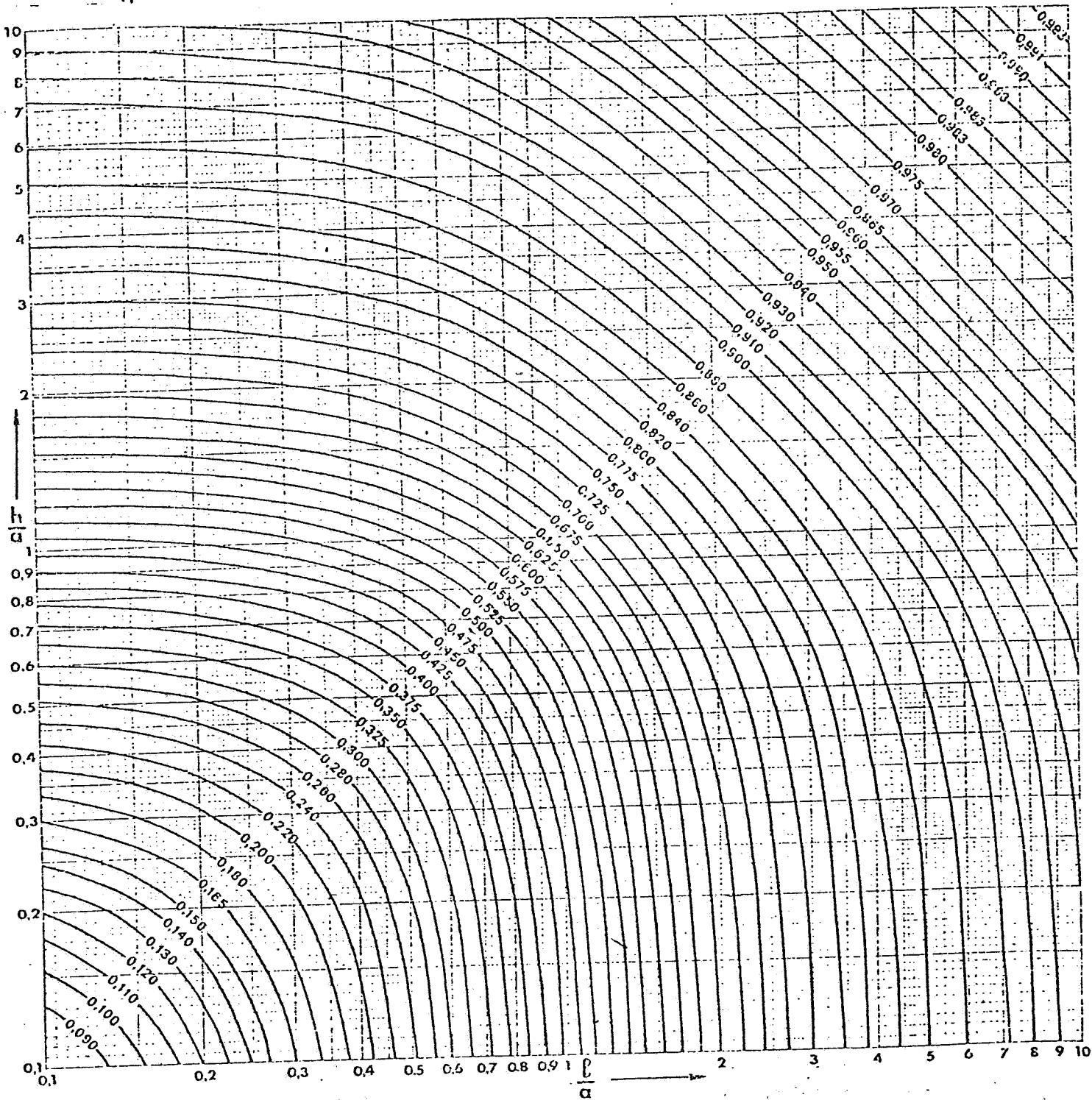
h	γ	γ_1	γ_3	γ_5	γ_{10}
1	1.40	0.91			
2	2.78	2.31			
3	4.14	3.71	2.73		
4	5.47	5.04			
5	6.74	6.3		4.34	
6	7.95	7.5	6.5		
7	9.09	8.6			
8	10.14	9.7			
9	11.09	10.6	9.7		
10	11.93	11.5		9.45	6.3
11	12.64	12.2			
12	13.22	12.7	11.8		
13	13.64	13.1			
14	13.91	13.4			
15	14.00	13.5	12.5	11.5	
16	↓	13.6			
17		↓			
18	↓		12.6		
19	S	↓	↓		
20	I	S		11.7	9.5
21	L	I	↓	↓	
22	L	L	S		
23		L	I	↓	
24			L	S	
25			L	I	
				L	
				L	

If we compare these values h with the experimental ones we notice a fair agreement for distances $h < 10 = \frac{L}{5}$, which is one fifth of the total core length. For $h > 15 = \frac{L}{3}$ the fluctuations of the experimental variogram may not be significant.

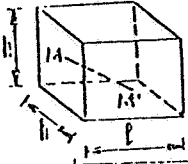




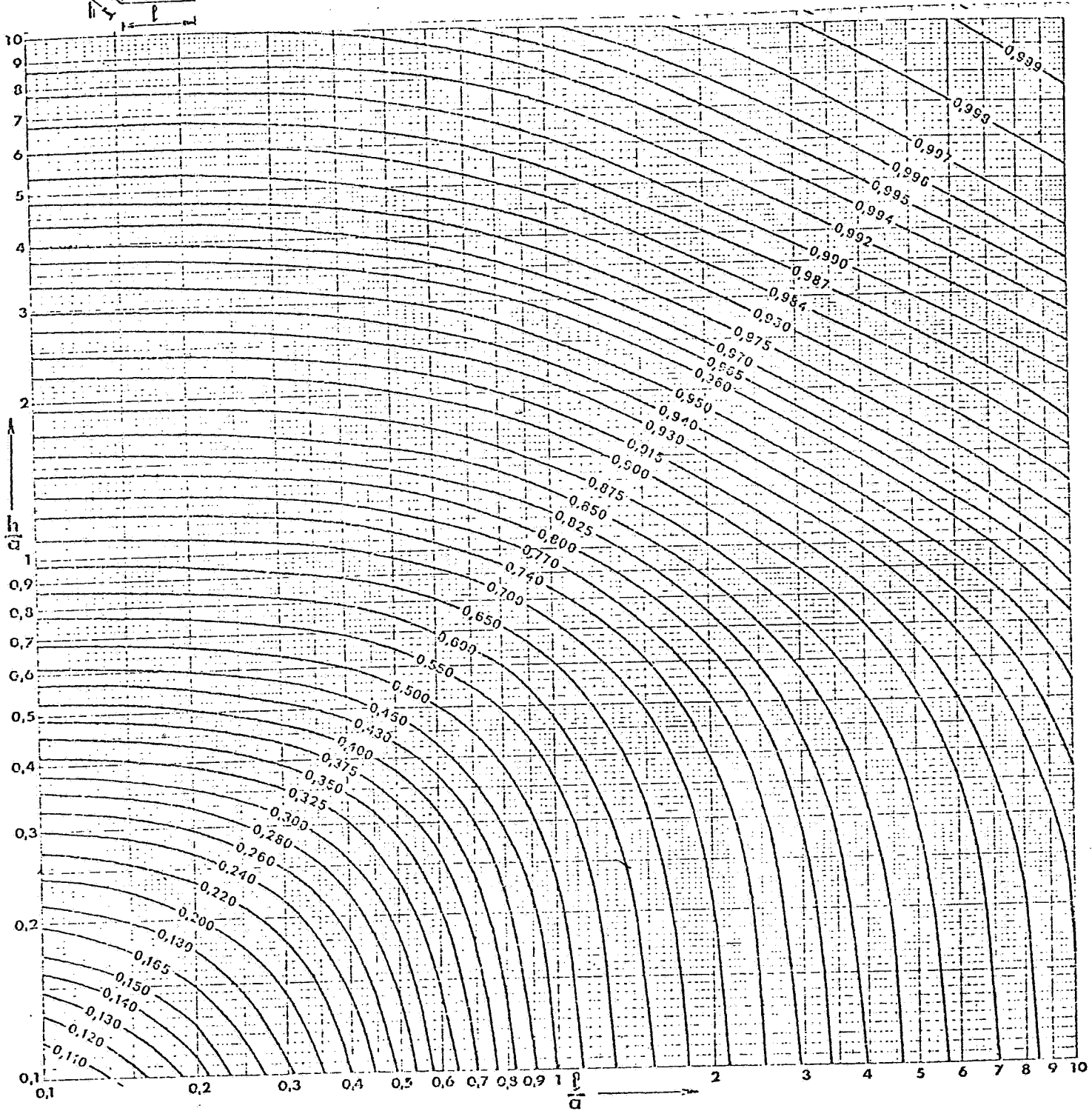
Function $\frac{1}{C} F\left(\frac{l}{a}, \frac{h}{a}\right)$ = variance of a point within the rectangle $l \times h$.



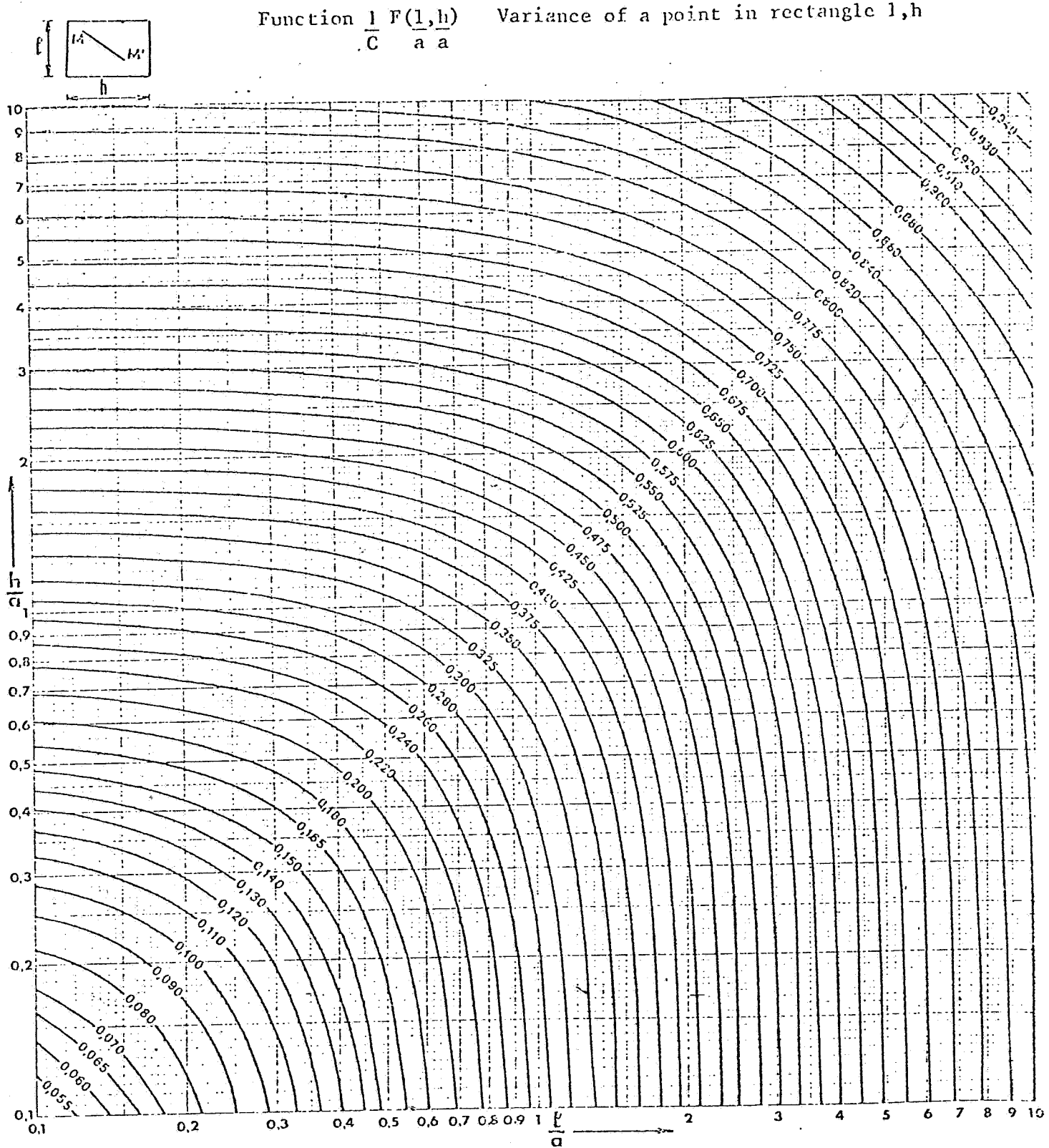
GRAPH 1 SPHERICAL VARIOGRAM



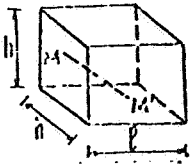
Function $\frac{1}{C} F\left(\frac{l}{a}, \frac{h}{a}\right)$ Variance of a point in parallelepiped l, h, a



GRAPH 2 SPHERICAL VARIOGRAM

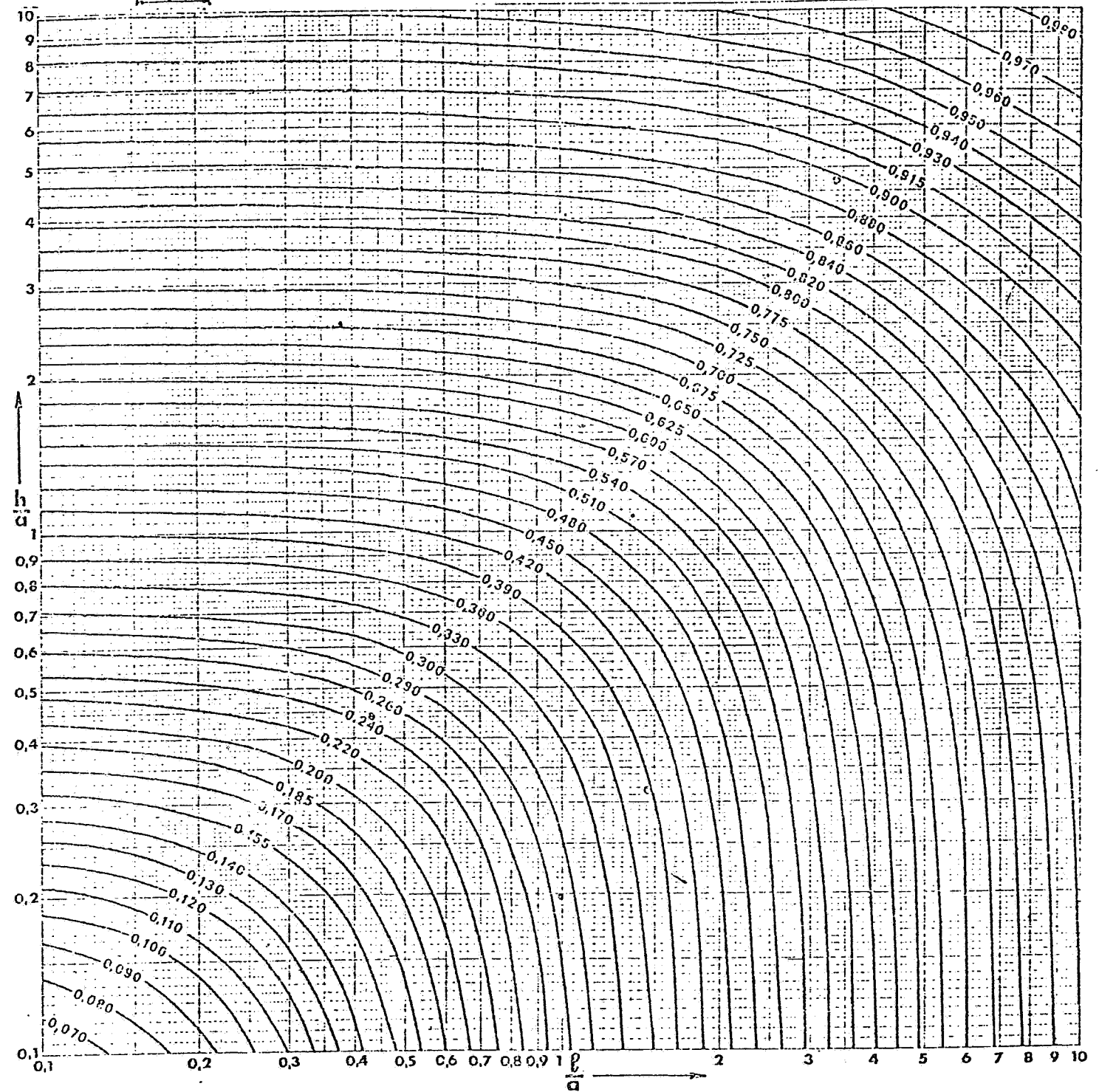


GRAPH 3 EXPONENTIAL VARIOGRAM



Function $\frac{1}{C} F\left(\frac{l}{a}, \frac{h}{a}, \frac{h}{a}\right)$

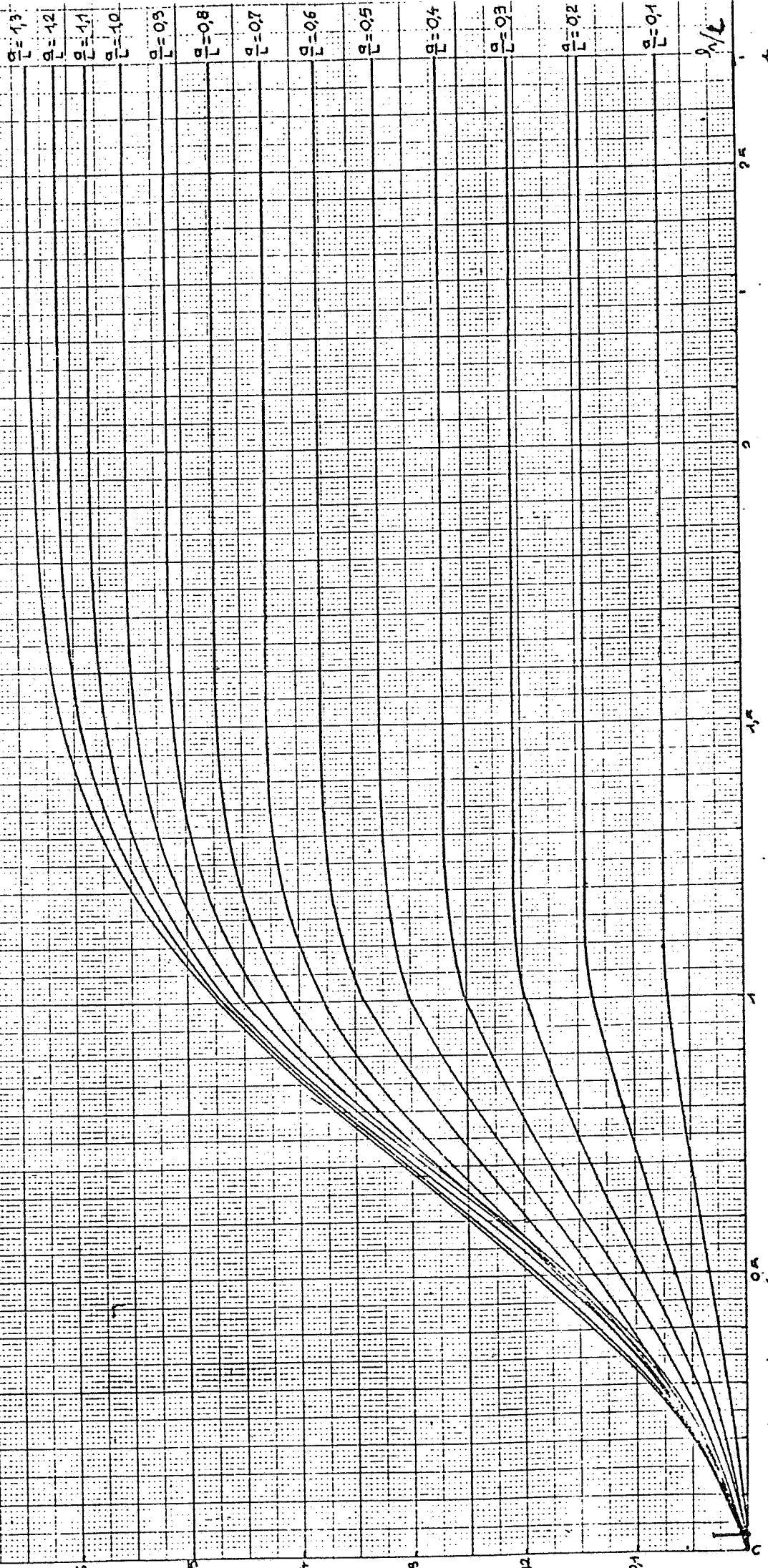
Variance of a point in parallelepiped l, h, h

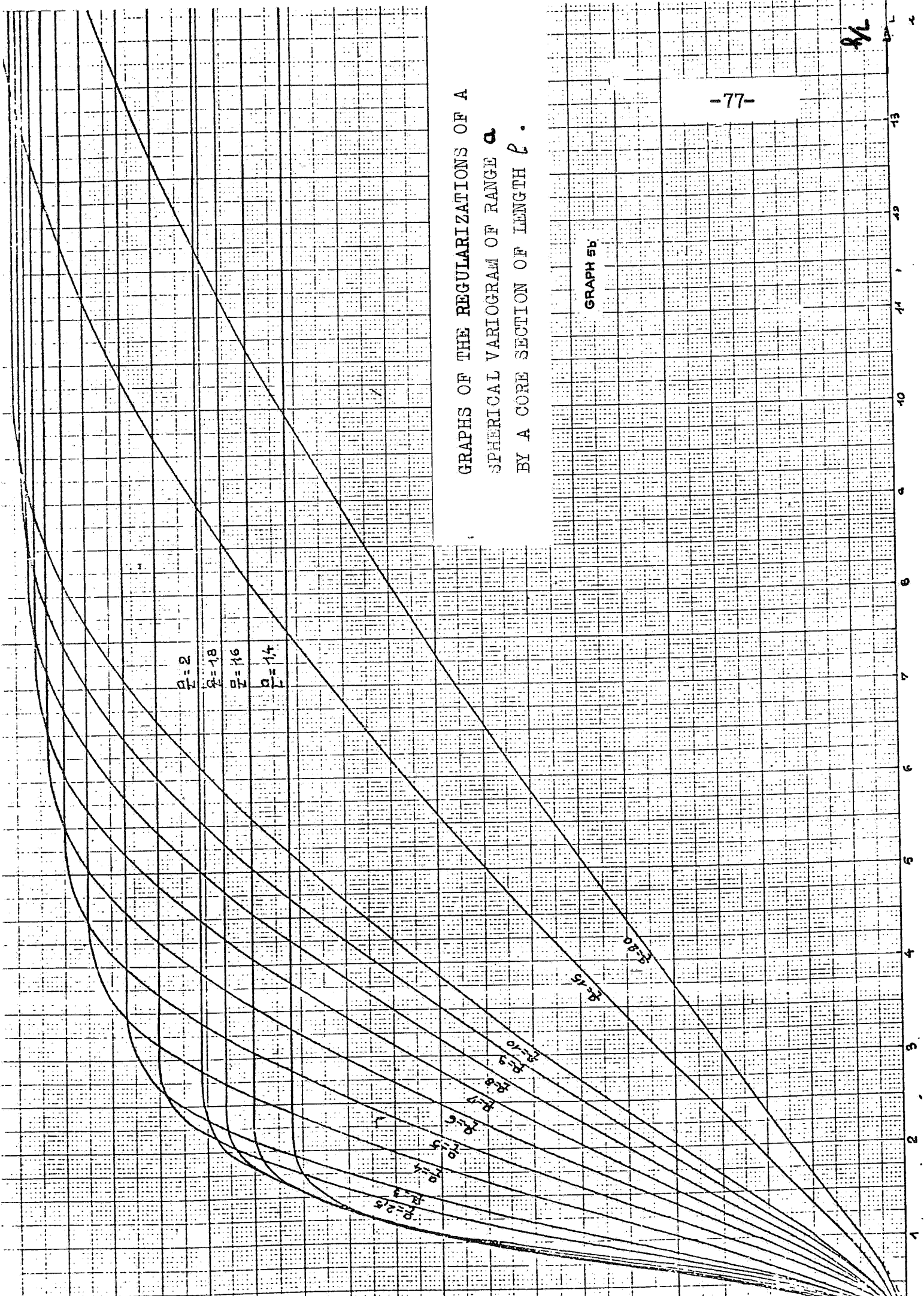


GRAPH 4 EXPONENTIAL VARIOGRAM

GRAPH 5

Graphs of the regularizations of a spherical variogram of range a by a core section of length c .





$\frac{a}{L} = 2$

$\frac{a}{L} = 1.8$

$\frac{a}{L} = 1.6$

$\frac{a}{L} = 1.4$

GRAPHS OF THE REGULARIZATIONS OF A
SPHERICAL VARIOGRAM OF RANGE a
BY A CORE SECTION OF LENGTH ℓ .

GRAPH 5b

8/2

- CHAPTER IV -

VARIANCES OF EXTENSION AND THEIR APPLICATIONS TO

ESTIMATION AND SAMPLING PATTERNS

1 - THE CONCEPT OF VARIANCE OF EXTENSION AND ITS RELATIONSHIP TO THE VARIANCE OF DISPERSION.

Suppose that we want to assess the average value of a parameter $Z(x)$ over a given domain V of the field, i.e. the integral :

$$Z(V) = \frac{1}{V} \int_V Z(x) dx$$

Now, the information is available only on a domain v - for example V can be a block and v a drill-hole ; or V can be a core section and v a plug, etc.

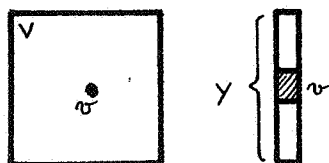


Fig. 1

So, we actually know the value of

$$Z(v) = \frac{1}{v} \int_v Z(x) dx$$

In many cases a natural step is to simply take $Z(v)$ as an estimate of $Z(V)$. What error do we make?

First of all, under the assumption that $Z(x)$ is intrinsic, $Z(v)$ is an unbiased estimator of $Z(V)$:

$$E[Z(v)] = \frac{1}{V} \int_V E(Z(x)) dx = \frac{1}{V} \int_V m dx = m = E[Z(V)]$$

$$\text{Then : } E[Z(v) - Z(V)]^2 = \text{Var}(Z(v) - Z(V)) = \sigma_E^2(v, V)$$

The variance of $Z(v) - Z(V)$ is called the "variance of extension" of v to V . It is the variance of the error that we incur in "extending" to the domain V the average value measured on the domain v . This variance is sometimes denoted by $\sigma_E^2(v, V)$ or σ_E^2 for short. (Note that $\sigma_E^2(v, V)$ is symmetric in v and V).

Conceptually, $\sigma_E^2(v, V)$ is simply the variance of estimation of Z_V by Z_v . In geostatistical practice however, the term "extension variance" is preferred in reference to elementary cases when sample values are extended to their "zones of influence". The expression "estimation variance" is used for more general situations where several samples are combined to estimate a given quantity.

The theoretical value of the variance of extension is given by :

$$\sigma_E^2(v, V) = \frac{2}{vV} \int_V dx \int_v \gamma(x-y) dy - \frac{1}{V^2} \int_V dy \int_V \gamma(y-y') dy' - \frac{1}{v^2} \int_v dx \int_v \gamma(x-x') dx'$$

Or, in a more concise form :

$$\sigma_E^2(v, V) = 2 \bar{\gamma}(v, V) - \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$$

where $\bar{\gamma}(V, V) = F(V)$ and $\bar{\gamma}(v, v) = F(v)$, i.e. the average value of the variogram when the end points x and x' (or y and y') are sweeping independently throughout V and v (cf. Chap. III).

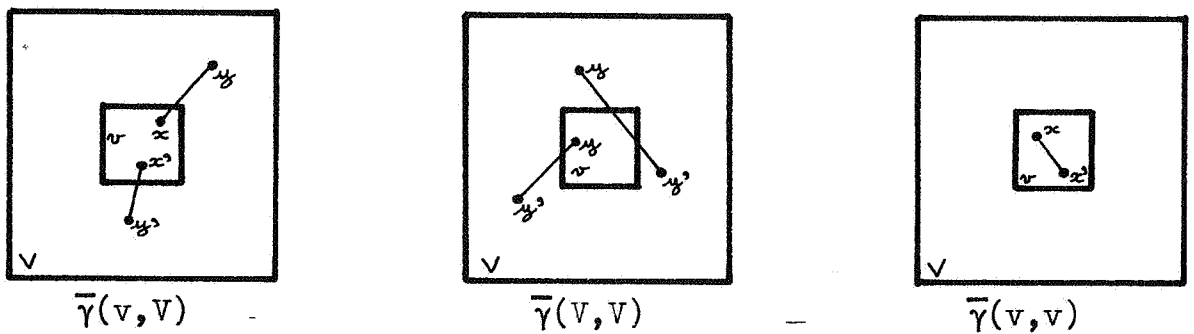


Fig. 2

The above formula holds for any shape of v and V ; in particular v need not be included into V . The factors influencing the variance of extension are seen to be :

- (i) the regularity of the variable : through γ
- (ii) the geometry of V : through $\bar{\gamma}(V, V)$
- (iii) the geometry of v : through $\bar{\gamma}(v, v)$
- (iv) the location of v with respect to V : through $\bar{\gamma}(v, V)$

Furthermore, rewriting the formula as :

$$\sigma_E^2(v, V) = [\bar{\gamma}(v, V) - \bar{\gamma}(V, V)] + [\bar{\gamma}(v, V) - \bar{\gamma}(v, v)]$$

makes it clear that the variance decreases as :

- the sampling v is more representative of the domain V to be estimated (In the limit, when $v = V$: $\sigma_E^2(v, V) = 0$)
- the variogram is more regular, i.e. the Re. V . is more continuous in its variations (cf. Exercice p. IV-15).

Also, an obvious but important property of variances of extensions is that they involve only the variogram and the geometry of the problem, but not the actual values taken by the variable under study. This fact, which is a consequence of considering second order models, is the basis of the design of measurement networks.

Relationship to the variance of dispersion.

Even when v is included into V the extension variance $\sigma_E^2(v, V)$ should not be confused with the dispersion variance $\sigma^2(v|V)$ presented in Chap. III. $\sigma^2(v|V)$ has a physical significance : it measures the dispersion of samples of size v within the domain of definition V . The extension variance for its part, is an operational concept, characterizing the error attached to a given sampling pattern. Theoretically, the two types of variances are related in the following way : the dispersion variance $\sigma^2(v|V)$ is the average of extension variances $\sigma_E^2(v, V)$ when the sample v takes all possible positions within V .

To see this, consider first the case when v is reduced to a point :

$$\sigma^2(0|V) = E \left(\frac{1}{V} \int_V (Z(x) - m_V)^2 dx \right) = \frac{1}{V} \int_V E(Z(x) - m_V)^2 dx$$

But $E(Z(x)-m_V)^2$ is the extension variance of x to V , so that :

$$\sigma^2(O|V) = \frac{1}{V} \int_V \sigma_E^2(x, V) dx$$

This result can also be obtained from the explicit expression of $\sigma_E^2(x, V)$:

$$\sigma_E^2(x, V) = \frac{2}{V} \int_V \gamma(x-y) dy - \frac{1}{V^2} \int_V dy \int_V \gamma(y-y') dy'$$

so that :

$$\begin{aligned} \frac{1}{V} \int_V \sigma_E^2(x, V) dx &= \frac{2}{V^2} \int_V dx \int_V \gamma(x-y) dy - \frac{1}{V^2} \int_V \gamma(y-y') dy' \\ &= \frac{1}{V^2} \int_V dy \int_V \gamma(y-y') dy' = F(V) \end{aligned}$$

Similarly, if we consider that V is the union of domains v_i

$$V = \bigcup_{i=1}^N v_i$$

then

$$\sigma^2(v|V) = \frac{1}{N} \sum_{i=1}^N E[Z_{v_i} - m_V]^2 = \frac{1}{N} \sum_{i=1}^N \sigma_E^2(v_i, V)$$

which is the announced result.

2 - ELEMENTARY VARIANCES OF EXTENSION.

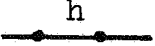
One dimension.

Let us first introduce the auxiliary functions χ and F

$$\chi(h) = \frac{1}{h} \int_0^h \gamma(u) du \quad \begin{array}{c} h \\ \text{---} \end{array} \quad \begin{array}{l} \text{average value of } \gamma \text{ when} \\ \text{origin fixed and end point} \\ \text{moves along segment } h \end{array}$$


$$F(h) = \frac{2}{h^2} \int_0^h u \chi(u) du$$

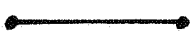
$$= \frac{2}{h^2} \int_0^h (h-u) \gamma(u) du$$



average value of γ when
two end points move inde-
pendently along segment
h.

Then, the extension variances of the following sampling pattern are :


 $\sigma_E^2 = 2 \chi(l/2) - F(l)$


 $\sigma_E^2 = 2 \chi(l) - F(l) - \frac{1}{2} \gamma(l)$

Example : The average porosity of a 1-foot core has been assessed by means of a central plug. Up to 1 foot the variogram can be modeled as : $\gamma(h) = 1.5 \cdot 10^{-4} |h|$. What is the accuracy of our assessment?

Assuming the dimensions of the plug to be negligible, we just apply the above formula. In this case when $\gamma(h) = \omega h$ we get :

$$\chi(h) = \omega \frac{h}{2} \quad \text{and} \quad F(h) = \omega \frac{h}{3} \quad \text{so that} \quad \sigma_E^2 = \omega \frac{h}{6}$$

Numerically :

$$\sigma_E^2 = 0.25 \cdot 10^{-4} \quad \text{so} \quad \sigma_E \sim 0.5\%$$

Higher dimensions - Graphs.

As in the 1-D case auxiliary functions can be defined. However, it is easier to use graphs, which cover most cases. The others can be treated by numerical integration on computer.

Graphs 1, 2 and 3 are typical examples for the spherical variogram. Their use is identical to that of the F-function. Remember that graphs are given for standardized variograms with $C = 1$ and that distances are in range units.

When a nugget effect C_0 is there, add to σ_E^2 the quantity

$$C_0 \left[\frac{v_0}{v} - \frac{v_0}{V} \right]$$

where v and V are the estimating and estimated volumes, respectively, and v_o the volume of basic sample data. By "basic" samples we mean samples having the same support as those which were used to compute the variogram, and therefore to fit C_o . When the support is a point $\frac{v_o}{v} = \frac{1}{N}$ (N number of sample points) and $\frac{v_o}{V} = 0$ (unless V is itself discrete).

EXAMPLES

1 - For the porosity data of the previous example, another experimenter fitted a spherical variogram with $a = 10'$ and $C = 10^{-3}$. Reevaluate σ_E^2 .

Using Graph 1 with $\frac{\ell}{a} = \frac{1}{10} = 0.1$ we read on the curve $\sigma_{E_1}^2$ the value 0.0255 so that :

$$\sigma_E^2 = 10^{-3} \times 0.0255 = 0.255 \cdot 10^{-4} \text{ and } \sigma_E \sim 0.5\%$$

We find roughly the same answer as before. No wonder : at small distances the spherical model is linear with slope $\omega = \frac{3}{2} \frac{C}{a} = 1.5 \cdot 10^{-4}$, which is precisely the model used before.

2 - The average porosity of a square cross sectional block ($100' \times 300' \times 300'$) was assessed by a central vertical core ($100'$ long). The variogram was computed using half-foot long core samples and was fitted to a spherical model with the following parameters :

$$a = 30' \quad C = 20 \cdot 10^{-4} \quad C_o = 10^{-4}$$

Assuming the variogram is isotropic in 3-D, compute the variance of extension of the core to the block.

For

$$\frac{h}{a} = \frac{300}{30} = 10 \quad \text{and} \quad \frac{\ell}{a} = \frac{100}{30} = 3.33$$

we read on graph 3 the value 0.21. Also, $100'$ of core length represent 200 basic half-foot samples. Thus :

$$\sigma_E^2 = 0.21 \times 20 \cdot 10^{-4} + \frac{10^{-4}}{200} = 4.2 \cdot 10^{-4}$$

i.e. $\sigma_E = 2\%$

We have neglected the term $C_0 \frac{v_0}{V}$; in fact the other term $C_0 \frac{v_0}{V}$ is also negligible.

3 - APPROXIMATION PRINCIPLE.

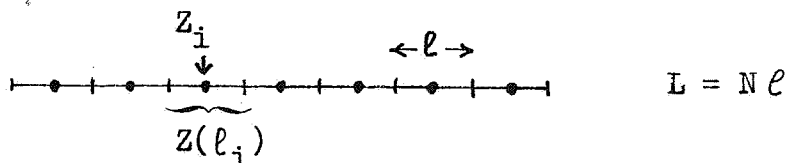
Consider now the practical problem of estimating an average value $Z(N)$ by means of N samples Z_1, Z_2, \dots, Z_N . Suppose we choose $Z(v) = \frac{1}{N} \sum_i Z_i$ as an estimate of $Z(V)$. What is the variance of estimation ?

A rigorous answer is readily provided by the general formula for extension variances, namely here :

$$\sigma_E^2 = \frac{2}{NV} \sum_i \int_V \gamma(x_i - y) dy - \frac{1}{V^2} \int_V dy \int_V \gamma(y - y') dy' - \frac{1}{N^2} \sum_i \sum_j \gamma(x_i - x_j)$$

But if N is large, this formula is quite cumbersome. Instead, we shall appeal to an approximation principle which postulates the absence of correlation between extension errors relative to disjoint domains.

Let us look for example at the following 1-D problem.



There are N samples centrally located in segments of length ℓ . Let $Z(\ell_i)$ be the true average over segment i . Then, the average over L is simply :

$$Z(L) = \frac{1}{N} \sum_i Z(\ell_i)$$

If Z_i is the value assumed by sample i and if we take the average of these $\hat{Z}(L) = \frac{1}{N} \sum_i Z_i$ as an estimate of $Z(L)$, the error is

$$\hat{Z}(L) - Z(L) = \frac{1}{N} \sum_i [Z_i - Z(\ell_i)]$$

i.e. the average of partial errors $Z_i - Z(\ell_i)$.

The approximation principle consists in assuming that these partial errors are uncorrelated. It can be proved that, for usual variogram models, this approximation is quite good. Then :

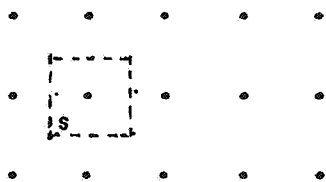
$$\sigma_N^2 = \text{Var}[\hat{Z}(L) - Z(L)] = \frac{1}{N} \sum_i \text{Var} [Z_i - Z(\ell_i)]$$

But $\text{Var} [Z_i - Z(\ell_i)] = \sigma_E^2(0, \ell)$ so that :

$$\sigma_N^2 = \frac{1}{N} \sigma_E^2(0, \ell) = \frac{1}{N} (2 \chi(\ell/2) - F(\ell))$$

The estimation variance is simply computed by dividing the elementary extension variance of each sample into its zone of influence by the number of samples N.

In 2 or 3-D the same method is applied. In the case of a regular grid :



$$\sigma_N^2 = \frac{1}{N} \sigma_E^2(0, s)$$

EXAMPLE

200 samples were taken along a core at a 1-foot interval. The average porosity of these samples and their standard deviation were found to be :

$$\bar{\Phi} = 16.2\%$$

$$\sigma = 4.47\%$$

Compute the variance of $\bar{\Phi}$:

(i) by the σ^2/n rule

(ii) by geostatistics, knowing that γ is spherical with $a = 10'$ and $C = 10^{-3}$.

$$(i) \quad \frac{\sigma^2}{n} = \frac{20 \cdot 10^{-4}}{200} = 10^{-5} \Rightarrow \sigma_N = 0.316\%$$

(ii) From the previous examples we know that :

$$\sigma_E^2(0, \rho) = 0.255 \cdot 10^{-4} \quad \text{so} \quad \sigma_N^2 = \frac{0.255 \cdot 10^{-4}}{200} = 0.0127 \cdot 10^{-5}$$

and $\sigma_N = 0.036\%$.

To our surprise we find a variance of estimation smaller than that given by the $\frac{\sigma^2}{n}$ rule by a factor of a 100! This is due to the fact that, with the adopted variogram model, a central sample is a very good representative of its 1-foot long influence zone.

The composition principle.

In more complicated situations the calculation of a variance of estimation must be broken down into several variances of extension calculations, and some weighting may be required.

Suppose for example that we want to assess the average oil porosity of an homogeneous reservoir (i.e. the volume of oil in place), on the basis of well measurements. The wells have been drilled on a regular pattern, but they have unequal depths d_i .

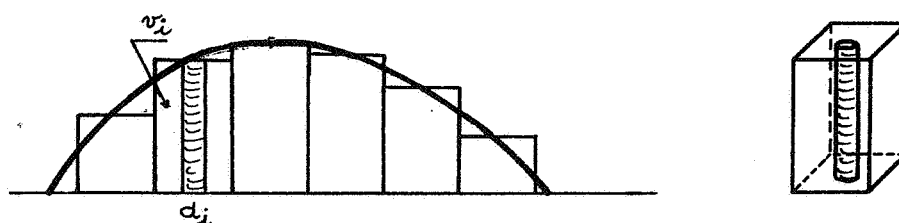


Fig. 5

Let

$Z(V)$ = true average oil porosity over the reservoir

$Z(v_i)$ = true average oil porosity over block v_i

$Z(d_i)$ = true average oil porosity over core length d_i

Assume, at first, that the cores have been thoroughly analyzed, so that we know $Z(d_i)$ exactly. Clearly, to estimate the average over V we need to weigh each $Z(d_i)$ by the volume V_i of its block of influence. As the cross-sectional area is a constant, this

amounts to weighting by the length d_i :

$$\hat{Z}(V) = \frac{\sum_i V_i Z(d_i)}{\sum_i V_i} = \frac{\sum_i d_i Z(d_i)}{\sum_i d_i}$$

Now $Z(V)$ itself can be written as :

$$Z(V) = \frac{\sum_i V_i Z(V_i)}{\sum_i V_i} = \frac{\sum_i d_i Z(V_i)}{\sum_i d_i}$$

The error is therefore a weighted average of elementary extension errors $Z(d_i) - Z(V_i)$:

$$\hat{Z}(V) - Z(V) = \frac{\sum_i d_i [Z(d_i) - Z(V_i)]}{\sum_i d_i}$$

By the approximation principle we get :

$$\sigma_E^2 = \text{Var} [\hat{Z}(V) - Z(V)] = \frac{\sum_i d_i^2 \sigma_{E_i}^2}{(\sum_i d_i)^2}$$

where $\sigma_{E_i}^2 = \text{Var} [Z(d_i) - Z(V_i)]$, i.e. the elementary extension variance of a central bore-hole to its block of influence (cf. Graph 3).

Note that these variances are weighted by the squares of the lengths d_i .

More realistically now, the average core porosities $Z(d_i)$ are themselves estimated by means of samples. To simplify matters, we admit that these were taken at regular intervals of length ℓ . So, each estimate $\hat{Z}(d_i)$ is the average of $n_i = \frac{d_i}{\ell}$ samples :

$$\hat{Z}(d_i) = \frac{1}{n_i} \sum_k Z_{i_k}$$

The total estimation error can be split into two terms :

$$\hat{z}(V) - z(V) = \underbrace{\frac{\sum_i d_i [\hat{z}(d_i) - z(d_i)]}{\sum_i d_i}}_{\text{Extension of samples to cores}} + \underbrace{\frac{\sum_i d_i [z(d_i) - z(V_i)]}{\sum_i d_i}}_{\text{Extension of cores to blocks}}$$

Denoting by $z(\ell_{i_k})$ the true average oil porosity of the k^{th} section of the i^{th} core we can rewrite the first term as :

$$\frac{\sum_i \frac{d_i}{n_i} \sum_k [z_{i_k} - z(\ell_{i_k})]}{\sum_i d_i}$$

As usual, the elementary extension errors are considered uncorrelated so that variances add up. We finally get :

$$\sigma_E^2 = \frac{1}{(\sum d_i)^2} \sum_i d_i^2 \frac{\sigma^2(0, \ell)}{n_i} + \frac{\sum d_i^2 \sigma_{E_i}^2}{(\sum d_i)^2}$$

$$\sigma_E^2 = \frac{\sigma^2(0, \ell)}{N} + \frac{\sum_i d_i^2 \sigma_{E_i}^2}{(\sum d_i)^2}, \quad \text{where } N = \sum_i n_i = \text{total number of core samples}$$

The total variance of estimation is the composition of a "line term" $\frac{\sigma^2(0, \ell)}{N}$ - the variance of the error made when extending the samples to the cores - and a "block term", accounting for the extension of the cores to their blocks of influence.

4 - SAMPLING PATTERNS.

The preceding theory enables us to compare the efficiencies of 3 usual sampling patterns.

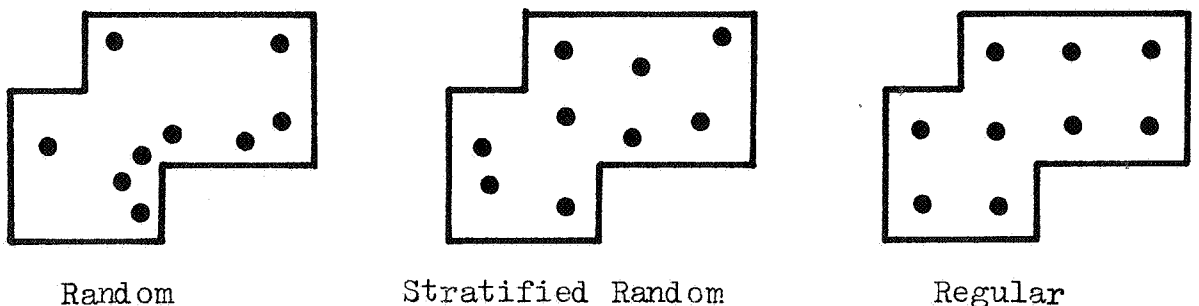


Fig. 4

Random Network.

In order to estimate the average value $Z(V)$ over V we take the average of N samples $Z(x_i)$ randomly scattered within V . More specifically, the x_i have been independently located in V with uniform density $1/V$. We wish to evaluate the variance of the error.

An elegant way to derive the result is the following. For a given realization $Z(x) = z(x)$ the values $z(X_i)$ are random through X_i and are independent. Furthermore,

$$E[\hat{Z}(V) - z(V)] = 0 \quad \text{and}$$

$$\begin{aligned} E[\hat{Z}(V) - z(V)]^2 &= \text{Var} [\hat{Z}(V) - z(V)] = \frac{1}{N^2} \sum_i \text{Var} [z(X_i) - z(V)] \\ &= \frac{1}{N} s^2(C|V) \end{aligned}$$

Randomizing the realization we find : $\sigma_E^2 = \frac{1}{N} \sigma^2(C|V)$

Stratified Random Network.

This time V is divided into N similar disjoint zones of influence v_i . Within each v_i a sample is placed at random with uniform density and independently of other samples.

The error is

$$\hat{Z}(V) - Z(V) = \frac{1}{N} \sum_i [Z_i - Z(v_i)]$$

and by an argument similar to that just seen, the partial errors $Z_i - Z(v_i)$ are conditionally independent and :

$$\sigma_E^2 = \frac{1}{N} \sigma^2(O|v)$$

The variance of estimation has the same form as in the purely random case, except that $\sigma^2(O|V)$ is replaced by $\sigma^2(C|v)$, i.e. the variance of a point sample within its zone of influence v .

Now, by the additivity relationship of Chap. III :

$$\sigma^2(0|V) - \sigma^2(0|v) = \sigma^2(v|V) \geq 0$$

which proves that the stratified random pattern is always more efficient than the purely random one.

Regular Grid.

We have seen that in this case the variance of estimation can be approximated by $1/N$ times the variance of extension of a sample to its zone of influence :

$$\sigma_E^2 = \frac{1}{N} \sigma_E^2(0, v)$$

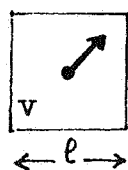
for the usual variogram model $\sigma_E^2(0, v)$ is smaller than $\sigma^2(0|v)$. To get order of magnitude information, let us consider the 2-D case when v is a square of side ℓ .

For the linear variogram $\gamma(h) = \omega h$ it was stated in Chap. III that :

$$F(\ell, \ell) = 0.5214 \omega \ell (= \sigma^2(0|v))$$

Recall now that :

$$\sigma^2(0, v) = 2 \bar{\gamma}(0, v) - \bar{\gamma}(v, v)$$



$$\bar{\gamma}(0, v) = \omega \times \frac{\ell}{3} \times \frac{1}{2} [\sqrt{2} + \text{Log}(1 + \sqrt{2})] = 0.3826 \omega \ell$$

$$\text{so that } \sigma_E^2(0, v) = 0.2438 \omega \ell$$

The ratio

$$\frac{\sigma^2(0|v)}{\sigma_E^2(0, v)} = 2.1387$$

shows that for a linear variogram the regular grid pattern is twice more efficient than the random stratified pattern itself more efficient than the purely random scheme .

In order to have a complete comparison between the sampling patterns, let us further assume that the domain V is itself a square of area S . Then the variance of a sample within V is :

$$\sigma^2(0|V) = F(L, L) = 0.5214 \times \omega \times L \quad \text{with } L = \sqrt{S}$$

for the stratified random and the regular network each square block has a side of length $\ell = \sqrt{\frac{S}{N}}$ and we can express the estimation variances in term of $\sigma^2 = \sigma^2(0|V)$ and the total number of samples N :

<u>Random</u>	<u>Stratified random</u>	<u>Regular</u>
$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{N^{3/2}}$	$\frac{\sigma^2}{2.14 \times N^{3/2}}$

TABLE I - Estimation variances for the average over a square when the variogram is linear.

One can appreciate the benefit of a sampling pattern exploiting spatial correlations over the crude random sampling : the variance is brought down by a factor of $N\sqrt{N}$ instead of N .

Note that if V is not a square but an arbitrary domain of area S (that we divide into square blocks), the formulae for stratified random and regular remain valid with $0.5214 \times \omega \times \sqrt{S}$ substituted for σ^2 .

The same type of comparison can be made in the case of a bounded variogram. But the results now depend on an additional parameter : the range a of the variogram. Suppose that a 2-D domain V has been divided into square blocks of side ℓ , and that the variogram is an isotropic spherical model.

From readings of Graph 1, Chapter III, and Graph 1, Chapter IV, curve $\sigma_{E_2}^2$, we can build Table 2. The variance $\sigma^2(0|V)$ has not been included as it depends on the shape of the domain V . However, for reference purposes, one may keep in mind that if the dimensions of V are large with respect to a ($L/a > 10$), then $\sigma^2(0|V) = 1$ in the units of Table 2.

TABLE 2

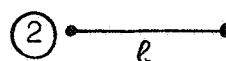
ℓ/a	$\sigma^2(0 v)$	$\sigma_E^2(C,v)$	Ratio
0.15	0.118	0.056	2.107
0.2	0.155	0.074	2.094
0.4	0.31	0.15	2.067
0.6	0.45	0.235	1.915
0.8	0.56	0.32	1.75
1.	0.66	0.41	1.61
1.5	0.81	0.65	1.25
2	0.88	0.80	1.1
3	0.94	0.92	1.02
5	0.977	0.96	1.017
8	0.9905	0.99	1

Variances for the Spherical Model

Again, the regular grid network performs better than the random stratified one. However, this advantage fades away as the size of the grid box becomes large with respect to the range of the variogram. The reason is that for a large ℓ/a ratio, the influence of a sample is purely local anyway and the center of the square loses its strategic superiority.

5- EXERCISE.

1 - In one dimension, compare the two following sampling patterns for the family of variograms : $\gamma(h) = |h|^\lambda$ $0 < \lambda < 2$



Answer :


$$\chi(h) = \frac{h^\lambda}{\lambda+1} ; \quad F(h) = \frac{2 h^\lambda}{(\lambda+1)(\lambda+2)} ;$$

$$\sigma_1^2 = \frac{2 l^\lambda}{\lambda+1} \left(\frac{1}{2^\lambda} - \frac{1}{\lambda+2} \right) ; \quad \sigma_2^2 = \left(\frac{2}{\lambda+2} - \frac{1}{2} \right) l^\lambda$$

λ	σ_1^2	σ_2^2
0	1	0.5
0.5	$0.409 \sqrt{l}$	$0.3 \sqrt{l}$
1	$0.167 l$	$0.167 l$
1.5	$0.054 l^{1.5}$	$0.071 l^{1.5}$
2	0	0

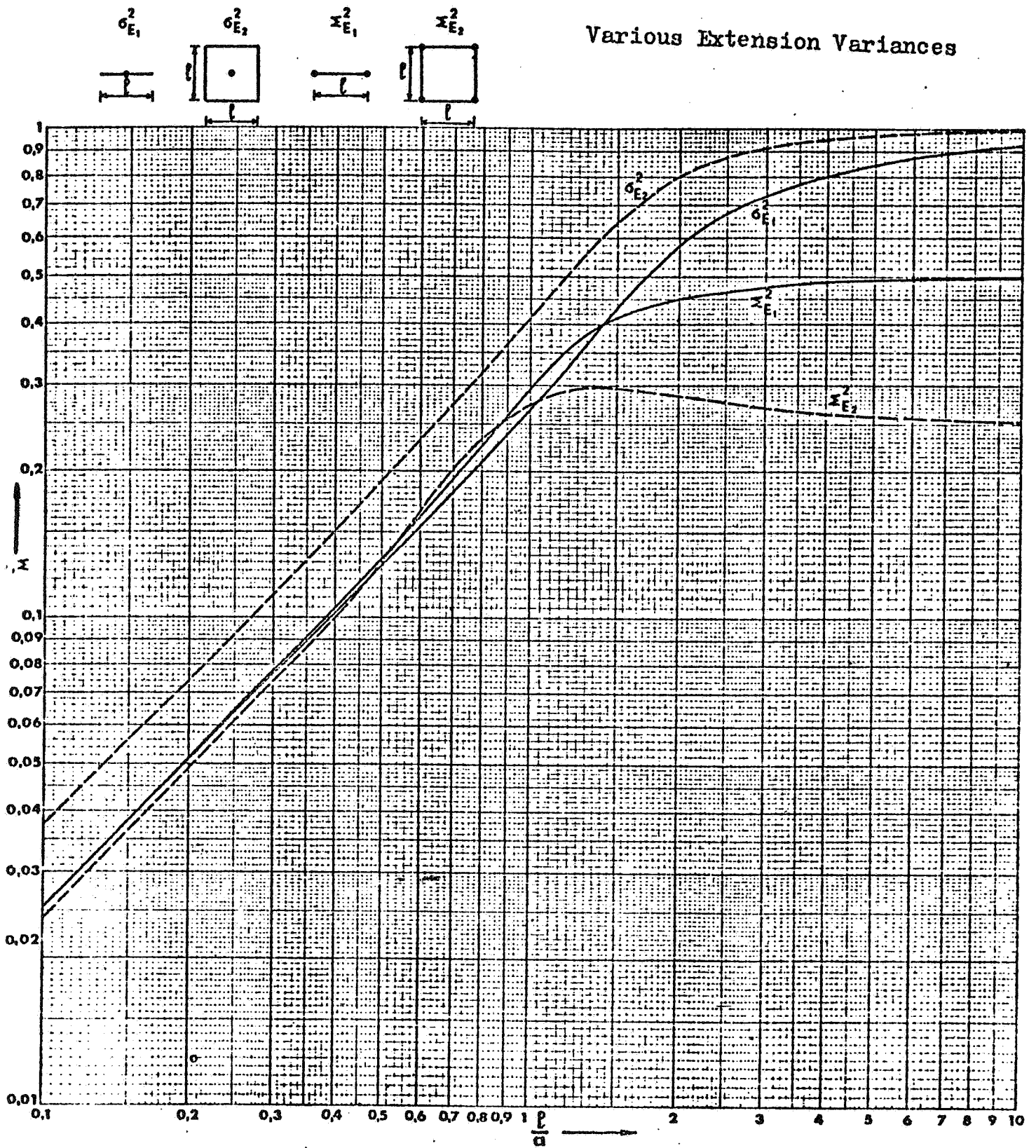
For $\lambda < 1$  better than 

For $\lambda = 1$  equivalent to 

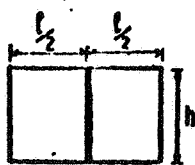
For $\lambda > 1$  better than 

For $\lambda = 0$, limiting case of a pure nugget effect, only the number of samples matters.

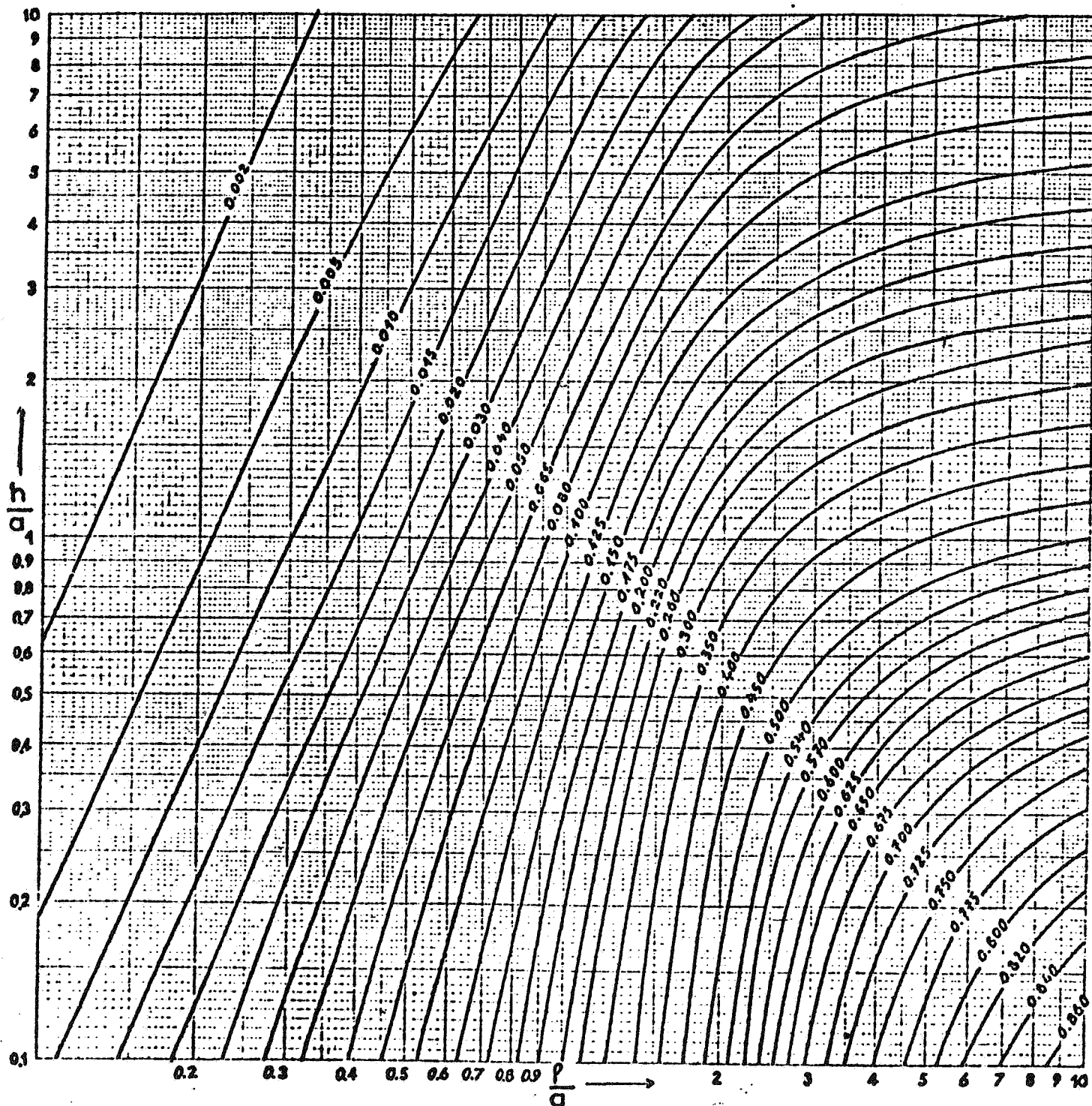
As λ increases, the variogram becomes more regular, and a single centrally located sample becomes better than two ill-placed ones. For $\lambda = 1$ the two patterns are equivalent. For $\lambda = 2$ the variances are 0, but $\lambda = 2$ is not an admissible value for a variogram model.



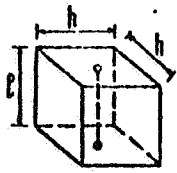
GRAPH 1 SPHERICAL VARIOGRAM



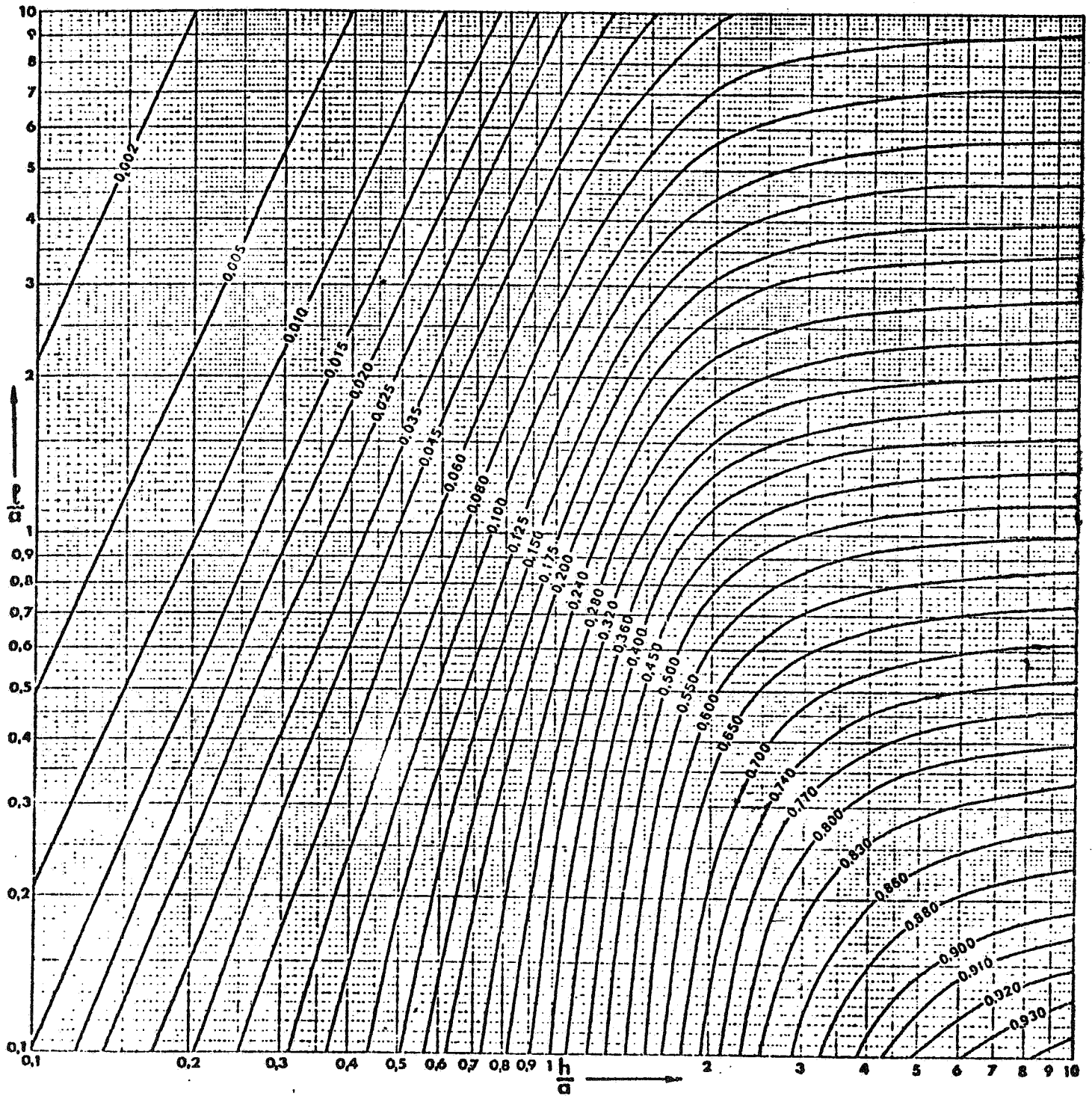
Extension variance of median drift h to rectangle l, h



GRAPH 2 SPHERICAL VARIOGRAM



Extension variance of central bore-hole 1 to square block of influence 1,h,h



GRAPH 3 SPHERICAL VARIOGRAM

- CHAPTER V -

K R I G I N G

1 - THE PURPOSE OF KRIGING.

Measurements on core data or log analysis results provide the reservoir engineer with detailed information on a single well basis. The problem is then to extend these studies to a field basis.

Clearly, the answer can only be in the form of "estimates". No data processing method, however sophisticated, will ever tell us exactly what would be found at a well that has not been drilled. However, what a good method can do is to handle the precious available data in an efficient way so as to get the most out of them.

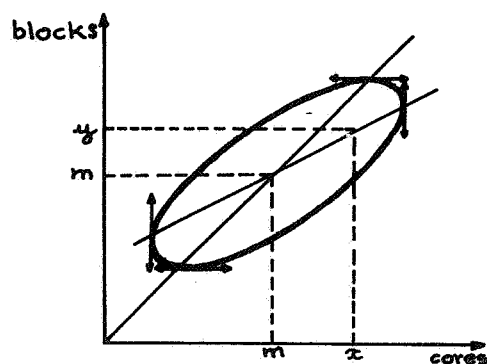
The accuracy of estimates depends on several factors :

- (i) the number of wells and the quality of the data for each well : this quality may vary from well to well and therefore all wells should not be granted the same importance.
- (ii) the positions of the wells within the field : from the point of view of a general knowledge of the field, wells that are evenly spaced throughout the field achieve a better coverage and thus give us more information than if they were clustered in a particular region.
- (iii) the distances between the wells and the area of interest : if we are interested in a particular block it is natural to rely more on wells close to the block than on distant ones. Likewise, if we want reservoir parameter values on a point basis (thickness for example), we expect the accuracy to be good in the vicinity of the wells and to deteriorate as the distances from the wells increase.
- (iv) the spatial continuity of the interpolated variables : obviously enough, a quantity with smooth variations can be interpolated better than a quantity with erratic

fluctuations. For example, with the same pattern of data points, it is usually easier to assess the depth of the reservoir at a point than its local porosity or permeability. In fact, these are two very different problems and reservoir engineers are well aware of it since they generally use different methods to solve them : "exact surface interpolation" for depths and "trend surface analysis" for porosities and permeabilities. This is a first step. Optimally, the interpolation algorithm should be tailored to each particular variable.

"Kriging" is a method of estimation which takes into account all these factors. The term was derived by G. Matheron from the name of D.G. Krige, a mining specialist who first introduced the use of moving averages to avoid systematic overestimation of ore reserves in mineral deposits. Kriging essentially consists in taking the best linear unbiased estimator (abbreviation : B.L.U.E.) of the quantity to be estimated, i.e. to give a weight to the different available samples so as to obtain an estimator with minimum mean squared error. Before giving a detailed account of the method it may be interesting to present an hypothetical example designed to illustrate the rationale of the weighting procedure.

Suppose we want to estimate the average porosity of blocks in a homogeneous layer of a reservoir. Blocks are sampled by cores and it is assumed that the sampling is unbiased, i.e. the expected value of the porosity of a core taken inside a block is equal to the



average porosity of that block. We assume now for simplicity of the argument that the porosities have a Gaussian distribution ; then the bivariate distribution of block porosity versus core porosity is also Gaussian. The regression of Φ_{core} on Φ_{block} , i.e. the expected value of Φ_{core} given Φ_{block} is the unit slope line. Then, the other regres-

sion line, the interesting one, giving Φ_{block} as a function of Φ_{core} has a slope β necessarily smaller than one. The regression formula writes : $y = m + \beta(x-m)$.

When the porosity x of a core is greater than the general average m of the layer the expected porosity y of the block is then smaller than x , and vice versa. In other words, there is a correction to weight down values of x larger than the overall mean m , and conversely.

In fact, this regression approach suggests a more general formulation. Rewriting y as :

$$y = \beta x + (1-\beta)m$$

we see that the estimate y is a linear combination of the sample value x and the general average m , the sum of weights being 1. But in practice, m itself is known only through an estimate $m^* = \frac{1}{N} \sum_i X_i$, i.e. the average of cores in the layer. Altogether the estimate y^* of the block porosity is of the form

$$y^* = \sum_i a_i X_i$$

with $\sum_i a_i = 1$, and $a_i = \frac{1-\beta}{N}$ for all cores outside the block, whether or not they are close to the estimated block. The kriging procedure will improve that weighting by allocating to each core the weight a_i that it really deserves.

2 - DERIVATION OF KRIGING EQUATIONS

The general situation is the following_: we have N data values $z(x_1), z(x_2), \dots, z(x_N)$ measured at points x_i , $i = 1, 2, \dots, N$. We want to estimate a quantity y_0 which is any linear functional of the studied variable $z(x)$. For example ::

$$(i) \quad y_0 = z(x_0) \quad : \text{value taken by } z \text{ at } x = x_0$$

- (ii) $y_o = \frac{1}{v} \int_{v_o} z(x) dx$: average of $z(x)$ over the block v_o of volume v and centered at $x = x_o$
- (iii) $y_o = \frac{1}{V} \int_V z(x) dx$: average value of $z(x)$ over the whole field V
- (iv) $y_o = \int z(x_o+u) p(u) du$: weighted moving average of $z(x)$
- (v) $y_o = \text{grad } z(x)|_{x=x_o}$: components of the gradient at $x = x_o$ (hence dip and strike of a geological layer)

For the sake of simplicity we shall give explicit solutions in the first three cases only. furthermore, we shall pool them in a single problem by writing y_o as an average value over an unspecified domain V , with the understanding that if V is reduced to a point x_o the average value of z over the point is $z(x_o)$ itself :

$$y_o = \frac{1}{V} \int_V z(x) dx$$

To estimate y_o we consider a weighted average of the data :

$$y_o^* = \sum_{i=1}^N \lambda_i z(x_i)$$

(By convention the star * will denote kriging estimates). The problem is to choose the weights λ_i in the best possible way. This is where we appeal to our statistical setup. Considering now the estimator

$$Y_o^* = \sum_{i=1}^N \lambda_i Z(x_i)$$

we determine the weights λ_i so that Y_o^*

1) is unbiased : $E[Y_o^* - Y_o] = 0$

2) has minimum mean squared error : $E[Y_o^* - Y_o]^2$ minimum

When condition 1 is satisfied the mean squared error is also the variance and we shall refer to the variance or the standard deviation of the kriging error, rather than the mean squared error.

To find the λ_i 's that satisfy 1) and 2) we shall distinguish two cases : the so-called stationary case where the means of $z(x)$ are constant and the non-stationary case where there is a drift.

A - The Stationary Case.

Here $E[Z(x_i)] = m$ and $E[Y_o] = m$

Condition 1 entails :

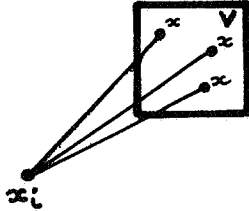
$$E\left[\sum_{i=1}^N \lambda_i Z(x_i) - Y_o\right] = \sum_{i=1}^N \lambda_i m - m = m \left(\sum_{i=1}^N \lambda_i - 1\right) = 0$$

Therefore for the estimator to be unbiased the weights must add up to 1, which is a rather intuitive condition.

The variance of the error can be expressed in terms of the variogram :

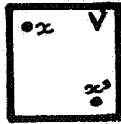
$$E[Y_o^* - Y_o]^2 = \text{Var}(Y_o^* - Y_o) = 2 \sum_{i=1}^N \lambda_i \bar{\gamma}(x_i, V) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(x_i - x_j) - \bar{\gamma}(V, V)$$

where



$\bar{\gamma}(x_i, V)$ is the average of the variogram between x_i and the volume V , i.e.

$$\bar{\gamma}(x_i, V) = \frac{1}{V} \int_V \gamma(x_i - x) dx$$



$\bar{\gamma}(V, V)$ is the average of the variogram between any two points x and x' sweeping independently throughout the volume V (F-function) :

$$\bar{\gamma}(V, V) = \frac{1}{V^2} \int_V \int_V \gamma(x - x') dx dx' = F(V)$$

In order to minimize the above variance under the constraint that the sum of the weights must be 1, we introduce a Lagrange multiplier μ and set to 0 the partial derivatives $\frac{\partial \phi}{\partial \lambda_i}$, $i = 1, 2, \dots, N$ and $\frac{\partial \phi}{\partial \mu}$ of the quantity :

$$\phi = \text{Var}(y_o^* - y_o) - 2 \mu \left(\sum_{i=1}^N \lambda_i - 1 \right)$$

We get the following linear system :

Kriging System	$\sum_{j=1}^N \lambda_j \gamma(x_i - x_j) + \mu = \bar{\gamma}(x_i, V) \quad i = 1, 2, \dots, N$
	$\sum_{j=1}^N \lambda_j = 1$

The minimum of the variance or "kriging variance" is given by :

Kriging Variance	$\sigma_K^2 = \text{Var}(y_o^* - y_o) = \sum_{i=1}^N \lambda_i \bar{\gamma}(x_i, V) - \bar{\gamma}(V, V) + \mu$
---------------------	---

When the variogram is a pure nugget effect γ_o , we get : $\lambda_i = \frac{1}{N}$.
In the case where V is a point $\bar{\gamma}(x_i, V) = \gamma(x_i - x_o)$

and

$$\bar{\gamma}(V, V) = \gamma(0) = 0$$

To solve the system numerically, it is convenient to write it in a matrix form. We get simply :

$$\Gamma X = B$$

i.e. :

$$\underbrace{\begin{bmatrix} 0 & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1N} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \dots & \gamma_{2N} & 1 \\ \gamma_{31} & \gamma_{32} & 0 & \dots & \gamma_{3N} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \dots & \dots & 0 & 1 \\ 1 & 1 & \dots & \dots & 1 & 0 \end{bmatrix}}_{\Gamma} \times \underbrace{\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix}}_X = \underbrace{\begin{bmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \bar{\gamma}(x_3, V) \\ \vdots \\ \bar{\gamma}(x_N, V) \\ 1 \end{bmatrix}}_B$$

If γ is a proper variogram model, Γ is always non-singular and the solution is simply :

$$X = \Gamma^{-1} B$$

and the variance

$$\sigma_K^2 = X^T B - \bar{\gamma}(V, V) \quad (X^T = X \text{ transposed})$$

Beware : Γ is not positive definite (zeroes on the diagonal)

B - Non-Stationary Kriging ("Universal Kriging")

In cases where the variable shows a systematic increase or decrease along certain directions we say that there is a drift. Then, the assumption that $E[Z(x)] = \text{constant}$ is obviously violated and a more sophisticated model must be used.

A natural one is to consider that the drift $m(x) = E[Z(x)]$, a slowly varying function, can be well approximated by a polynomial of the form :

$$m(x) = \sum_{\ell=1}^k a_{\ell} f^{\ell}(x)$$

The f^ℓ are monomials. For example, in 2-D the explicit expression of a linear drift in terms of the coordinates x and y would write :

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

and a quadratic drift :

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

The coefficients a_0, a_1, \dots, a_k are of course unknown. A surprisingly nice fact is that we do not need to estimate them, but just to introduce into the kriging system some more side conditions. To see why that is so, let us just examine the case of point estimation, i.e. when $Y_0 = Z(x_0)$. The unbiasedness of Y_0^* implies this time :

$$E[Y_0^* - Y_0] = \sum_{i=1}^N \lambda_i m(x_i) - m(x_0) = 0$$

But as

$$m(x_i) = \sum_{\ell=1}^k a_\ell f^\ell(x_i)$$

we have, after rearrangement :

$$\sum_{\ell=1}^k a_\ell \left[\sum_{i=1}^N \lambda_i f^\ell(x_i) - f^\ell(x_0) \right] = 0$$

This has to be an identity in a_ℓ so that we get the unbiasedness conditions :

$$\sum_{i=1}^N \lambda_i f^\ell(x_i) = f^\ell(x_0) \quad \ell = 1, 2, \dots, k$$

The rest of the derivation is the same as for stationary kriging. Again the weights λ_i are solutions of a linear system but this time with k Lagrange multipliers $\mu_1, \mu_2, \dots, \mu_k$ instead of just one. This linear system has a unique solution, provided that the k vectors $f^\ell(x_i)$ ($i = 1, \dots, N$) are linearly independent on the set of observations; i.e. :

$$\sum_{\ell} c_\ell f^\ell(x_i) = 0 \quad \text{for all } i \Rightarrow c_\ell = 0 \quad \text{for all } \ell$$

NON-STATIONARY

KRIGING SYSTEM

KRIGING VARIANCE

$$\sum_{j=1}^N \lambda_j \gamma(x_i - x_j) + \sum_{\ell=1}^k \mu_{\ell} f^{\ell}(x_i) = \gamma(x_i - x_0) \quad (i = 1, 2, \dots, N)$$

$$\sum_{j=1}^N \lambda_j f^{\ell}(x_j) = f^{\ell}(x_0) \quad (\ell = 1, 2, \dots, k)$$

$$\sigma_K^2 = \text{Var}(Z^*(x_0) - Z(x_0)) = \sum_{j=1}^N \lambda_j \gamma(x_j - x_0) + \sum_{\ell=1}^k \mu_{\ell} f^{\ell}(x_0)$$

3 - PROPERTIES AND USAGE OF KRIGING ESTIMATORS. APPLICATIONS TO CONTOUR MAPPING.

Concerning what we mentioned in the introduction on the relationships between an estimator and the data, we note that the kriging system takes the following elements into account :

- (i) relative position of the estimated domain V and the sample points x_i : through the terms $\bar{\gamma}(x_i, V)$ or $\gamma(x_i - x_0)$.
- (ii) distances between sample points : through the terms $\gamma(x_i - x_j)$.
- (iii) structure of the variable : through the semi-variogram γ .

The influence of the structure on the kriging weights is illustrated by Fig. 1. The value at a grid node has been kriged with the same data points but with two different variograms.

We can see what was meant by an algorithm tailored to each particular variable. When the variable is continuous, kriging relies a lot on close data points **precisely because continuity** means that close points assume close values ; this is no longer true if the variable is irregular and consequently kriging **damps** down the influence of nearest neighbors.

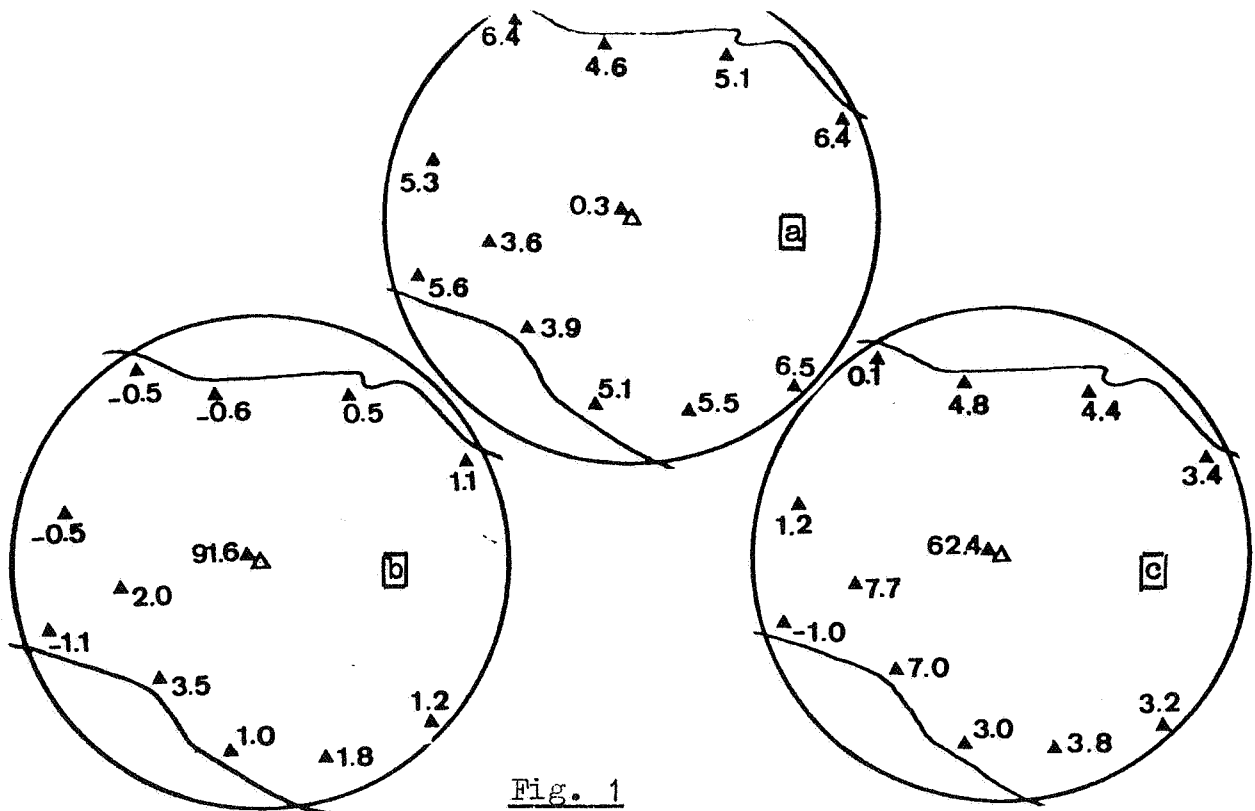


Fig. 1

- (a) distances of data points to grid node (in Km)
- (b) kriging weights in % computed with a continuous variogram $\gamma(h) = 15 |h|$: a large weight is ascribed to the closest data (91.6%)
- (c) kriging weights in % computed with a variogram with nugget effect $\gamma(h) = 20.4 + 11.2 |h|$: the nearest neighbor gets only 62.4%

A very important property of kriging is to be an exact interpolator : when estimating point values, kriging restores at data points the measured values. This can be checked on the kriging system : when x_0 coincides with x_i , then the solution is $\lambda_i = 1$ and $\lambda_j = 0$ for $j \neq i$. But it is even simpler to go back to the first principles and see that the minimum of $E[Z^*(x_0) - Z(x_0)]^2$ is obviously achieved when $Z^*(x_0) = Z(x_0)$ and this minimum is 0. This makes kriging an appropriate method for contour mapping. Moreover, as a by-product, kriging provides the variance of the error σ_K^2 from which an error map can be drawn. If the error is assumed to be Gaussian, a 95% confidence interval for the true value $z(x_0)$ is $[z^*(x_0) - 2 \sigma_K, z^*(x_0) + 2 \sigma_K]$ so that by mapping $2 \sigma_K$ we get a picture of the spatial distribution of the 95% error level. An example is shown in Fig. 2 and 3. Fig. 2

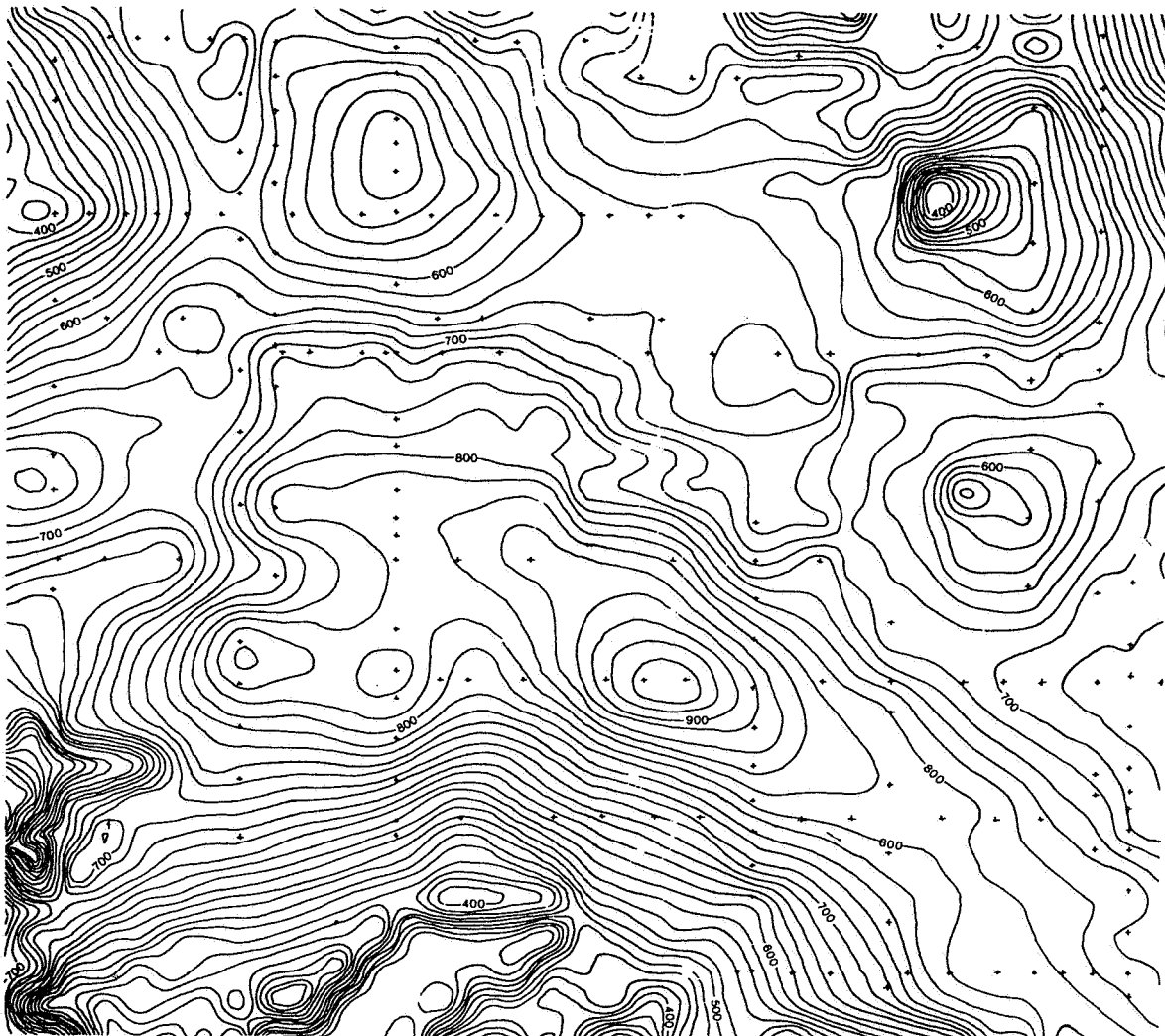


Fig. 2

Time Contour Map (C. HUIJBREGTS, 1971)

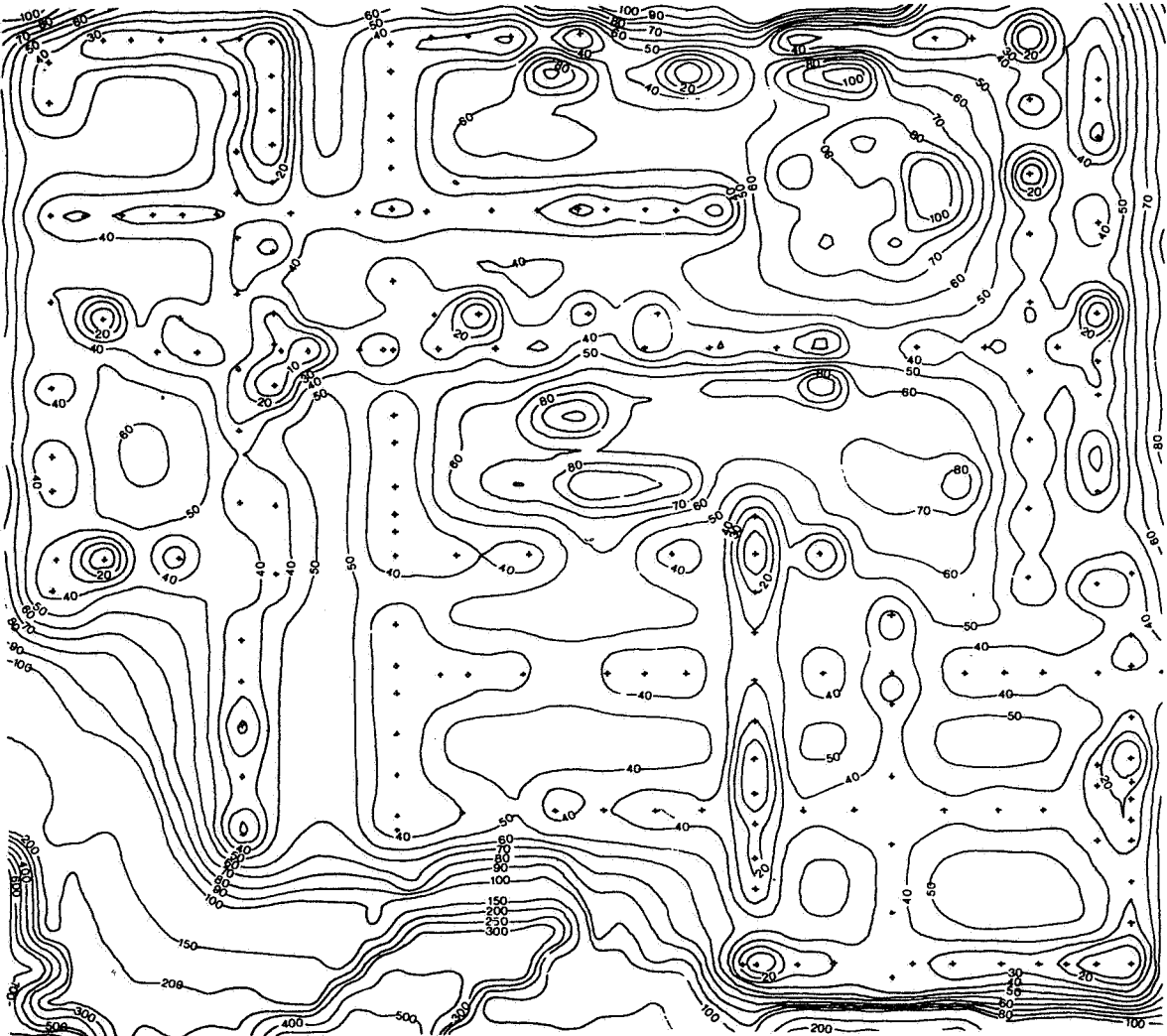


Fig. 3

Error Contour Map ($2\sigma_K$) (C. HUIJBREGTS, 1971)

is a time contour map (in msec) computed by kriging. The crosses indicate the position of the measurements. Certain interesting structures can be observed, for example there is one on the right upper part of the map. Looking at the error map shows that in that area $2 \sigma_K$ is around 100 msec, which casts doubts on the reality of that structure. Likewise in the left lower corner, where there is no data, we observe a high standard deviation urging us not to give credit to the wriggling contours in this region. **On the other hand**, when the standard deviation is low the map can be trusted.

4 - HOW TO TAKE INTO ACCOUNT THE UNCERTAINTY OF THE DATA.

Porosities and fluid saturations are obtained after a complex process involving several measurements techniques, empirical formulae and coefficients deduced from laboratory experiments. In general we shall assume that the uncertainty about a value $z(x_i)$ is summarized by a variance σ_i^2 . Sometimes σ_i^2 will have been computed by statistical or geostatistical methods, or just roughly figured out for example by dividing the likely range of variation by 4 and squaring the result. We shall assume that the errors $\varepsilon(x_i)$ are :

a) unsystematic : $E[\varepsilon(x_i)] = 0$ for all i

b) uncorrelated : $\text{cov}[\varepsilon(x_i), \varepsilon(x_j)] = 0$ for all $i \neq j$

c) uncorrelated with the studied variable :

$$\text{cov}[\varepsilon(x_i), Z(x)] = 0 \quad \text{for all } i \text{ and all } x.$$

Then, the uncertainty on the data just introduces a slight modification in the kriging system. Instead of $z(x_i)$ our data are in fact $z(x_i) + \varepsilon(x_i)$, i.e. the real value plus an error, and our estimator is

$$Y_o^* = \sum_{i=1}^N \lambda_i [Z(x_i) + \varepsilon(x_i)]$$

Since $E[\varepsilon(x_i)] = 0$ nothing is changed in the unbiasedness conditions. But the variance of the error becomes :

$$\begin{aligned} \text{Var}(Y_o^* - Y_o) &= \text{Var}\left[\left(\sum_{i=1}^N \lambda_i Z(x_i) - Y_o\right) + \sum_{i=1}^N \lambda_i \varepsilon(x_i)\right] \\ &= \text{Var}\left[\sum_{i=1}^N \lambda_i Z(x_i) - Y_o\right] + \text{Var}\left[\sum_{i=1}^N \lambda_i \varepsilon(x_i)\right] \\ &= \text{Var}\left[\sum_{i=1}^N \lambda_i Z(x_i) - Y_o\right] + \sum_{i=1}^N \lambda_i^2 \sigma_i^2 \end{aligned}$$

When minimizing this expression we get the kriging system except for the N diagonal terms $\gamma_{ii} = 0$ which are replaced by $-\sigma_i^2$. The stationary kriging system for example writes :

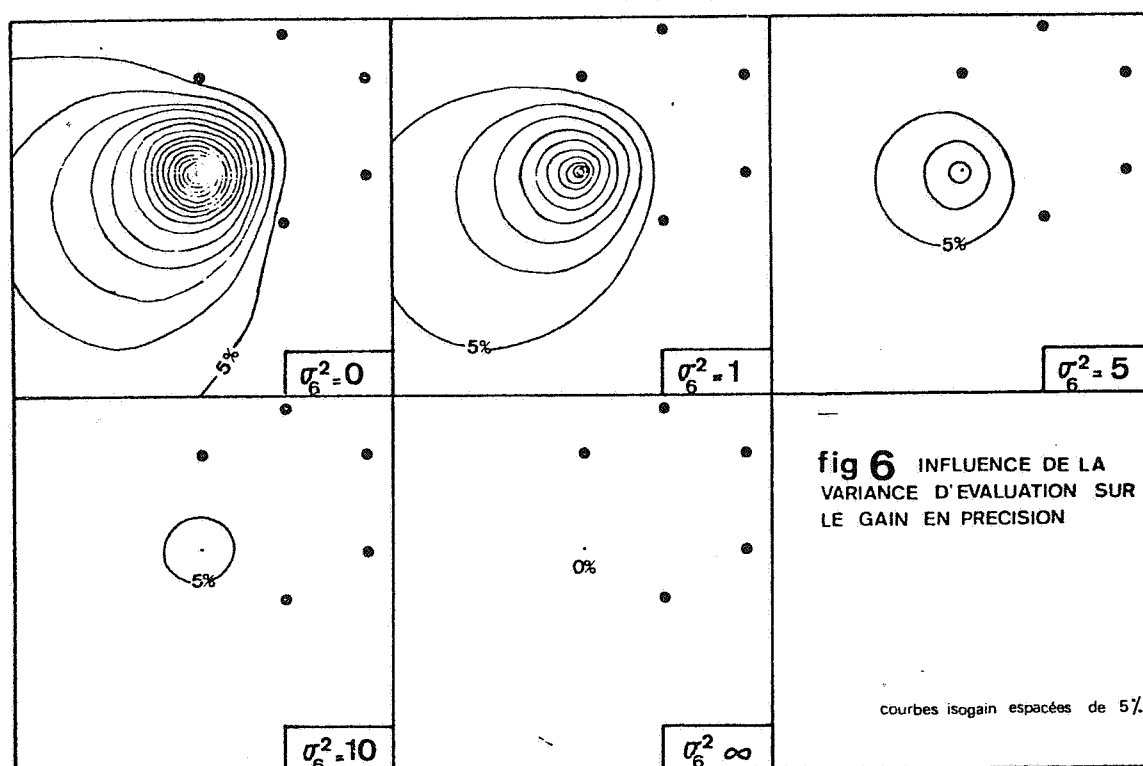
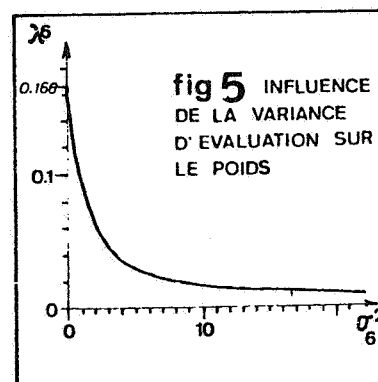
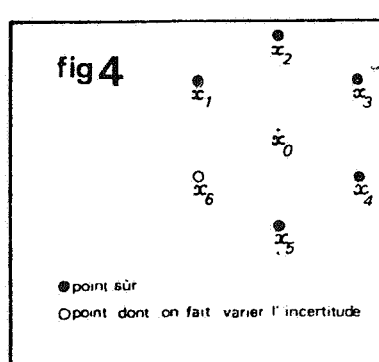
STATIONARY KRIGING WITH ERRORS AT DATA	$\sum_{j=1}^N \lambda_j \gamma(x_i - x_j) - \lambda_i \sigma_i^2 + \mu = \bar{\gamma}(x_i, V) \quad i = 1, 2, \dots, N$
	$\sum_{j=1}^N \lambda_j = 1$
	$\sigma_K^2 = \sum_{j=1}^N \lambda_j \bar{\gamma}(x_i, V) + \mu - \bar{\gamma}(V, V)$

It is interesting to get a feeling for the practical effects of the introduction of an uncertainty variance σ_i^2 . For this, let us consider the following elementary example (from J.P. DELHOMME, 1974) : the variable has no drift, a linear variogram $\gamma(h) = |h|$, and the data are located at the vertices of a regular hexagon with sides of unit length (Fig. 4). Five of these data are supposed to be without error $\sigma_i^2 = 0$ for $i = 1, 2, \dots, 5$, and we want to study the influence of σ_6^2 .

One way is to look at the weight λ_6 ascribed to $z(x_6)$ when estimating the value at the center of the hexagon (Fig. 5). At the beginning $\sigma_6^2 = 0$ so $z(x_6)$ gets a weight of $1/6 = 0.166$. As σ_6^2 increases, λ_6 decreases and in the limit vanishes when σ_6^2 is infinite : a totally imprecise data is like no data at all.

Another way to study the information brought about by the sixth data point is to look at the relative variance reduction due to the presence of this sixth value. This time we want to estimate $z(x)$ at any point x (not necessarily the center of the hexagon). Using the five data only, we would have a kriging variance $\sigma_{K_5}^2$, while using also the sixth value we have $\sigma_{K_6}^2$ which is smaller. In Fig. 6 we can see the contour lines of the relative variance reduction (or "gain in precision") :

$$g = \frac{\sigma_{K_5}^2 - \sigma_{K_6}^2}{\sigma_{K_5}^2}$$



When $\sigma_6^2 = 0$, the maximum variance reduction is 100% at x_6 (since then $\sigma_{K_6}^2 = 0$) and decreases as the kriged point moves away from x_6 . For larger and larger values of σ_6^2 the zone of influence of the sixth data point shrinks and in the limit disappears when σ_6^2 is infinite. Again, a totally uncertain data would be of no help at all.

5 - APPLICATIONS OF KRIGING TO THE ESTIMATION OF HYDROCARBON IN PLACE.

Kriging techniques can be used for reserves evaluation either directly to calculate the volume of hydrocarbon in place or indirectly to provide dynamic reservoir simulation models with grids of initial formation parameters values.

A - Volumetric Calculations.

The volume of hydrocarbon in place, denoted by Q , is given by the following triple integral :

$$Q = \int_x dx \int_y dy \int_{\text{Top}(x,y)}^{\text{Min}(WL, \text{Bot}(x,y))} [1 - S_w(x,y,z)] \Phi(x,y,z) dz$$

WL = water level, Bot = Bottom, Φ = porosity, S_w = water saturation. x and y are the two horizontal coordinates and z is the depth. If required, the formation volume factor (F.V.F.) can be introduced in the formula or the correction be made later.

Ideally this integral should be calculated by numerical integration in 3-D. This would require estimation of the hydrocarbon porosity

$$\Phi_h(x,y,z) = [1 - S_w(x,y,z)] \Phi(x,y,z)$$

at the nodes of a 3-D grid limited by the reservoir boundaries. On account of the vertical heterogeneity of a reservoir, this would be a heavy task. A shortcut consists in reducing the

problem to a 2-D one by considering cumulated hydrocarbon porosities :

$$\begin{aligned} H\Phi S_h(x,y) &= \int_{\text{Top}(x,y)}^{\text{Min}[WL, \text{Bot}(x,y)]} [1 - S_w(x,y,z)] \Phi(x,y,z) dz \\ &= [\text{Min}[WL, \text{Bot}(x,y)] - \text{Top}(x,y)] \times \bar{\Phi}_h(x,y) \end{aligned}$$

where H = thickness and h is a subscript for hydrocarbon porosity or saturation.

$$Q = \int_x \int_y H \Phi S_h(x,y) dx dy$$

Now it suffices to grid in 2-D the variable $H \Phi S_h(x,y)$ and sum over the grid points. This can be done :

- (i) either by working on the variable $H \Phi S_h$ directly
- (ii) or by gridding H , Φ and S_h independently and multiplying the grids.

The first approach requires that measurements of H , Φ and S_h are available for all wells. This is usually not the case and the second method is preferred.

An intermediate method is to divide the reservoir into homogeneous layers and work layer by layer, as for independent reservoirs. These layers can be horizontal ones when the profile of water saturation plays a major rôle, or determined by lithologic studies. Variograms can also be a help for the zonation of the reservoir by indicating those layers which maximize spatial correlations. At this stage no general recommendation can be made. To a certain extent, one can say that each reservoir is a particular case and kriging techniques, as any other tool, should be used with insight.

The complexity of reserves evaluation and the fact that several variables are involved makes it difficult to perform rigorous error calculations. However, kriging provides some indications. Symbolically we may represent the reserves in place

in a certain volume (reservoir or layer) as :

$$Q = \underbrace{V}_{\text{volume}} \times \underbrace{\overline{\Phi_h}}_{\text{average hydrocarbon porosity over } V}$$

The errors can be classified in two categories :

- (i) errors of estimation of Φ_h : they depend on the variability of the variables, on the number and location of the wells and on the uncertainties at the wells.
- (ii) errors on the geometry of the reservoir : errors about the reservoir top, the water level depth, and possibly the presence of faults and wedges.

If the geometrical error may be neglected then :

$$\sigma_Q^2 = V^2 \sigma_{\Phi_h}^2$$

with $\sigma_{\Phi_h}^2$ given by kriging. In the case of several layers a crude approximation is to simply add the estimation variances of each layer as if the errors were independent. When the boundaries of the reservoir are very imprecise we must resort to other methods (cf. infra : "conditional simulations").

B - Initialization of dynamic reservoir simulation models.

Kriging of course can provide the input grid for reservoir simulation models (Φ, K, H). Studies are going on to combine kriging with history-matching techniques based on optimal control theory. The idea is to realize that, although they are highly variable, permeabilities are usually not totally erratic, and that their structure can be taken into account by means of a geostatistical model. This approach will reduce the number of parameters to be determined, and is expected to yield more realistic solutions to permeability distribution in the reservoir.

6 - HOW DO KRIGING ESTIMATES COMPARE WITH REALITY ? A CASE STUDY.

In the early years when calculating a variogram and laying down kriging equations was still considered pioneering work, some enthusiastic geostatisticians claimed that kriging estimates were the "best" because they minimized estimation variances. This argument never quite convinced practitioners, and rightly so. Optimality is a property established within a specified model and its relevance to a real situation depends on well the model describes this reality.

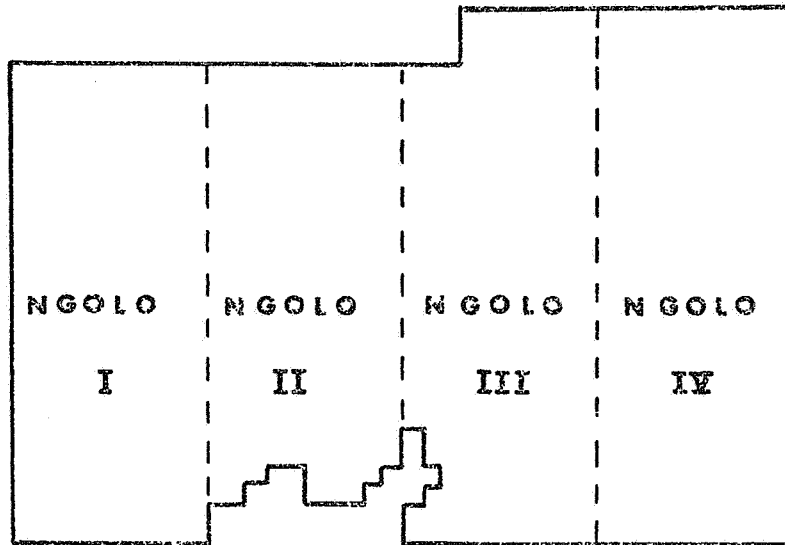
Another means of assessing the quality of kriging estimates is to check them against the true values. Unfortunately - or perhaps fortunately for the geostatistician, who knows ? - the true values are seldom known. One may instead work on Monte Carlo simulations but these do not have the flavor of real case studies.

In that respect a recent report by Ph. Narboni offers a very interesting picture of the performance of kriging estimators in forestry estimation. He conducted an exhaustive survey of 20,000 ha in Gabon (1 hectare = 10,000 m² = 2.471 acres), concentrating on the predominant type of wood called "okoumé". The basic document was a map at a scale of 1/10,000 showing the position of every tree with a diameter greater than 60 cm. A 5 mm square mesh grid was superimposed on the map, defining sampling units of 50 m × 50 m (patches) in which trees were counted. The studied variable is the count of trees per hectare.

The sampling scheme may be described as a transect survey, but with contiguous transects of 50 m width.

A total of about 80,000 sampling units were analyzed, a task that took two man-months.

Variograms were calculated : (a) using the complete survey, (b) using continuous transect surveys sampled at a rate of 1 %, 5 % and 10 % of the total. Because of heterogeneity of the forest the area was divided into 4 zones of about 5,000 ha each (Fig. 7). An example of the variograms obtained is given in Fig. 8.



Parting the area into 4 zones.

Fig. 7

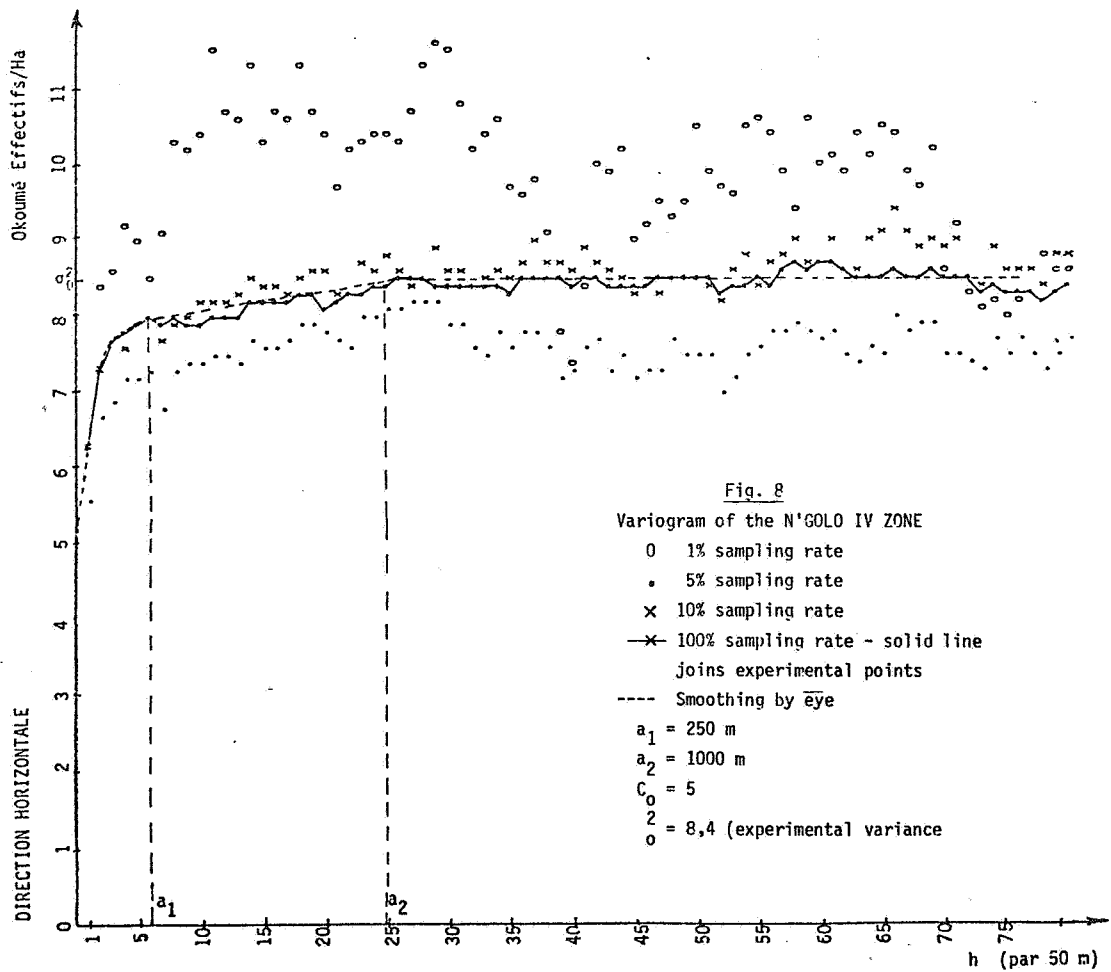


Fig. 8 (From Ph. NARBONI , 1979)

One interesting outcome of the study is the evidence of the fluctuations of the experimental variograms (at 10 % sampling rate it is possible to compute 10 replicates). We will not discuss this aspect here but simply note that ranges are stable while sills vary more (factor of 1.2 at 10 % rate, of 2 at 1 % rate). The fitted variogram is a sum of two spherical models plus a nugget effect.

Estimation of 20 ha panels (500 m \times 400 m) was carried out at a 10 % sampling rate. Panels are centered on transects (Fig. 10). The classical method used in forestry is simply to take the mean value of patches inside the panel as an estimate of the panel itself. Kriging for its part also uses information from outside the panel. In case of a regular sampling pattern such as this the same kriging configuration can be used for all panels, and weights calculated once for all. Fig. 9 shows the layout employed here, with data grouped into 5 rings.

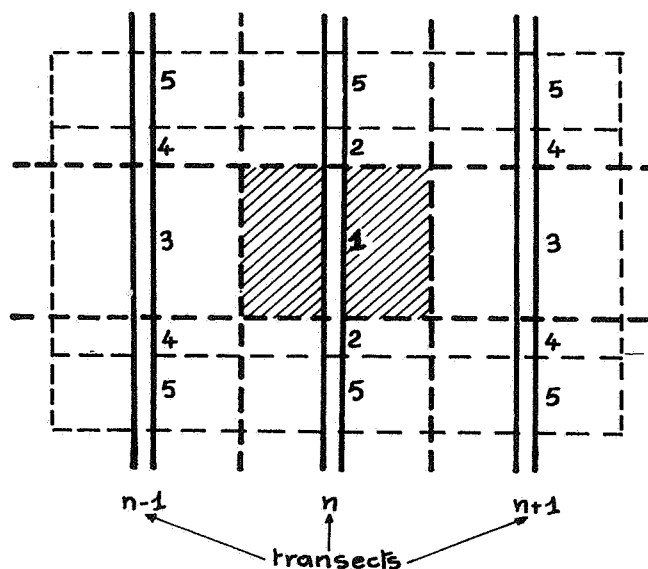


Fig. 9 : Kriging layout

Table 1 compares mean estimates over the 4 zones ; multiplication by the areas give the total number of trees.

Zones NGOLO	\bar{Z}_v	\bar{Z}_1^*	Pondérateurs					\bar{Z}^*	σ_k^2	$2\sigma_k/\bar{Z}^*$
			λ_1	λ_2	λ_3	λ_4	λ_5			
I	1,18	1,25	0,38	0,10	0,18	0,08	0,26	1,25	0,20	72 %
II	1,26	1,29	0,38	0,08	0,18	0,08	0,28	1,28	0,23	75 %
III	1,53	1,58	0,44	0,10	0,16	0,06	0,24	1,53	0,34	76 %
IV	1,69	1,61	0,41	0,10	0,17	0,07	0,25	1,62	0,36	74 %

TABLE 1

\bar{Z}_v : true value \bar{Z}_1^* : mean of inside samples

\bar{Z}^* : mean of kriging estimates σ_k^2 : kriging variance

λ_i : weights ascribed to the rings of Fig. 9

It may be noticed that outside samples are weighted more than inside samples (λ_1 less than 0.5). This is due to the presence of a large nugget effect.

Table 1 basically indicates that the two estimates \bar{Z}_1^* and \bar{Z}^* are comparable as far as global estimation is concerned. This does not remain true however when turning to local estimation. Fig. 10 is the cross-plot of true values versus kriging estimates : the cloud is well-centered about the unit slope line, with a few points outside the 95 % confidence interval (there should be some !). This plot illustrates a property of the kriging estimator that can be proved mathematically in the case of a Gaussian random function with a known mean, namely conditional unbiasedness :

$$E[Z_v | Z^*] = Z^*$$

If we go and exploit all panels estimated to be Z^* their mean will indeed be Z^* . By contrast, Fig. 11 shows that this property does

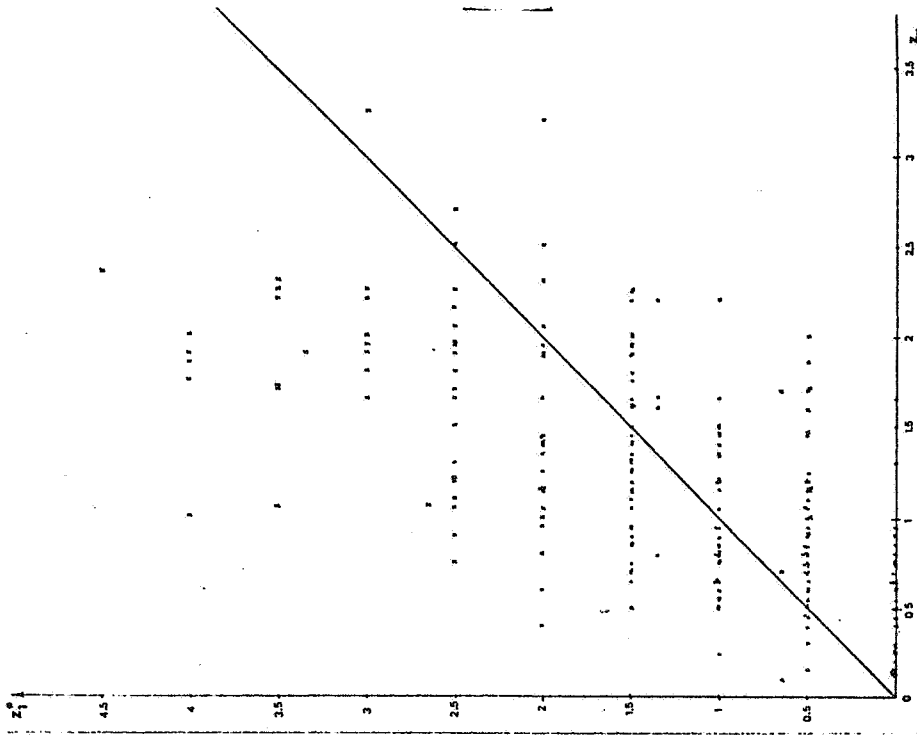


Fig. 11

Classical estimates/true values
Same panel geometry as in Fig. 10

(From Ph. NARBONI, 1979)

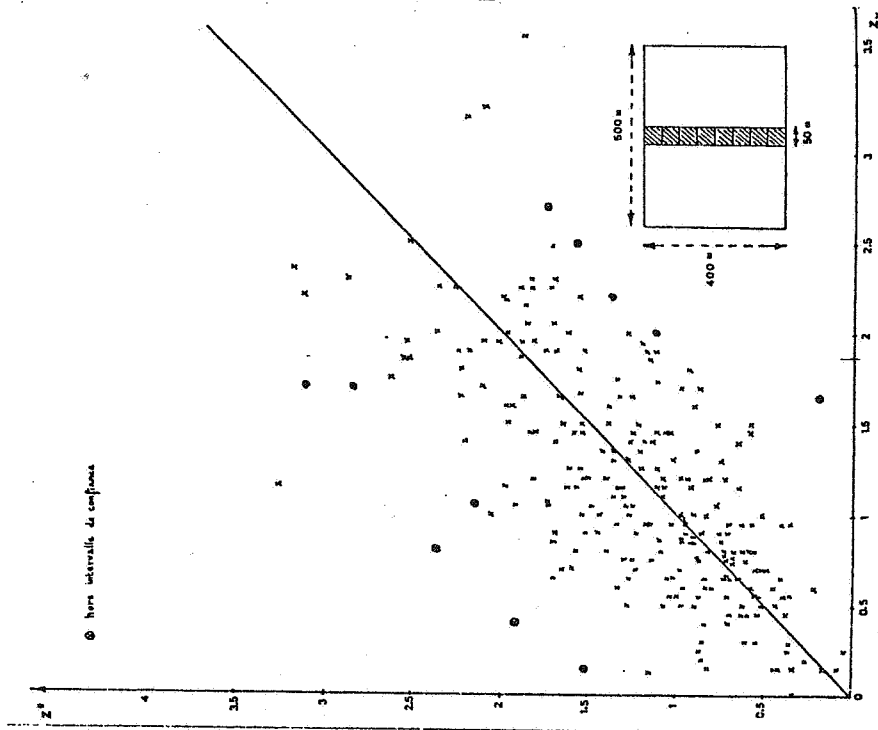
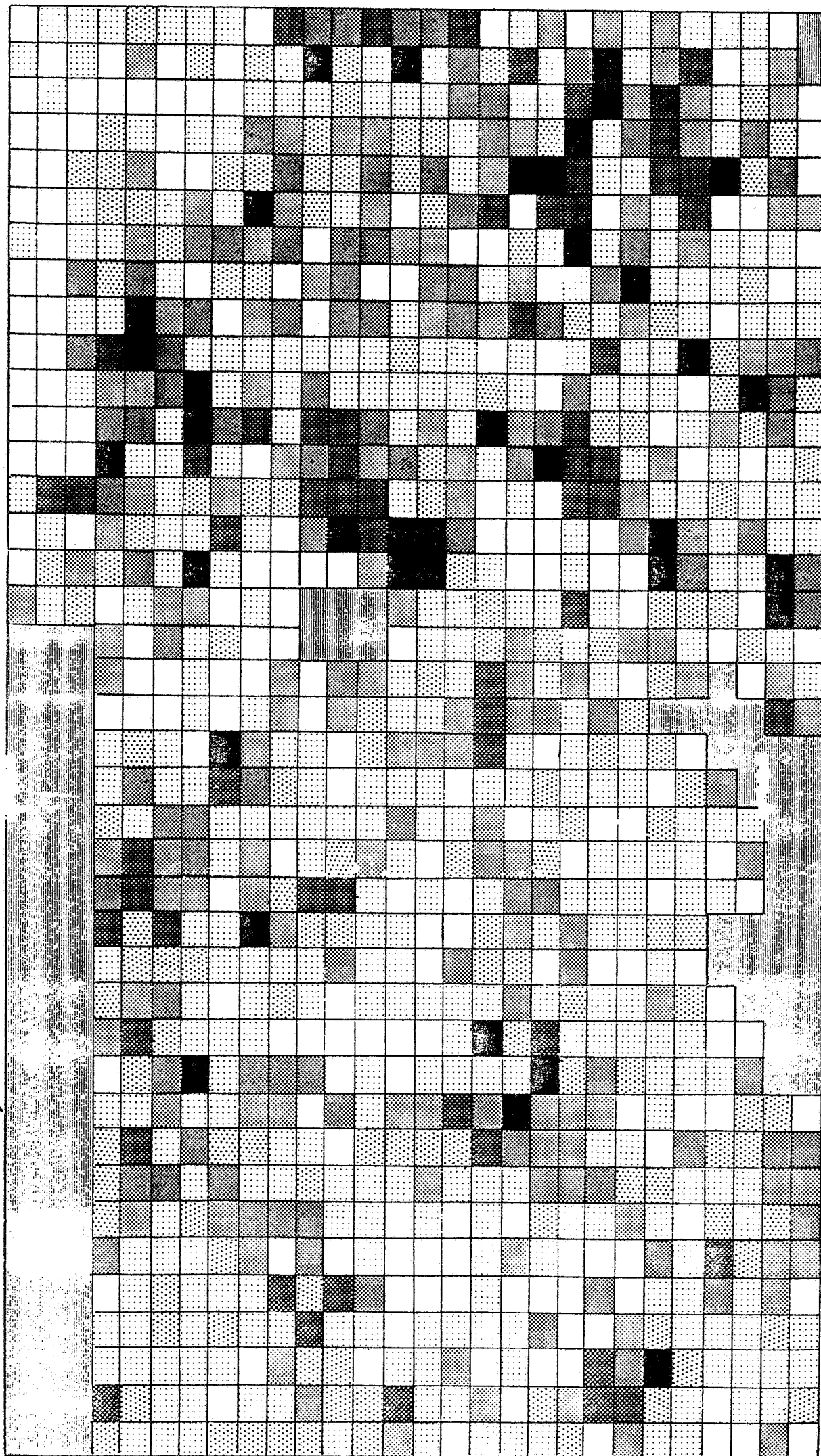


Fig. 10

Kriging estimates/true values;
Each point represents a 20 ha panel
10% sampling rate on NG0101



Densité Okoumé / ha

hors
comptage

> 4

3,01 à 4

2,01 à 3

1,51 à 2

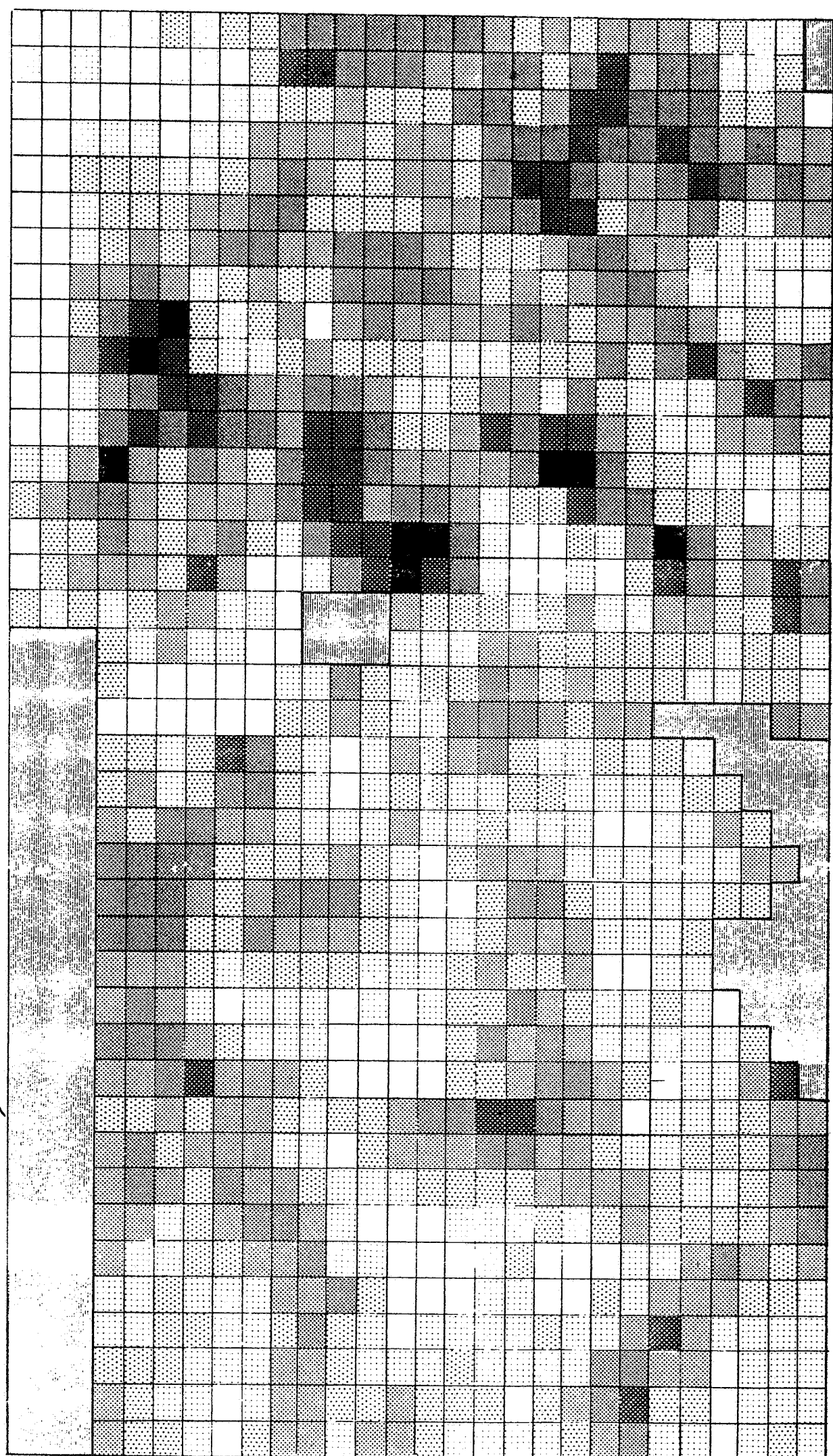
1,01 à 1,50

0,5 à 1

< 0,5

NGOIO Area - Kriging

Nord ↙



Densité Okoumé / ha



> 4

hors
comptage

3,01 à 4

2,01 à 3

1,51 à 2

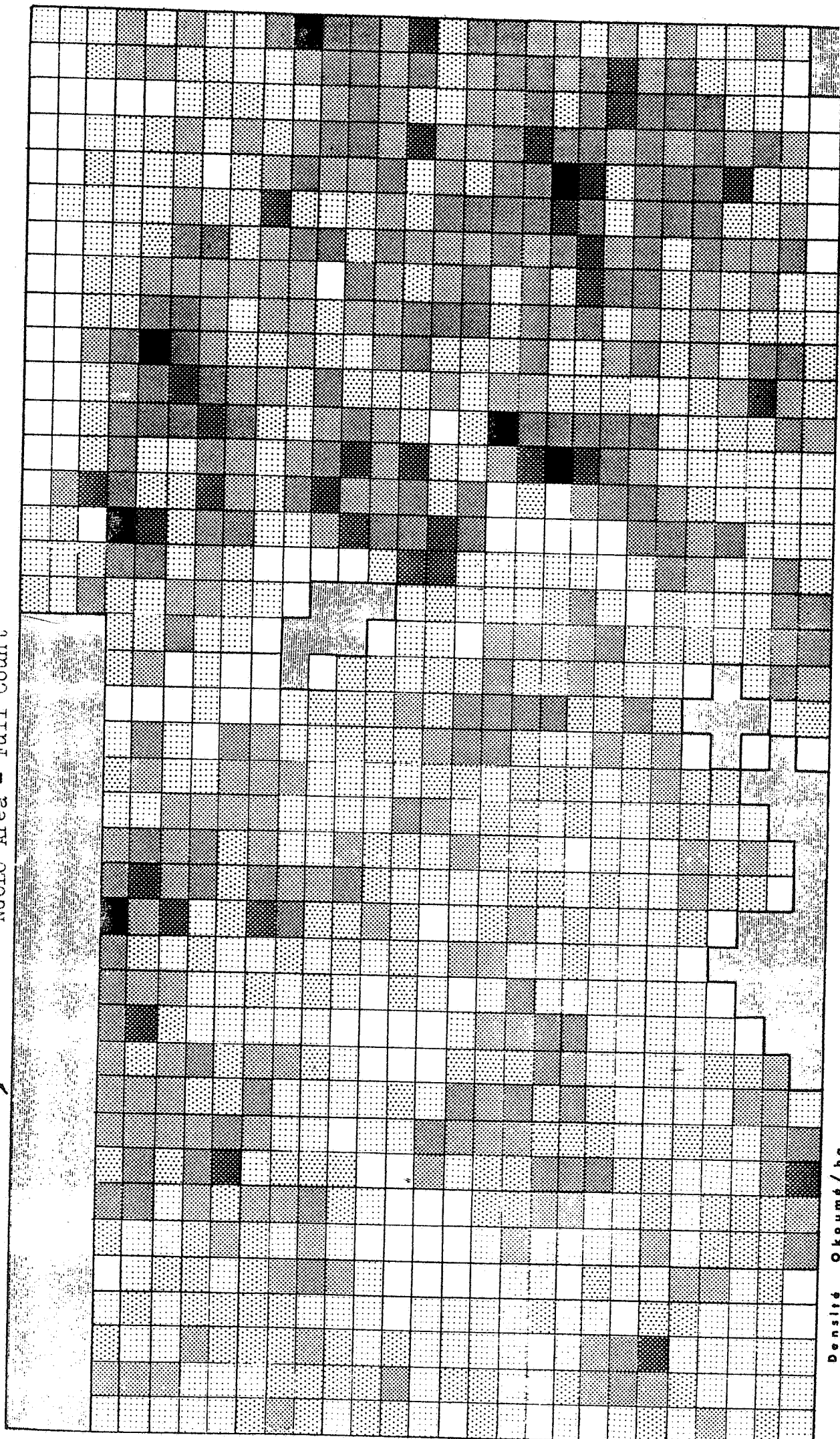
1,01 à 1,50

0,5 à 1

< 0,5

(From Ph. NARBONI, 1979)

NGOLO Area - Full Count



Densité Okoumé / ha

hors
comptage

> 4

3,01 à 4

2,01 à 3

1,51 à 2

1,01 à 1,50

0,5 à 1

< 0,5

not hold with the classical estimator Z_1 based on inside values. Even though it is globally unbiased, locally it has a tendency to overestimate panels such that $Z_1^* > m$ and to underestimate them when $Z_1^* < m$. This is exactly the regression effect mentioned in the introduction to kriging (attention : true values are now plotted horizontally).

It is even more spectacular to compare the estimates locally. Fig. 12 a, 12 b, 12 c are density plots symbolizing the counts in the panels. As Narboni puts it "the map obtained by the classical estimates is a patchwork that the eye can't integrate". Kriging produces smoother values, easier to interpret, while at the same time faithful to the true values.

7 - VARIANCE REDUCTION AS A GUIDE TO LOCATE A NEW OBSERVATION.

An interesting possibility is to use the kriging variance as an indicator of the information brought about by a new observation. (We have already seen this idea along with the example in V-4). The crux of the argument is that the kriging variance does not depend on the actual data values but only on their geometrical arrangement and on the variogram. As a consequence it is possible for example to "drill" a fictitious well and compute how much the variance decreases with this new information. The place where to drill is where the variance reduction is a maximum (assuming of course that the objective of drilling is to gain information. Whether that is the case is another problem).

The reader may try his own skill on Fig. 13 where the objective is to estimate the total rainfall over a basin.

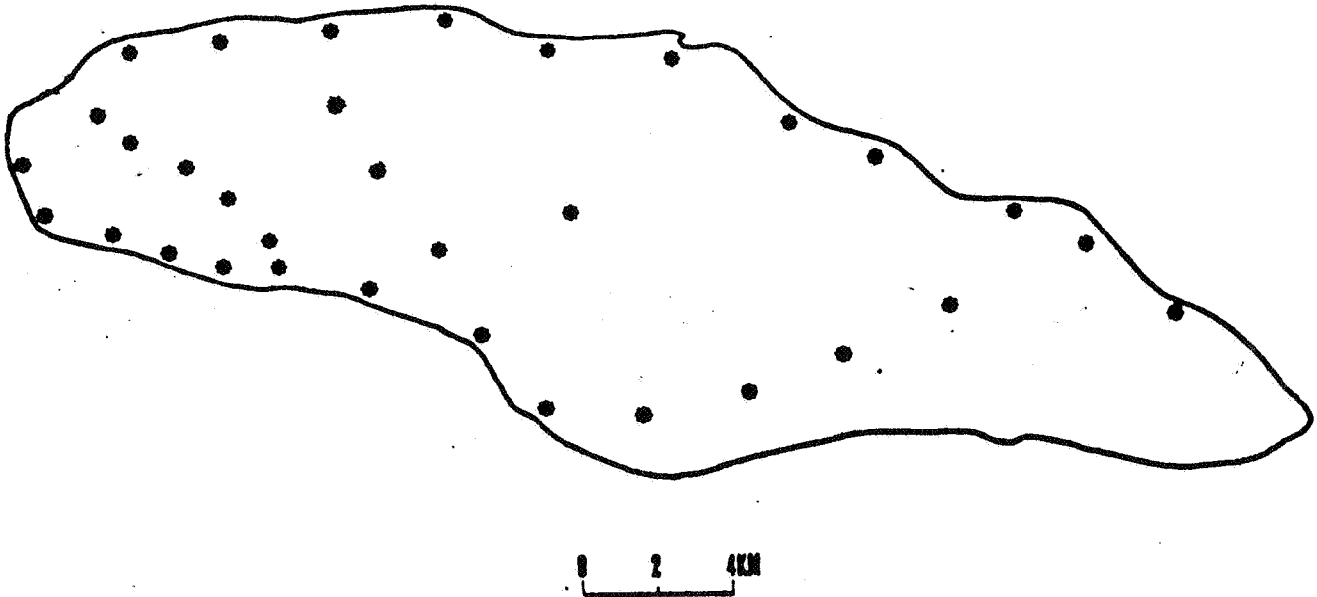


Fig. 13

"Where do you think a new rain gauge would be most informative?"

The gain of information (= relative variance reduction) has been contoured in Fig. 14. It shows that the best place is in the South-Eastern part of the basin (18.8 %) and not in the middle as could have been thought at first. Of course the objective of variance reduction may be counterbalanced by other factors, such as ease of access. Then Fig. 14 may be used to find a constrained optimum.

REINFORCEMENT OF A MEASUREMENT NETWORK

RELATIVE VARIANCE REDUCTION

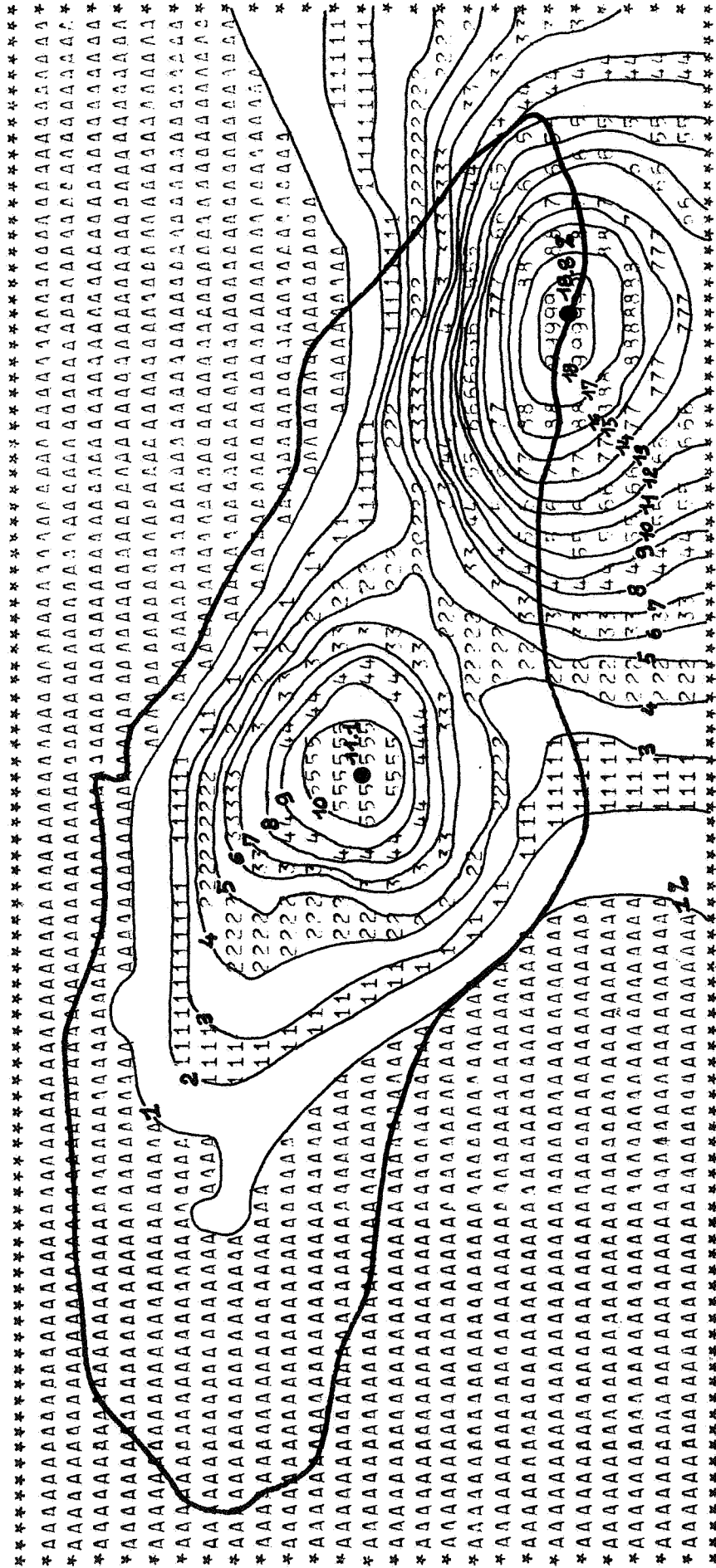


Fig. 14

Contour lines of the relative variance reduction on the estimation of average rainfall over the basin.

REFERENCES

The following list includes only the references which have actually been used to write this course.

- CAPON, J., 1969. "High-Resolution Frequency Wavenumber spectrum Analysis". Proc. IEEE, vol. 57, pp. 1408-1418.
- CAPON, J., R.J. GREENFIELD, and R.J. KOLKER, 1967. "Multidimensional Maximum-Likelihood Processing of a large Aperture Seismic Array". Proc. IEEE, Vol. 55, pp. 192-211.
- CHILES, J.P., 1976. "How to adapt kriging to non-classical problems - three case studies". In M. Guarascio et al, Eds., Advanced Geostatistics in the Mining Industry., Reidel, Dordrecht, the Netherlands, pp. 69-90.
- CHILES, J.P., 1977. Géostatistique des phénomènes non stationnaires. Unpublished Thesis., University of Nancy I, 152 p.
- DAVID, M., 1977. Geostatistical Ore Reserve Estimation. Elsevier, the Netherlands, 364 p.
- DELFINER, P., 1976. "Linear estimation of non-stationary spatial phenomena". In M. Guarascio et al., Eds., Advanced Geostatistics in the Mining Industry., Reidel, Dordrecht, The Netherlands, pp. 49-68.
- DELFINER, P., and J.P. DELHOMME, 1973. "Optimum Interpolation by Kriging". In J.C. Davis and M.J. McCullagh, Eds., Display and Analysis of Spatial Data, Wiley, London, pp. 96-114.
- DELFINER, P., J. ETIENNE, and J.M. FONCK, 1972. "Application de l'analyseur de textures à l'étude morphologique des réseaux poreux en lames minces". Rev. de l'Inst. Franç. du Pétrole, Vol. XXVII, n° 4, pp. 535-588.
- DELHOMME, J.P., 1974. "La cartographie d'une grandeur physique à partir de données de différentes qualités". In Mémoires de l'Assoc. Int. des Hydrogéologues, Tome X (1), pp. 185-194.
- DELHOMME, J.P., and P. DELFINER, 1973. "Application du krigeage à l'optimisation d'une campagne pluviométrique en zone aride". Proc. of the Symposium on the Design of Water Resources Projects with Inadequate Data (2), UNESCO, Madrid, pp. 191-210.

- DOOB, J.L., 1953. Stochastic Processes, Wiley, 654 p.
- HAAS, A. and C. JOUSSELIN, 1976. "Geostatistics in Petroleum Industry". In M. Guarascio et al., Eds, Advanced Geostatistics in the Mining Industry., Reidel, Dordrecht, the Netherlands, pp. 333-350.
- HUIJBREGTS, Ch., 1971. "Courbes d'Isovariance en Cartographie Automatique". Sciences de la Terre, Vol. XVI, pp. 291-301.
- HUIJBREGTS, Ch., 1975. "Three-day Short Course on Practical Mining Geostatistics". Clausthal, 2-4 Oct. 1975. Unpublished, 120 p.
- MATERN, B., 1960. Spatial Variation. Medd. Skogsforskningsinst., 49, 144 p.
- MATHERON, G., 1965. Les Variables Régionalisées et leur Estimation. Masson & Cie, Paris, 305 p.
- MATHERON, G., 1969. Le Krigeage Universel. Les Cahiers du Centre de Morphologie Mathématique, n° 1, Ecole Nat. Sup. des Mines de Paris, 82 p.
- MATHERON, G., 1971. The Theory of Regionalized Variables and its Applications. Les Cahiers du Centre de Morphologie Mathématique, n° 5, Ecole Nat. Sup. des Mines de Paris, 212 p.
- MOOD, A.M., and F.A. GRAYBILL, 1963. Introduction to the Theory of Statistics. McGraw and Hill, 443 p.
- NARBONI, Ph., 1979. "Application de la méthode des Variables Régionalisées à deux forêts du Gabon". Statistical Report n° 18, Centre Technique Forestier Tropical, 94130 Nogent-sur-Marne, France, 51 p.
- SERRA, J., 1967. Echantillonnage et estimation locale des phénomènes de transition miniers. Thesis, published by Institut de Recherche de la Sidérurgie, Maizières-lès-Metz, France, 670 p.
- SERRA, J., 1968. "Morphologie Mathématique et genèse des concrétions carbonatées des minerais de fer de Lorraine". Sedimentology, 10, pp. 183-208.
- YAGLOM, A.M., 1962. Stationary Random Functions. Dover Publications, New York, 253 p.

ADDITIONAL REFERENCES

Two bibliographic reviews of geostatistical publications available in English have been compiled. The first one covers the literature until mid 1977, and the second one until end of 1977.

PAUNCZ, I., 1978. "English Language Publications on the French School of Geostatistics". Mathematical Geology, Vol. 10, n°:2, pp. 253-260.

BELL, G.D. and M. REEVES., 1979. "Kriging and Geostatistics : A Review of the Literature Available in English"., Proc. Australas. Inst. of Min. Metall., n° 269, pp. 17-27.

More recent publications not mentioned in the above lists include

DEIHOME, J.P., 1978. "Kriging in the Hydrosiences", Advances in Water Resources, Vol. 1, n° 5, pp. 251-266.

JOURNEL, A.G. and Ch. HUIJBREGTS, 1978. Mining Geostatistics. Academic Press, 600 p.

MATHERON, G., 1979. Estimer et Choisir - Essai sur la Pratique des Probabilités. Gauthier-Villars, Paris. In press.

A RECOLLECTION OF SOME ELEMENTARY STATISTICAL RESULTS

I - SAMPLE POINTS, SAMPLE SPACE, EVENTS, PROBABILITIES, INDEPENDENCE.

Every possible outcome of an experiment is called a sample point. If we toss a coin there are two possible outcomes : heads or tails. If we roll a die there are six outcomes : 1, 2, 3, 4, 5 or 6. If we measure the length of an object there are theoretically infinitely many possible outcomes : all numbers from zero to infinity.

The set of all possible outcomes (or sample points) is called the sample space (usually denoted by Ω).

An event is a collection of sample points. For example in the die rolling experiment we may consider the event : "an odd face turns up" ; this event consists of 3 possible sample points : 1, 3 or 5. If we measure a length L we may consider the event : L is between 1.53 m and 1.79 m", which consists of infinitely many possible values in the interval $[1.53, 1.79]$.

Events may be transformed or combined and it is useful to view them as sets. For example :

LOGICAL NOTATION

NOT A (A false)
A or B
A and B
A impossible
A certain
A and B mutually exclusive
A implies B

SET NOTATION

A^c : A complement
 $A \cup B$: A union B
 $A \cap B$: A intersection B
 $A = \emptyset$: A is an empty set
 $A = \Omega$: A is the whole space
 $A \cap B = \emptyset$: A and B have an empty intersection
 $A \subset B$: A included in B

etc...

To every event A we may assign some numbers $P(A)$ between 0 and 1 called the probability $P(A)$ of that event. By convention we assign a probability 0 to impossible events and a probability 1

to any event which must happen with certainty. Also, the probability that one or the other of two mutually exclusive events A and B occurs is the sum of the probabilities that A occurs and B occurs. This must remain true if we consider the sum of any countable collection of events A_i . Mathematically these axioms of probability write :

$$a) \quad 0 \leq P(A) \leq 1$$

$$b) \quad P(\Omega) = 1 \text{ and } P(\emptyset) = 0$$

$$c) \quad P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$

provided $A_i \cap A_j = \emptyset$ if $i \neq j$

From b) it follows that the probability that A does not occur is simply $P(A^c) = 1 - P(A)$.

It is important to note that the additivity property c) only holds for mutually exclusive events. In general, if A and B are not mutually exclusive, we have :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

In words : the probability that A or B occurs is the probability that A occurs plus the probability that B occurs minus the probability that A and B occur simultaneously. This is made clear by Fig. 1

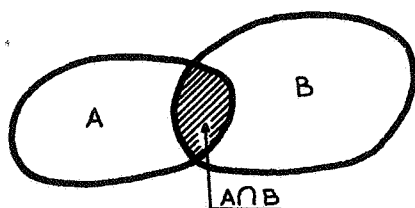


Fig. 1

When adding A and B we count the common region $A \cap B$ twice and must therefore subtract it out.

A fundamental notion in probability theory is that of independence. Two events A and B are independent if the probability that A and B occur is the product of $P(A)$ and $P(B)$:

$$A, B \text{ independent} \Leftrightarrow P(A \cap B) = P(A) \times P(B)$$

For example if we flip a fair coin twice the probability of getting Heads, Heads is $1/4$ (one out of the four sample points (H,H), (T,T), (H,T), (T,H)). But $1/4$ is equal to the product $1/2 \times 1/2$ of individual

probabilities so that the events "Heads at first toss", "Heads at second toss" are independent.

The intuitive idea of the independence of two events A and B is that knowing that A has occurred conveys no information on the event B. It is usually assumed for example that measurement errors are independent.

2 - RANDOM VARIABLES AND THEIR DISTRIBUTIONS - THE HISTOGRAM.

A random variable is a function which assigns a numerical value to each outcome of an experiment (= to each sample point in the sample space). For example if the experiment consists of tossing a coin, the score 1 may be assigned to the outcome Heads, while 0 is assigned to Tails. Then we have a random variable X which assumes the values :

$$\begin{array}{llll} 1 & \text{with probability} & 1/2 \\ 0 & " & " & 1/2 \end{array}$$

So, in an intuitive manner a random variable can be defined as a function which takes on certain values with certain probabilities. Usually a random variable is denoted by a capital letter, say X, while one of its "realizations", i.e. the result of a particular drawing, is denoted by a lower-case letter x.

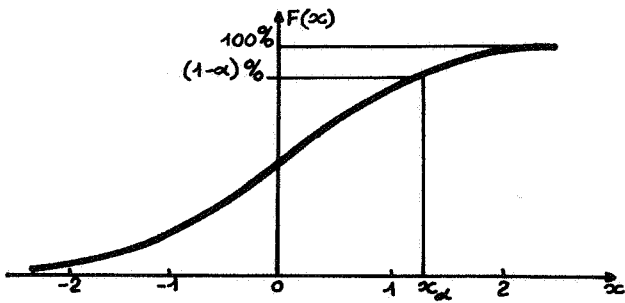
Independence of random variables goes as the independence of events concerning them. Let B_i be an interval on the line, or a countable union or intersection of intervals ; then the statement " X_i belongs to B_i ", denoted $(X_i \in B_i)$, is an event. By definition the random variables X_1, X_2, \dots, X_n are independent if probabilities multiply :

$$P\{(X_1 \in B_1) \cap (X_2 \in B_2) \cap \dots \cap (X_n \in B_n)\} = P(X_1 \in B_1) \times \dots \times P(X_n \in B_n)$$

Any random variable X is defined by its cumulative distribution function

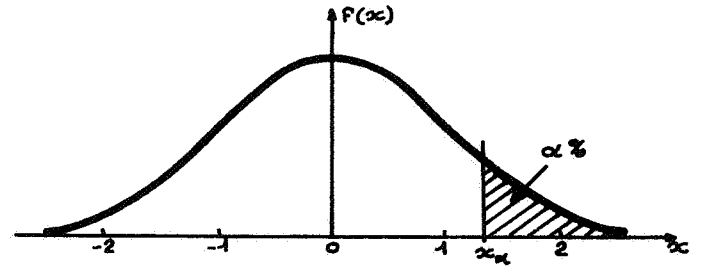
$$F(x) = P[X \leq x] \quad -\infty < x < \infty$$

which is the probability that X is less than or equal to some given value x. Fig. 2a shows an example of such a function F(x). By construction it increases monotonically from 0 to 1.



-Fig:2a- $P(X \leq x_\alpha) = (1-\alpha)\%$

Fig. 2a



-Fig:2b- shaded area = $P(X > x_\alpha)$

Fig. 2b

If we are interested in the probability that X lies in a given interval (a, b) , we have

$$P[a < X \leq b] = F(b) - F(a)$$

Now if the interval is small, say of length dx , the probability that X lies in $(x - \frac{dx}{2}, x + \frac{dx}{2})$ is :

$$P[x - \frac{dx}{2} < X \leq x + \frac{dx}{2}] = F(x + \frac{dx}{2}) - F(x - \frac{dx}{2}) \approx F'(x)dx$$

provided X is continuous. $F'(x)$ is the derivative of $F(x)$ and we call it $f(x)$:

$$f(x) = F'(x)$$

$f(x)$ is the density of probability at the point x . It should be pointed out that $f(x)$ itself is not a probability (it can well be greater than 1) ; only the area under the curve $f(x)$ is a probability. The relationship between F and f is illustrated on Fig. 2. $F(x)$ is the area under the curve $f(x)$ up to the point x :

$$F(x) = \int_{-\infty}^x f(x)dx \quad (\text{hence the name "cumulative"})$$

The way to estimate $f(x)$ from the data is to build a histo-gram, i.e. divide the range of values into classes and count the number of data in each class. The standard question is then : how many classes? If the classes are too wide, much information is lost

due to the smoothing of the histogram. On the other hand, if classes are too narrow the variability of the density estimate is too high. Statisticians formulate this by saying that "there is a conflict between bias and variance reduction". A practical recommendation is often to take classes with no less than 5 points in them.

Incidentally, for reservoir data the estimation of a probability density is not as simple as it may appear. The sampling may be very biased because the wells are drilled in areas that are believed to be favorable (anticlines, zones of high porosity and permeability). Also, as it will be seen later on, the various measurements in a field are not independent and do not carry the same amount of information so that some weighting of the data is necessary.

3 - MOMENTS : MEAN, VARIANCE, COVARIANCE AND CORRELATION.

The kth moment of a continuous random variable is by definition :

$$E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx \quad (\text{provided } E(|X|^k) < \infty)$$

The first moment called the mean or expected value is thus simply :

$$m = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

The symbol E denotes the operation of computing the expected value. This operation is linear, namely :

$$E(c X) = c E(X) \quad c = \text{constant}$$

$$E(X+Y) = E(X) + E(Y) \quad X, Y \text{ random variables}$$

Also we have of course

$$E(c) = c$$

The mean value of a random variable is a sort of theoretical average, or more precisely, the limit of the sample average when the sample size increases indefinitely. The mean can also be interpreted as the center of the distribution. Think of the x-axis as a bar with a variable density $f(x)$: then $E(X)$ is the center of gravity of the bar. For this reason the mean is often referred to as a location parameter - it tells where the center of the distribution is.

Other location parameters are also used, for example the median, e.g. the value such that 50% of the data are above and 50% are below. The mean however has a very strong physical appeal when we are dealing with additive quantities.

Another property of the expectation worth noticing is the following :

If X and Y are independent : $E(XY) = E(X)E(Y)$

We now turn to the key concept which we will use in the future : the concept of variance. First the formal definition : "the variance of a random variable X is the expected value of the squared deviations about its mean". It is denoted by $\text{Var}(x)$ or σ^2 for short. We have :

$$\begin{aligned}\text{Var}(X) = \sigma^2 &= E(X-m)^2 = \int_{-\infty}^{+\infty} (x-m)^2 f(x)dx = \int x^2 f(x)dx - m^2 \\ &= E(X^2) - [E(X)]^2\end{aligned}$$

An estimator of the variance is :

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{where } \bar{x} = \text{sample mean.}$$

(Statisticians sometimes recommend to divide by $(n-1)$ instead of n , but this is a debatable issue and it has no importance when n is large).

The main properties of the variance are the following :

- (a) $\text{Var}(c) = 0$ $c = \text{constant}$
- (b) $\text{Var}(X+c) = \text{Var}(X)$
- (c) $\text{Var}(cX) = c^2 \text{Var}(X)$
- (d) $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X,Y)$

(a) and (b) make sense if we are told that the variance is a measure of dispersion, or spread. (a) says that the dispersion of a constant is 0. (b) says that the dispersion is unaltered by a shift in all the data. (c) shows that the variance is in squared units since a change in the scale of X by a factor c changes the variance

by c^2 . To have the dispersion in the same unit as the data, one considers the square root σ of the variance, called the standard-dévi-
ation :

$$\sigma = \sqrt{\text{Var}(x)} = \text{standard deviation}$$

If most of the probability mass lies near the mean, the variance will be small, whereas if it is spread out over a large range, the variance will be large (cf. Fig. 3)

This statement is made more precise by Chebyshev's inequality. For any positive value t :

$$P(|X-m| \geq t \sigma) \leq \frac{1}{t^2}$$

If we take $t = 2$ for example we find that the probability that X deviates from the mean by more than 2 standard deviations is less than 25%. This is true whatever the distribution of X may be. Of course if we have more precise ideas about the distribution of X , the statement itself becomes more accurate (e.g. for a normal, the probability is in fact 5% instead of 25%).

Formula (d) is of capital importance and it is worthwhile to derive it from the definitions. By definition :

$$\text{Var}(X+Y) = E[(X+Y)-(m_X+m_Y)]^2 = E[(X-m_X) + (Y-m_Y)]^2$$

But $[(X-m_X)+(Y-m_Y)]^2 = (X-m_X)^2 + (Y-m_Y)^2 + 2 (X-m_X) (Y-m_Y)$. By linearity of the expectation we thus have :

$$\underbrace{E[(X-m_X)+(Y-m_Y)]^2}_{\text{Var}(X+Y)} = \underbrace{E(X-m_X)^2}_{\text{Var}(X)} + \underbrace{E(Y-m_Y)^2}_{\text{Var}(Y)} + 2 \underbrace{E[(X-m_X)(Y-m_Y)]}_{\text{Cov}(X,Y)}$$

The term $E[(X-m_X)(Y-m_Y)]$ is named the covariance between X and Y . It can be denoted by σ_{XY} . A property of the covariance is that :

$$|\text{cov}(X,Y)| \leq \sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}$$

so that the ratio :

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

is always between -1 and +1. ρ is the well known coefficient of correlation. It measures the mutual relationship between two variables. A positive value of ρ indicates a tendency for X and Y to increase together. When ρ is negative, large values of X are associated with small values of Y. When $\rho = \pm 1$ there is a perfect linear relationship between X and Y : $Y = \alpha X + \beta$. When X and Y are independent their correlation coefficient ρ (and likewise their covariance σ_{XY}) is 0. Caution! the converse is generally not true.

Formula (d) can be rewritten using ρ :

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2 \rho \sigma_X \sigma_Y$$

As a particular case when X and Y are independent we simply have :

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

Also by formula (c) if we take the average of n independent random variables with equal variances σ^2 we get :

$$\text{Var}\left(\frac{1}{n} \sum_i X_i\right) = \frac{1}{n^2} \sum_i \text{Var}(X_i) = \frac{n \text{Var}(X_1)}{n^2} = \frac{\sigma^2}{n}$$

The variance is divided by n. This is only true if the variables are independent, or at least if their correlation is 0.

We now give two approximate formulae that are useful.

. Variance of a product.

$$P = X Y \quad \text{with } E(Y) = m_Y > 0, E(X) = m_X > 0, E(P) = m_P > 0$$

$$\frac{\sigma_P^2}{m_P^2} \approx \frac{\sigma_X^2}{m_X^2} + \frac{\sigma_Y^2}{m_Y^2} + 2 \rho \frac{\sigma_X \sigma_Y}{m_X m_Y} \quad \left(\frac{\sigma_X^2}{m_X^2}, \frac{\sigma_Y^2}{m_Y^2} \text{ small} \right)$$

. Variance of a ratio.

$$R = \frac{X}{Y} \quad \text{with } E(R) = m_R > 0, E(X) = m_X > 0, E(Y) = m_Y > 0$$

$$\frac{\sigma_R^2}{m_R^2} \approx \frac{\sigma_X^2}{m_X^2} + \frac{\sigma_Y^2}{m_Y^2} - 2 \rho \frac{\sigma_X \sigma_Y}{m_X m_Y} \quad \left(\frac{\sigma_X^2}{m_X^2}, \frac{\sigma_Y^2}{m_Y^2} \text{ small} \right)$$

4 - THREE COMMON DISTRIBUTIONS

(a) Bernoulli.

It is a discrete random variable describing the presence or absence of a character.

$$\begin{aligned} X &= 1 && \text{with probability } p \\ X &= 0 && " \quad " \quad 1 - p \end{aligned}$$

Its mean and variance are :

$$E(X) = p$$

$$\text{Var}(X) = p(1-p)$$

(b) Normal.

$$\text{The density if } f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

where m is the mean and σ^2 the variance . A normal variable is often denoted by $N(m, \sigma^2)$. By using tables of the cumulative distribution function we can make statements like the following :

There are 90% chances that X lies in the interval	$[m-1.645\sigma, m+1.645\sigma]$
" " 95% " " " " " "	$[m-1.96\sigma, m+1.96\sigma]$
" " 97,5% " " " " " "	$[m-1.96\sigma, +\infty]$
" " 98% " " " " " "	$[m-2.326\sigma, m+2.326\sigma]$

(c) Lognormal.

A random variable X is lognormal if its logarithm $\log X$ is normal. Let $\log \gamma$ and σ^2 be the mean and variance of $\log X$:

$$\log X \sim N(\log \gamma, \sigma^2)$$

Then the transformation

$$Z = \frac{1}{\sigma} \log \frac{X}{\gamma}$$

transforms X into a standard normal $N(0,1)$. X itself is related to Z by :

$$X = \gamma e^{\sigma Z}$$

The density of X is :

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \frac{1}{x} e^{-\frac{1}{2} \left(\frac{\log x - \log \gamma}{\sigma} \right)^2}$$

Some examples are plotted on Fig. 4. However it is preferable to work on Z .

γ is the median of X . Its mean and variance are given by :

$$\begin{cases} E(X) = m = \gamma e^{\frac{\sigma^2}{2}} \\ \text{Var}(X) = \Sigma^2 = m^2(e^{\sigma^2} - 1) \end{cases}$$

The first formula show the effect of the skewness of the lognormal distribution : the mean is greater than the median by a factor $e^{\sigma^2/2}$ which increases with the logarithmic variance σ^2 . The variance of X is proportional to the square of the mean so that :

$$\frac{\Sigma^2}{m^2} = e^{\sigma^2} - 1$$

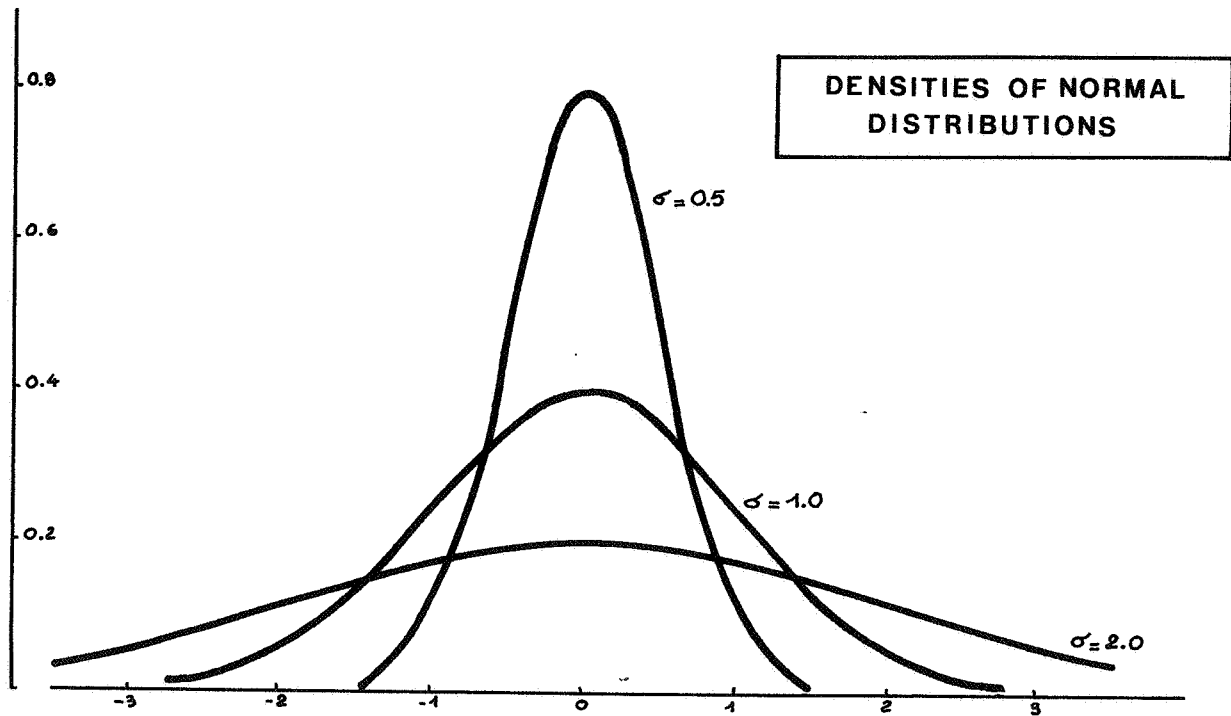
If σ^2 is small $e^{\sigma^2} - 1 \approx \sigma^2$ and the logarithmic variance is approximately equal to Σ^2/m^2 , an index of relative dispersion. ($\frac{\Sigma}{m}$ = coefficient of variation).

Probabilistic statements on X can easily be derived from those on the normal variable Z . For example there are 95% chances that

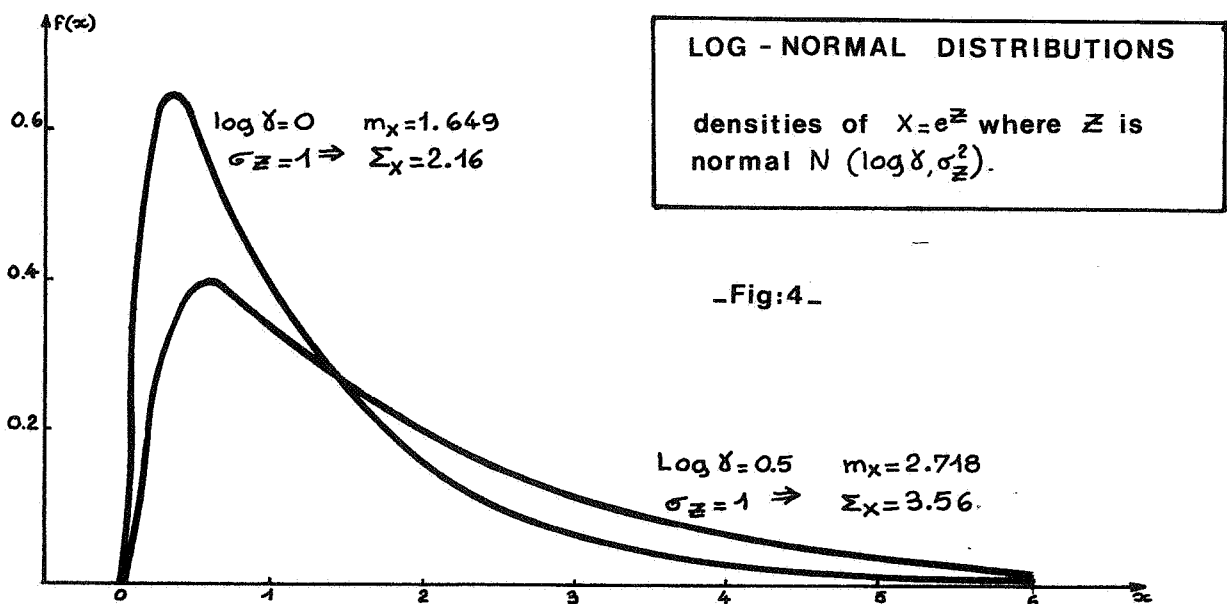
$$-1.96 \leq Z \leq 1.96 \Rightarrow \gamma e^{-1.96\sigma} \leq X \leq \gamma e^{1.96\sigma}$$

If σ is small $\sigma \approx \frac{\Sigma}{m}$ and the inequality is roughly :

$$\gamma e^{-2 \frac{\Sigma}{m}} \leq X \leq \gamma e^{2 \frac{\Sigma}{m}} \text{ etc..}$$



-Fig:3- The standard deviation as an index of dispersion



Similarly there are $97,5\%$ chances that X is greater than $\gamma e^{-2 \sum \frac{1}{m}}$
etc...

5 - THE BIVARIATE NORMAL DISTRIBUTION - REGRESSION - HIGHER DIMENSIONS.

Random variables can also be defined in several dimensions. A classical example is that of a rifle aimed at a target. The points of impact of the bullets on the target are characterized by 2 coordinates x and y . These can be considered as jointly defined random variables. Suppose that we divide the target into squares, count the number of hits in each square and compute the relative frequency. By plotting these frequencies we get a two-dimensional histogram (Fig. 5).

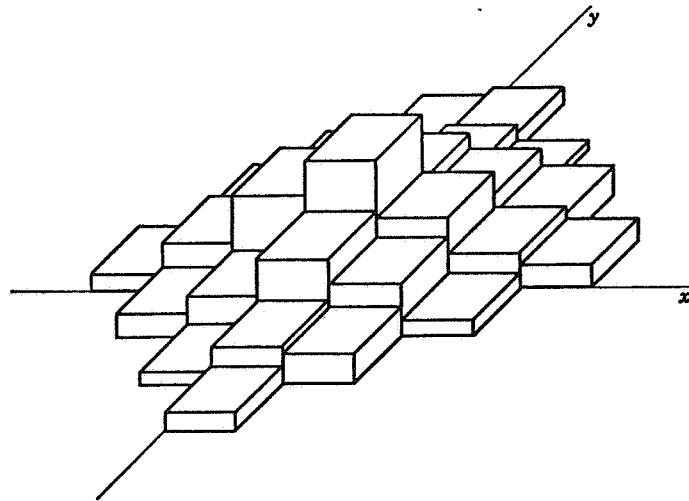


Fig. 5

Bivariate random variables can be defined in many cases of interpretation of reservoir data. A simple example is that of values of porosity as measured by two different types of logs (cf. "cross-plots").

A bivariate random variable is characterized by a two-dimensional cumulative distribution function :

$$F(x,y) = P[X \leq x \text{ and } Y \leq y]$$

Like in the univariate case there is generally a density $f(x,y)$ such that

$$F(x,y) = \int_{-\infty}^y dy \int_{-\infty}^x f(x,y) dx$$

The variable X considered alone has a distribution function

$$F_1(x) = P[X \leq x] = P[X \leq x, Y < \infty] = F(x, \infty)$$

Likewise Y alone has the distribution function

$$F_2(y) = P[Y \leq y] = P[X < \infty, Y \leq y] = F(\infty, y)$$

F_1 and F_2 are called the marginal distributions of X and Y respectively.

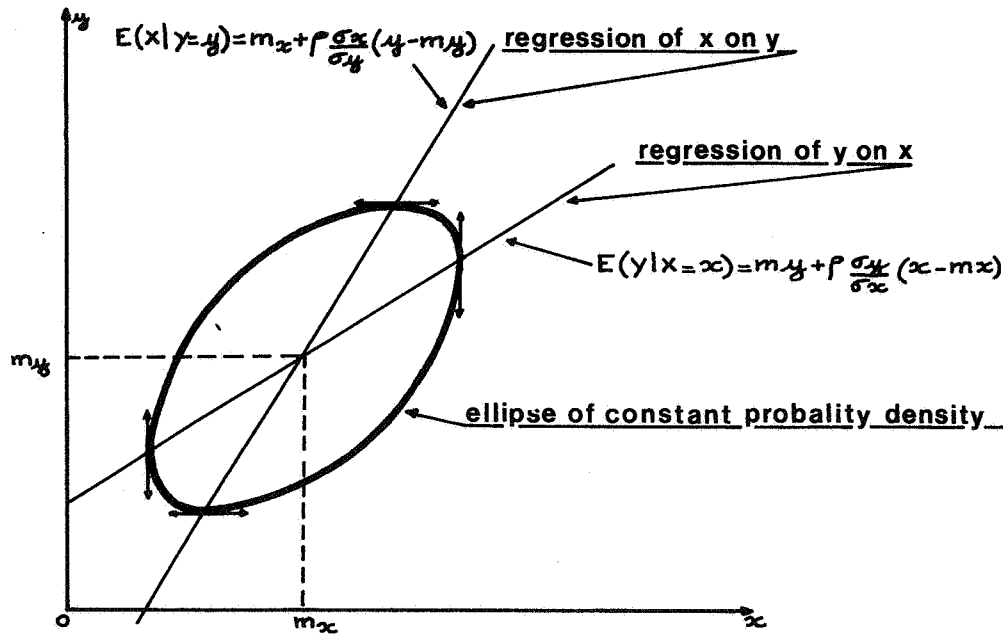
The bivariate normal has a density :

$$f(x,y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-m_x}{\sigma_x} \right)^2 - 2\rho \frac{(x-m_x)(y-m_y)}{\sigma_x \sigma_y} + \left(\frac{y-m_y}{\sigma_y} \right)^2 \right]}$$

where m_x and σ_x are the mean and standard-deviation of the marginal distribution of X which is itself univariate normal ; m_y and σ_y are the same for Y ; ρ is the correlation coefficient of X and Y . A nice property of the bivariate normal is that $\rho = 0$ implies independence of X and Y . On a cross-plot of X and Y lines of constant probability are ellipses centered at the point of coordinates (m_x, m_y) .

In many applications it is of interest to consider the mean value of one of the variables, say Y , when the other, say X , is fixed. This mean represents the best possible prediction of Y by a function of X , in the sense of least squares. For a bivariate normal the regression of Y on X is given by :

$$\underbrace{E(Y|X=x)}_{\text{read "expected value of Y given X"}} = m_y + \rho \frac{\sigma_y}{\sigma_x} (x - m_x)$$



-Fig:6- Regression lines for a bivariate normal

Fig. 6

It is a straight line passing through the center of gravity (m_x, m_y) and which cuts all vertical secants to the ellipses at their mid-points. It can be drawn easily by taking one ellipse and joining the two points where it admits a tangent parallel to the coordinate axis O_y (Fig. 6). Similarly the regression of X on Y is :

$$E(X|Y = y) = m_x + \rho \frac{\sigma_x}{\sigma_y} (y - m_y)$$

It must be noted that the two regression lines are different.

Instead of considering the average value of Y when X is known, we may also consider the whole conditional distribution of Y given $X = x$. Its density, denoted by $f(y|x)$ is related to the bivariate density $f(x,y)$ and the marginal density of X , $f_1(x)$ through :

$$f(y|x) = \frac{f(x,y)}{f_1(x)} \quad f_1(x) > 0$$

Applying this formula to the bivariate normal, we get :

$$f(y|x) = \frac{1}{\sqrt{2\pi} \sqrt{1-\rho^2} \sigma_Y} e^{-\frac{1}{2(1-\rho^2)\sigma_Y^2} [y - (m_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - m_X))]^2}$$

The conditional distribution of Y given X is again normal, with mean $E(Y|X=x)$ and variance $(1-\rho^2) \sigma_Y^2$. Note that this conditional variance does not depend on x.

Similar results are obtained for the conditional distribution of X given Y

Higher Dimensions.

The same approach can be used for higher dimensional random variables. To each point of the sample space one associates a set of k values $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$ (or a "vector" with k components). The joint cumulative distribution function and the joint density are defined as in 1 or 2 dimensions. The most common model is the multivariate normal. Call X an n-dimensional random vector, m its mean and V its variance-covariance matrix, i.e. :

$$V = E[(X-m)(X-m)^T]$$

then the density of X is :

$$f(X) = \frac{1}{(2\pi)^{n/2} |V|^{1/2}} e^{-\frac{1}{2} (X-m)^T V^{-1} (X-m)}$$

where $|V|$ stands for the determinant of the (nxn) matrix V.

Suppose now that X, m and V are partitioned as follows :

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad m = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

X_1 is a (kx1) vector and all other vectors and matrices are dimensioned accordingly.

Then the conditional distribution of X_1 given X_2 is a k-variate normal with mean vector :

$$E(X_1|X_2) = m_1 + V_{12} V_{22}^{-1} (X_2 - m_2)$$

and variance-covariance matrix :

$$\text{Cov}(X_1 | X_2) = V_{11} - V_{12} V_{22}^{-1} V_{21}$$

Note again :

- i) that the regression of X_1 against X_2 is linear in X_2
- ii) that the covariance matrix of X_1 given X_2 does not depend on X_2

As a consequence, if $X_1, X_2, \dots, X_n, X_{n+1}$ are jointly normally distributed, the regression of X_{n+1} on X_1, X_2, \dots, X_n is of the form :

$$E(X_{n+1} | X_1 = x_1, \dots, X_n = x_n) = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_n x_n$$

and the conditional variance of X_{n+1} about this value does not depend on x_1, x_2, \dots, x_n .

6 - A COUNTER-EXAMPLE OF LINEARITY : REGRESSION THEORY FOR THE BIVARIATE LOGNORMAL DISTRIBUTION.

To emphasize the fact that the above results are properties of the normal distribution, we consider the simple (and useful) case where X and Y are jointly normally distributed, i.e. $\text{Log } X$ and $\text{Log } Y$ are bivariate normal :

$$\text{Log } X \sim N(\text{Log } \gamma_x, \sigma_x^2)$$

$$\text{Log } Y \sim N(\text{Log } \gamma_y, \sigma_y^2)$$

$$\rho = \text{corr}(\text{Log } X, \text{Log } Y)$$

From the results on the bivariate normal we have that the conditional distribution of Y given X is lognormal with logarithmic mean :

$$E(\text{Log } Y | \text{Log } X = \text{Log } x) = \text{Log } \gamma_y + \rho \frac{\sigma_y}{\sigma_x} (\text{Log } x - \text{Log } \gamma_x)$$

$$= \text{Log} \left[\gamma_y \left(\frac{x}{\gamma_x} \right)^{\rho \frac{\sigma_y}{\sigma_x}} \right]$$

and logarithmic variance $(1-\rho^2) \sigma_y^2$.

Then, the arithmetic mean of this lognormal distribution is

$$E(Y|X=x) = \gamma_y \left(\frac{x}{\gamma_x}\right)^{\rho \frac{\sigma_y}{\sigma_x}} e^{\frac{(1-\rho^2)\sigma_y^2}{2}}$$

We see that in this case the regression curve is a power function and not a straight line (except when $\rho \frac{\sigma_y}{\sigma_x} = 1$).

To calculate the variance about the regression we make use of the following formula, easily established :

$$\text{if } \log X \sim N(\log \gamma, \sigma^2) \quad \text{then } E(X^\alpha) = \gamma^\alpha e^{\frac{\alpha^2 \sigma^2}{2}} \quad \text{for any } \alpha.$$

Therefore :

$$E(Y^2|X=x) = \left[\gamma_y \left(\frac{x}{\gamma_x}\right)^{\rho \frac{\sigma_y}{\sigma_x}} \right]^2 e^{2(1-\rho^2)\sigma_y^2}$$

and as :

$$[E(Y|X=x)]^2 = \left[\gamma_y \left(\frac{x}{\gamma_x}\right)^{\rho \frac{\sigma_y}{\sigma_x}} \right]^2 e^{(1-\rho^2)\sigma_y^2}$$

we have :

$$\text{Var}(Y|X=x) = [E(Y|X=x)]^2 [e^{(1-\rho^2)\sigma_y^2} - 1]$$

This time the variance does not depend on x : it is proportional to the squared mean.

Incidentally, we find, as could be expected, that what matters in the Lognormal case is the relative dispersion, i.e. ratio of variance to squared mean. When σ_y^2 is small this ratio is approximately $(1-\rho^2)\sigma_y^2$, something not unfamiliar.

If instead of the true regression curve, we took the linear formula of the bivariate normal, we would have :

$$\hat{y} = m_y + R_o \times \frac{\sum y}{\sum x} (x - m_x)$$

where R_o stands for the correlation between X and Y (ρ is the correlation between $\log X$ and $\log Y$). Explicitly :

$$R_o = \frac{e^{\rho \sigma_x \sigma_y} - 1}{(\sigma_x^2 - 1)^{\frac{1}{2}} (\sigma_y^2 - 1)^{\frac{1}{2}}}$$

When σ_x and σ_y are small, R_o is close to ρ and :

$$\hat{y} \approx m_y + \rho \frac{\sigma_y}{\sigma_x} \frac{m_y}{m_x} (x - m_x)$$

It is the same formula as in the normal case, except for the factor $\frac{m_y}{m_x}$. The graphical relationships between the true regression curve and its linear substitute are epitomized on Fig. 7

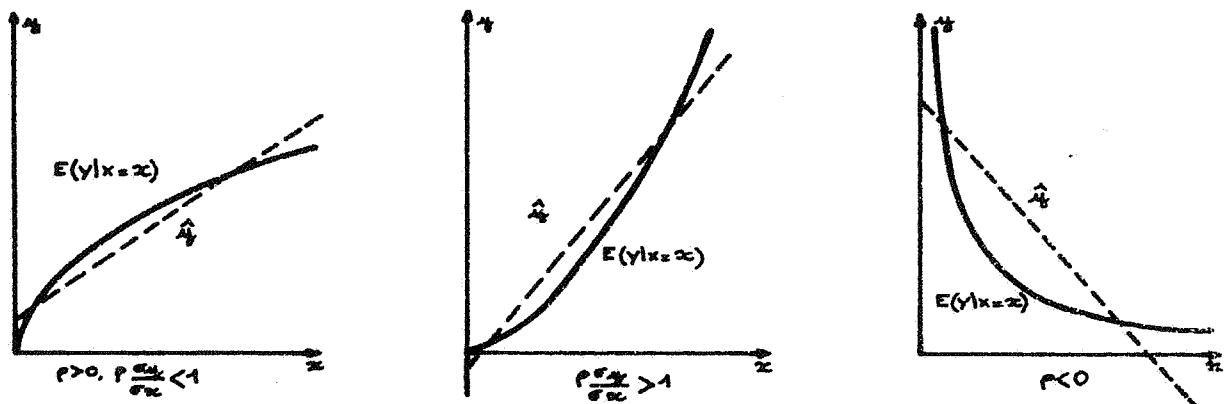


Fig. 7 : Regression curves for a bivariate lognormal

7 - ESTIMATION - CONFIDENCE INTERVALS.

Suppose that we have n values of a quantity, say Z_1, Z_2, \dots, Z_n at n points x_1, x_2, \dots, x_n . We want to estimate the value at a new point x_o . For this, we consider a function of the observations

$$\hat{Z}(x_o) = T(Z_1, Z_2, \dots, Z_n)$$

which will be an estimator of our unknown quantities $Z(x_0)$. When we insert particular numerical values for the Z_i 's we obtain a numerical value of the function $T(\cdot)$, which is called an estimate.

What criterion should we use for choosing an estimator?
Of course ideally we would like that

$$\hat{Z}(x_0) = Z(x_0) \text{ whatever } Z(x_0) \text{ is.}$$

But clearly this is asking for too much. So we are led to ask from $\hat{Z}(x_0)$ to be close to $Z(x_0)$ in some statistical sense. An obvious idea is to require that "on average" the magnitude of the error $|\hat{Z}(x_0) - Z(x_0)|$ be small. However this formulation is rather untractable mathematically and it is easier to concentrate on the squared error $(\hat{Z}(x_0) - Z(x_0))^2$. Our criterion is then to minimize the mean squared error :

$$E(\hat{Z}(x_0) - Z(x_0))^2$$

Now if we recall what we have learnt on the variance we see that :

$$\underbrace{E(\hat{Z}(x_0) - Z(x_0))^2}_{\text{mean squared error}} = \underbrace{\text{Var}(\hat{Z}(x_0) - Z(x_0))}_{\text{Variance}} + \underbrace{[E(\hat{Z}(x_0) - Z(x_0))]^2}_{(\text{bias})^2}$$

The quantity $E(\hat{Z}(x_0) - Z(x_0))$ is called the bias. Usually we shall require the bias to be zero, otherwise it would mean that our estimator is systematically smaller or larger than the true value. So if :

$$E[\hat{Z}(x_0) - Z(x_0)] = 0 \quad (\text{unbiasedness})$$

we have :

$$E(\hat{Z}(x_0) - Z(x_0))^2 = \text{Var}(\hat{Z}(x_0) - Z(x_0))$$

Confidence Intervals.

An estimate is not a true value and we may lead ourselves into error if we act as if the true value of the quantity were

equal to our estimate. To be safer, we want an interval which we can be reasonably confident will actually include the true value. This interval is called a confidence interval.

With Chebyshev's inequality we have a crude way to build a confidence interval. Assume σ_E^2 is the variance of $\hat{Z}(x_0) - Z(x_0)$, then :

$$\text{Prob}(|\hat{Z}(x_0) - Z(x_0)| \geq 2 \sigma_E) \leq 25\%$$

So, with 75% confidence we can assert that the interval

$$[\hat{Z}(x_0) - 2 \sigma_E, \hat{Z}(x_0) + 2 \sigma_E]$$

will contain the true value $Z(x_0)$. If we now assume that the error $\hat{Z}(x_0) - Z(x_0)$ is in fact normally distributed, our confidence in the previous interval rises to 95%.

It can be checked from probability tables (e.g. Biometrika Tables, Vol. II) that for many common distributions the interval $[m - 2\sigma, m + 2\sigma]$ contains 95% of the probability mass, so that this confidence interval will remain valid under a broad variety of assumptions on the distribution of the error.

8 - LINEAR PREDICTION BY LEAST-SQUARES - GENERAL THEORY - FITTING A STRAIGHT LINE

General Theory.

In the customary linear model each observation $y_i = y(x_i)$ is regarded as an outcome of a random variable $Y_i = Y(x_i)$, whose distribution depends on non-random parameters which are the coordinates of the point x_i . It is assumed that these random variables are uncorrelated, have a common variance σ^2 , and have means of the form :

$$E(Y_i) = \sum_{\ell=1}^k a_{\ell} f^{\ell}(x_i)$$

The f^{ℓ} ($\ell = 1, 2, \dots, k$) are given functions (monomials, for example)

and the a_ℓ are unknown coefficients. $E(Y_i)$ is thus linear with respect to the $f^\ell(x_i)$.

In full, the model can be written as the sum of a deterministic function and an error term ε :

$$Y_i = \sum_{\ell} a_{\ell} f_i^{\ell} + \varepsilon_i$$

with

$$f_i^{\ell} = f^{\ell}(x_i), E(\varepsilon_i) = 0, E(\varepsilon_i^2) = \sigma^2, E(\varepsilon_i \varepsilon_j) = 0 \quad \text{if } i \neq j$$

If the parameters f_i^{ℓ} are viewed as outcomes of random variables, $a_{\ell} f_i^{\ell}$ is the regression of Y_i on these variables.

When we substitute estimated \hat{a}_{ℓ} in place of the true a_{ℓ} we get predictors of the y_i 's :

$$\hat{y}_i = \sum_{\ell} \hat{a}_{\ell} f_i^{\ell}$$

Least squares estimation of the a_{ℓ} consists in finding those coefficients \hat{a}_{ℓ} which minimize the sum of squared deviations (also called "residuals") between the observed y_i and their predictions \hat{y}_i :

$$Q = \sum_i (y_i - \hat{y}_i)^2$$

To minimize $Q = \sum_i (y_i - \sum_{\ell} \hat{a}_{\ell} f_i^{\ell})^2$ it suffices to cancel the partial derivatives of Q with respect to the \hat{a}_{ℓ} :

$$-\frac{1}{2} \frac{\partial Q}{\partial \hat{a}_{\ell}} = 0 \Rightarrow \sum_i f_i^{\ell} (y_i - \sum_s \hat{a}_s f_i^s) = 0 \quad (\ell = 1, 2, \dots, k)$$

One gets the following set of linear equations (called "normal equations") :

$$\sum_s \hat{a}_s \sum_i f_i^{\ell} f_i^s = \sum_i f_i^{\ell} y_i \quad (\ell = 1, 2, \dots, k)$$

This system has a unique solution provided that the $k \times k$ matrix of terms $\sum_i f_i^{\ell} f_i^s$ is non singular. This condition is satisfied

if and only if the k vectors f_i^ℓ are linearly independent, i.e.
if

$$\sum_{\ell} c_{\ell} f_i^{\ell} = 0 \text{ for all } i \Rightarrow c_{\ell} = 0 \text{ for all } \ell$$

The least-squares estimators thus obtained are unbiased and moreover, among all linear unbiased estimators they are the best, in the sense that no other estimator has smaller variance (Gauss-Markov theorem).

As its minimum the sum of squares Q satisfies :

$$E(Q) = (n-k)\sigma^2$$

so that an unbiased estimate of the variance σ^2 is :

$$s^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{n-k}$$

It is of interest to note that premultiplication of the normal equations by \hat{a}_{ℓ} and summation over ℓ yields :

$$\sum_i \hat{y}_i (\hat{y}_i - y_i) = 0$$

In other words, the vector of residuals $\hat{y}_i - y_i$ is orthogonal to the vector of predicted values \hat{y}_i . This entails the following decomposition :

$$\sum_i y_i^2 = \sum_i \hat{y}_i^2 + \sum_i (y_i - \hat{y}_i)^2$$

If, as in normal practice, we include a constant term a_1 in the regression equation (and thus set $f^1(x) = 1$ for all x), we get :

$$\underbrace{\sum_i (y_i - \bar{y})^2}_{\text{sum of squares about the mean}} = \underbrace{\sum_i (\hat{y}_i - \bar{y})^2}_{\text{sum of squares due to regression}} + \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{\text{sum of squares about regression}}$$

The ratio $R^2 = \text{Sum of squares due to regression} / \text{Sum of squares about the mean}$ measures the proportion of total variation about

the mean explained by the regression. We are pleased if R^2 is close to unity. (Note that R is also called the coefficient of multiple correlation between the y_i 's and the \hat{y}_i 's).

A great deal of distributional results and hypothesis testing procedures can be derived under the assumption that the errors are normally distributed. But this would lead us too far.

Fitting a straight line.

An important application of the above approach is the derivation of parameter values by indirect means. Example : assessment of porosity using density, neutron or sonic logs. Often the relationship can be modeled simply by a straight line :

$$Y_i = a x_i + b + \varepsilon_i$$

It is convenient to reparametrize this model as :

$$y_i = E(\bar{Y}) + a(x_i - \bar{x}) + \varepsilon_i$$

(In the notation of the general theory we have here $a_1 = E(\bar{Y})$, $f_1^1 = 1$; $a_2 = a$, $f_1^2 = x_i - \bar{x}$). The least squares estimates of $E(\bar{Y})$ and a are :

$$\widehat{E(\bar{Y})} = \bar{y} \quad , \quad \hat{a} = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2} \quad (= \hat{\rho} \frac{s_y}{s_x})$$

These estimators are unbiased and their variances are :

$$\text{Var}(\bar{Y}) = \frac{\sigma^2}{n} \quad , \quad \text{Var}(\hat{a}) = E(\hat{a} - a)^2 = E \left(\frac{\sum_i (x_i - \bar{x}) \varepsilon_i}{\sum_i (x_i - \bar{x})^2} \right)^2 = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$$

Also, it can be seen that $\text{Cov}(\bar{Y}, \hat{a}) = 0$ (this is the benefit of reparametrization).

Consider now that we have measured x_0 but not y_0 . ($x_0 \neq x_i$, all i). As an estimate of y_0 we take :

$$\hat{y}_0 = \bar{y} + \hat{a} (x_0 - \bar{x})$$

This estimate is unbiased. Its variance is :

$$\text{Var}(\hat{Y}_0) = \text{Var}(\bar{Y}) + (x_0 - \bar{x})^2 \text{Var}(\hat{a})$$

So

$$E[\hat{Y}_0 - E(\hat{Y}_0)]^2 = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right]$$

This is the variance of the error that we make when predicting the expected value of $Y(x_0)$ at the point x_0 . It is a minimum at $x_0 = \bar{x}$ and increases as x_0 moves away from \bar{x} , in either direction. Our predictor is best in the "middle" of the observations.

However, it must be noticed that in practice we are interested in the actual value y_0 and not its mean $E(Y_0)$ ($= E(\hat{Y}_0)$). So the error we make in fact is :

$$\hat{Y}_0 - Y_0 = [\hat{Y}_0 - E(Y_0)] - [Y_0 - E(Y_0)] = [\hat{Y}_0 - E(\hat{Y}_0)] - \varepsilon_0$$

ε_0 is uncorrelated with Y_0 which depends on other ε_i 's, so :

$$E(\hat{Y}_0 - Y_0)^2 = E(\hat{Y}_0 - E(\hat{Y}_0))^2 + \sigma^2$$

Finally the estimation variance of Y_0 is :

$$E(\hat{Y}_0 - Y_0)^2 = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right]$$

For numerical calculations σ^2 is replaced by its estimate :

$$s^2 = \frac{1}{n-2} \sum_i (y_i - \hat{y}_i)^2$$

- CHAPTER VI -

CONDITIONAL SIMULATIONS OF RESERVOIR BOUNDARIES

I - WHY SIMULATIONS?

We have seen how kriging techniques provide efficient estimates in problems involving linear functions of the variable under study. But all problems are not linear and there are cases when we would like more than just estimates.

A typical non linear problem is that of the cut-off of the reservoir by a water level. The effective volume is determined by integration of the effective thickness h , i.e. a truncated quantity of the form :

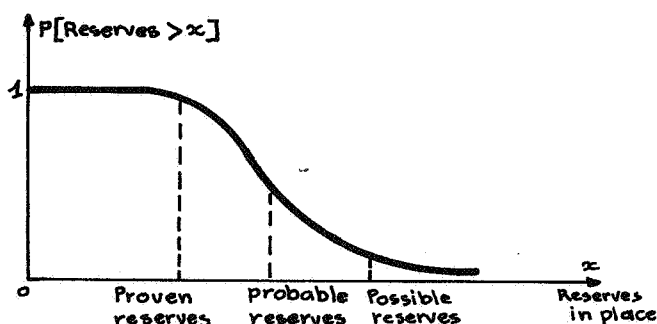
$$\begin{aligned} h &= Z_{WL} - Z(x) & \text{if } Z(x) \leq Z_{WL} \\ h &= 0 & \text{if } Z(x) > Z_{WL} \end{aligned}$$

where $Z(x)$ is the depth of the top at a point x and Z_{WL} the depth of the water level. Of course having values $Z(x_i)$ it is always possible to grid the reservoir top by kriging and compute the volume comprised between this upper boundary and a given water level. However, there is no guarantee of unbiasedness. Kriging only ensures that, on average, the estimate $Z^*(x)$ will be equal to the true depth $Z(x)$, irrespective of what this depth is. Nothing in the equations implies that this property remains true when we put restrictions on the values themselves. The equations have just not been designed for that. It goes without saying that there is no reason why any of the common methods of interpolation should perform better in this situation ; rather, we believe, it would perform worse.

A crude way to take into account the effect of geometric uncertainty on reserves evaluation is to work under 3 different hypotheses :

- (i) an optimistic one : top lifted up as much as possible, water level at its lowest.
- (ii) a pessimistic one : top pushed down, water level raised.
- (iii) a median one.

This gives an upper and a lower bound, and a median estimate. Ideally, what the company's economists would like for their calculations is the expectation curve of the oil or gas in place, i.e. for each quantity x the probability that the reserves in place are greater than x .



In order to compute such probabilities we must appeal to a model. One model, used implicitly in the so-called "Error-Analysis", is that of total independence between reservoir parameters values : independence between variables and between grid points. From the histograms of the different variables (thickness, porosity, water saturation, etc...) equally probable values are established and then drawn at random for each grid point. But in trying to simulate errors this procedure ignores completely the spatial structures of the variables. The model of random functions that we have been using throughout seems more adapted.

2 - THE PRINCIPLES OF CONDITIONAL SIMULATIONS.

The basic assumption in all geostatistical operations was to consider the phenomenon under study as a particular realization of a random function. This was a device to set the problems in a probabilistic framework and thus be able to make use of convenient tools like expectations, variances, covariances, correlations, variograms, etc... But in fact the profound nature of a random function, which is an ensemble of possible realizations, has not been utilized. The idea with simulations is to exhibit other possible realizations of the random function. While kriging, as any other interpolation procedure, gives a smoothed picture of reality, simulations display the same amount of spatial variability that can be expected from the actual phenomenon. Yet, for these simulations to be plausible candidates of reality, they also have to take at data points the values that have been measured. This leads to try to construct what is called "conditional simulations", i.e. functions (or, equivalently, surfaces) that satisfy two conditions :

- (i) have the same covariance or variogram as the data.
- (ii) pass through the given data points.

The term "simulation" here should be understood in a purely statistical context and has nothing to do with "dynamic reservoir simulation".

It is not possible here to enter into the technicalities of conditional simulations. We shall just outline the principles and then present an example of application to the estimation of hydrocarbon reserves.

We start from the following trivial decomposition :

$$\text{(Eq 1) : } \underbrace{z(x)}_{\text{true value}} = \underbrace{z^*(x)}_{\text{kriging estimate}} + \underbrace{[z(x) - z^*(x)]}_{\text{error}}$$

(We use lower-case letters for realizations and capital letters for the corresponding random functions)

In Eq 1 the error $z(x) - z^*(x)$ remains unknown because the true value $z(x)$ is not available. Suppose now that we are able to simulate a realization of a random function $S(x)$ having the same covariance as $Z(x)$. Then, on a particular simulation $s(x)$ it is possible, with the same pattern of data points, to compute kriging estimates and write similarly :

$$\text{(Eq 2) : } s(x) = s^*(x) + [s(x) - s^*(x)]$$

This time the error $s(x) - s^*(x)$ can be known exactly. The idea is to substitute in Eq 1 this error measured on $s(x)$ and define a function $z_s(x)$ as :

$$\text{(Eq 3) : } z_s(x) = z^*(x) + [s(x) - s^*(x)]$$

We claim that $z_s(x)$ is a conditional simulation. First we have to show that at a given point x_i , $z_s(x_i)$ is equal to the true value $z(x_i)$. This is obvious since at sample points kriging estimates coincide with the actual values, thus :

$$z^*(x_i) = z(x_i)$$

$$s^*(x_i) = s(x_i)$$

so in Eq 3 we have $z_s(x_i) = z(x_i)$. It remains to show that as a random function :

$$Z_s(x) = Z^*(x) + [S(x) - S^*(x)]$$

$Z_s(x)$ has the same covariance or covariogram as $Z(x)$. This

is true and we shall admit it without formal proof. The argument is based on a characteristic property of the kriging estimator to be uncorrelated with the error, so that we are entitled to pick an error from a simulation and bluntly add it to the actual estimate.

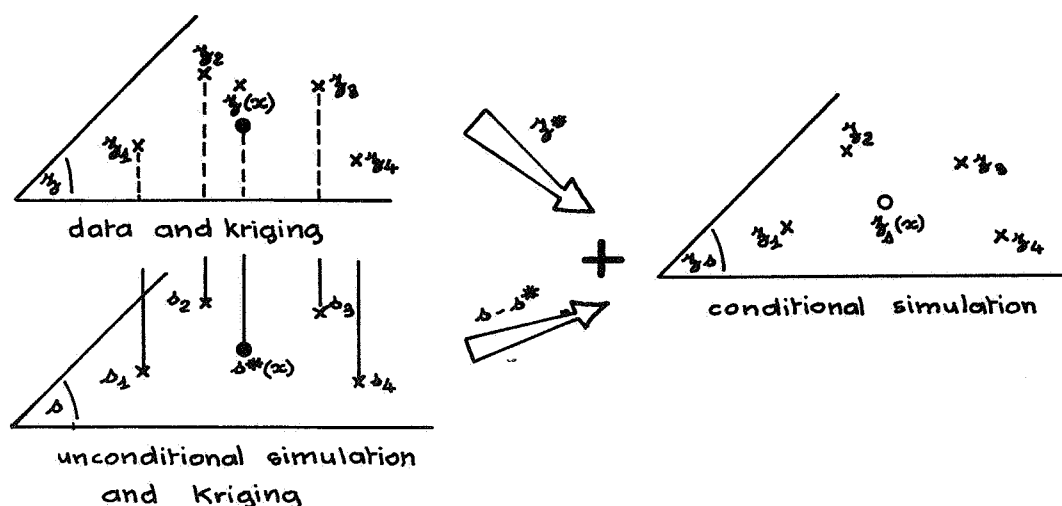


Fig. 1

The only problem left is how to obtain unconditional simulations with a given covariance or variogram. In one dimension the problem is relatively well solved. But in 2 or 3 dimensions it gets more complicated and the mere extension of 1-D methods would be very costly in terms of computer time. A very elegant and efficient method is the so-called "turning bands" method developed by G. Matheron and his group at Fontainebleau. With this technique it suffices to simulate 1-D functions, which is relatively easy, and then rotate the lines in space while adding up, for each simulated point x of space, the values taken by the projections of x on the lines.

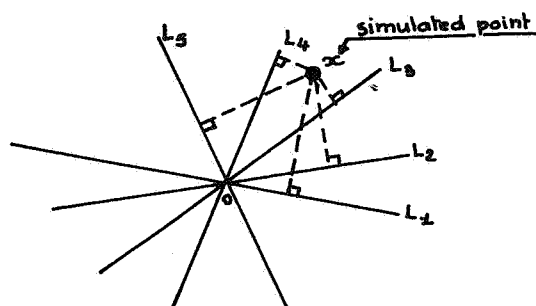


Fig. 2 : The turning bands method

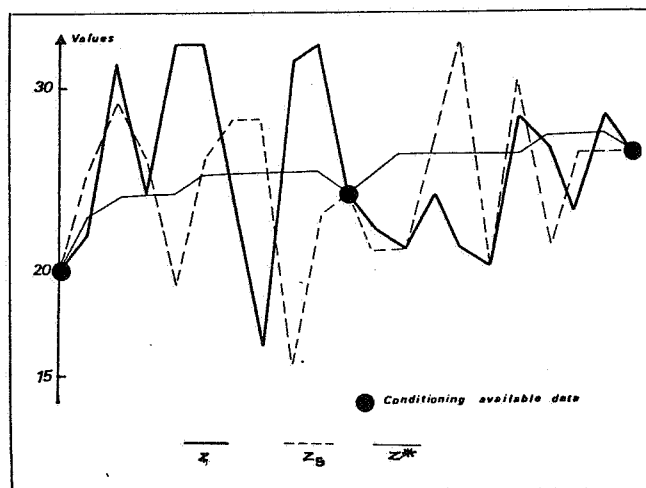


Fig. 3 : Reality - Simulation - Kriging

The relationships between reality, conditional simulation and kriging are well illustrated on the 1-D example of Fig. 3. The solid black line is the real curve ; it passes through data points marked by big dots. The thin line represents the kriging estimates : it goes through the data points but is much smoother than reality. The dashed line is a conditional simulation : it also goes through the data points but in addition has a variability similar to that of reality. We can notice that in general the kriging estimate is closer to the real value than the simulation. This point should be emphasized : conditional simulations do not purport to estimate reality, but simply to give plausible versions of what it can look like.

It is also interesting to note that if we take many conditional simulations at a point their average is simply the kriging estimate and the variance the kriging variance. This results immediately from the definition : (the estimates $z^*(x)$ being fixed)

$$z_s(x) = z^*(x) + [S(x) - S^*(x)]$$

Hence

$$E[z_s(x)] = z^*(x) + E[S(x) - S^*(x)]$$

But kriging is unbiased $E[S(x) - S^*(x)] = 0$, so

$$E[z_S(x)] = z^*(x)$$

Also

$$E[z_S(x) - z^*(x)]^2 = E[S(x) - S^*(x)]^2 = \sigma_K^2$$

The variability of a conditional simulation around its expected value (kriging) reflects precisely the uncertainty we have about this value.

3 - A CASE STUDY.

The reservoir we are dealing with is an oil reservoir which has the shape of a well marked dome, with a maximum thickness of around 140 m. and a lateral extension of the order of 3 or 4 Km. As the transition zone is particularly important the reservoir has been divided into 7 horizontal layers of 20 m. each (Fig. 4). The average porosity of each layer has been computed by kriging of the available porosity data (measured in 33 wells).

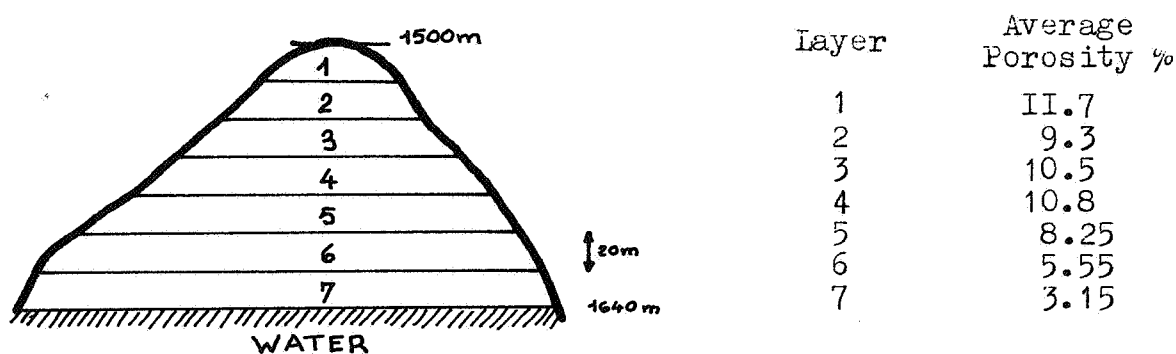


Fig. 4

To simplify matters in this case study we shall assume that these porosity values are correct and shall concentrate on the error due to the uncertainty about the reservoir boundaries. This uncertainty originates from two causes :

- (i) the reservoir top depth
- (ii) the water level depth

The reservoir top depth is known at 84 wells, the location of which is represented on Fig. 5. The structural analysis tells us that there is a quadratic drift and a variogram linear at the origin and without nugget effect. As for the water level, it is evaluated to be between 1620 m. and 1640 m. deep, i.e. within layer 7. We have then performed the following operations :

- (a) kriging of the top surface (Fig. 6) and calculation of the corresponding error standard deviation (Fig. 7).
- (b) generation of 30 different conditional simulations of the top. Six of them are shown below as they appear after cut-off by the lowest water level (1640 m.) : Fig. 8, 10, 11, 12, 13, 14. The digits represent the numbers of the layers ; the blanks between layers are just there to enhance the visualization. From these documents it is easy to see the effect of a cut-off with a higher water level. An example is shown in Fig. 9, where the water level is taken at its upper limit (1620 m.).
- (c) calculation of the averages of the 30 conditional simulations (Fig. 16) and of their standard deviations (Fig. 17). By comparing these to Fig. 6 and 7 we see that the simulations behave like predicted by theory, namely their average is the kriged map and their standard deviations coincide with those of the kriging errors.
- (d) evaluation of rock and hydrocarbon volumes by numerical integration on :
 - the kriged reservoir top
 - the 30 conditional simulations.

Three different depths of the water level have been considered : 1640 m, 1630 m, 1620 m. The results are recorded in Table I.

VOLUMES COMPUTED FROM CONDITIONAL SIMULATIONS

Water Level 1620 m		Water Level 1630 m		Water Level 1640 m	
Rock Volume 10^6 m^3	Oil Volume 10^6 m^3	Rock Volume 10^6 m^3	Oil Volume 10^6 m^3	Rock Volume 10^6 m^3	Oil Volume 10^6 m^3
325.7	26.46	391.1	28.52	463.9	30.81
332.7	27.74	395.9	29.73	467.0	31.98
337.9	27.96	402.3	29.99	473.7	32.24
331.4	27.16	395.9	29.19	467.6	31.45
324.7	26.58	393.1	28.73	468.2	31.10
350.7	28.88	417.0	30.97	489.6	33.25
314.7	25.82	377.0	27.79	446.6	29.98
354.3	29.20	423.0	31.37	498.7	33.75
317.5	25.85	382.7	27.90	454.6	30.17
325.7	26.95	388.6	28.93	459.8	31.18
317.8	26.13	382.3	28.16	450.7	30.31
316.7	26.32	376.6	28.20	445.0	30.36
352.7	29.04	419.3	31.13	490.2	33.37
332.9	27.25	401.1	29.40	475.9	31.76
320.8	26.52	387.8	28.63	466.3	31.10
290.6	23.67	351.3	25.58	419.1	27.72
333.9	27.68	400.7	29.79	473.4	32.08
316.8	25.77	379.8	27.76	447.7	29.89
346.8	28.60	416.0	30.78	492.4	33.19
335.3	27.76	398.4	29.75	466.8	31.90
316.2	25.92	386.3	28.13	465.0	30.61
332.0	27.27	397.3	29.33	467.8	31.55
312.9	25.59	375.3	27.56	443.6	29.71
305.6	25.07	368.2	27.04	436.2	29.19
330.3	27.06	399.6	29.24	476.2	31.65
341.8	28.05	411.7	30.25	488.1	32.65
318.9	26.15	383.0	28.17	453.3	30.39
323.0	26.72	387.8	28.76	461.4	31.08
391.0	32.01	470.1	34.51	555.6	37.20
391.2	32.00	468.9	34.45	554.1	37.13

m	331.4	27.24	397.6	29.32	470.6	31.62
σ	21.1	1.75	24.7	1.86	28.4	1.97

VOLUMES COMPUTED FROM KRIGED RESERVOIR TOP

332.1	27.41	396.2	29.43	465.7	31.62
-------	-------	-------	-------	-------	-------

In this case, we can note that the estimates of rock and oil volumes based on the kriged reservoir top are quite comparable to those given by the averages of the simulations. However, there is no guarantee that it should always be so.

From the figures on Table I we can build histograms and expectation curves of the oil reserves under the 3 hypotheses of water levels (Table II).

It turns out that for a fixed water level the distribution of reserves is not very dispersed : the coefficient of variation $\frac{\sigma}{m}$ is of the order of 6 % while the maximum observed variation (i.e. maximum minus minimum) represents 30% of the reserves. In other words, in this case and for a given water level, the uncertainty on the depth of the reservoir top does not have a large influence on the reserves estimate. This is due to the unusually large amount of wells. When the water level varies from 1640 m to 1620 m the rock volume decreases by 30% or so but fortunately, due to the low porosity of layer 7, the reserves only vary by about 14%. Of course the uncertainty on the water level must be combined to that of the reservoir top. To do this it suffices to cut the top by several water levels and pool the histograms thus obtained.

CONCLUSIONS.

This deliberately simplified case study has shown how we can proceed to find a probability distribution of the hydrocarbon reserves in place. Naturally these reserves are purely static ones. The evaluation of the recoverable reserves would involve the hydro-dynamic characteristics of the reservoir as well as the technology applied for production. This is a different problem and we do not touch it. The sole objective of geostatistical estimation of hydrocarbon reserves is to tell with the utmost possible accuracy how much there is down there in the field.

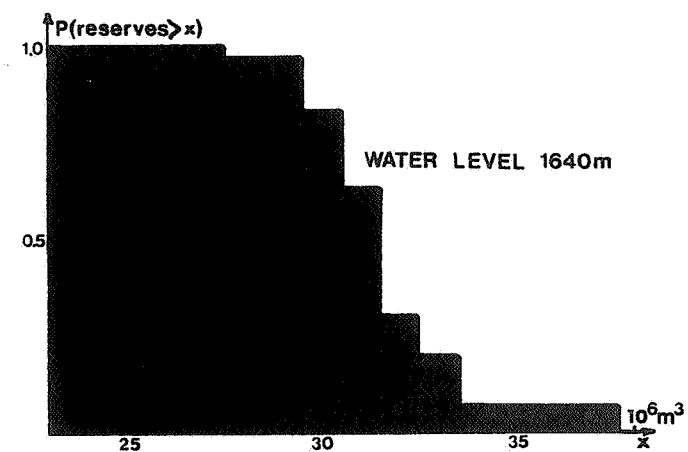
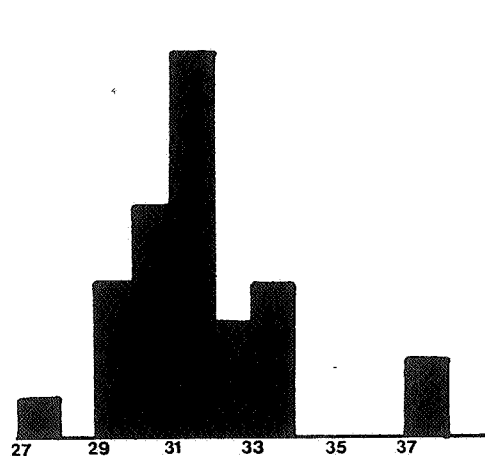
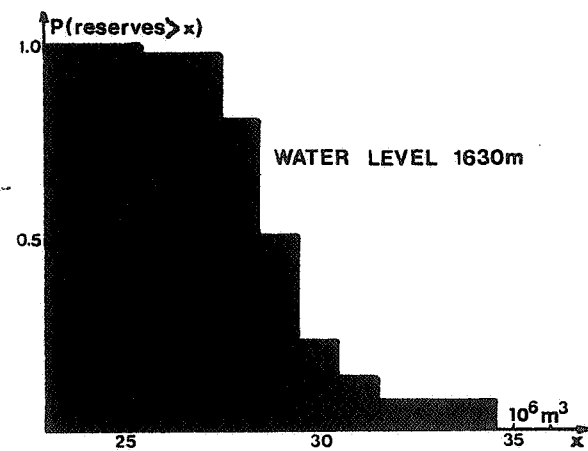
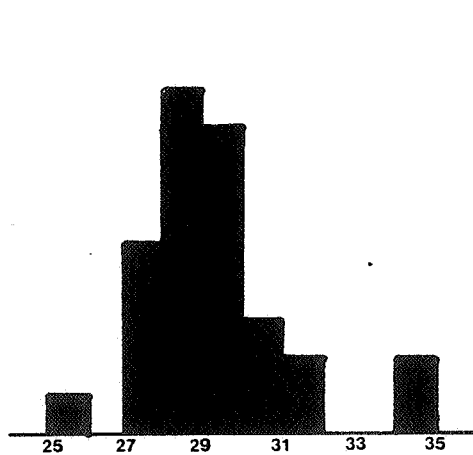
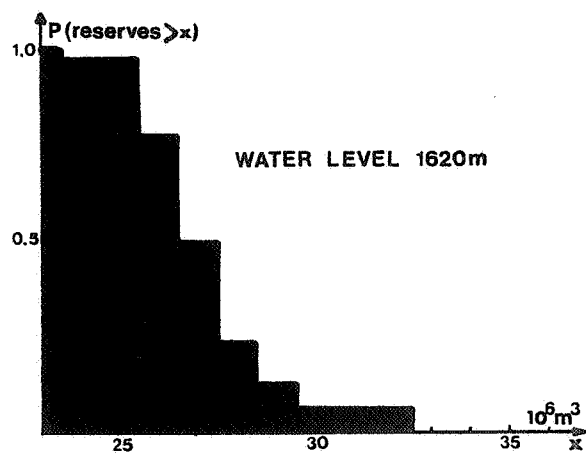
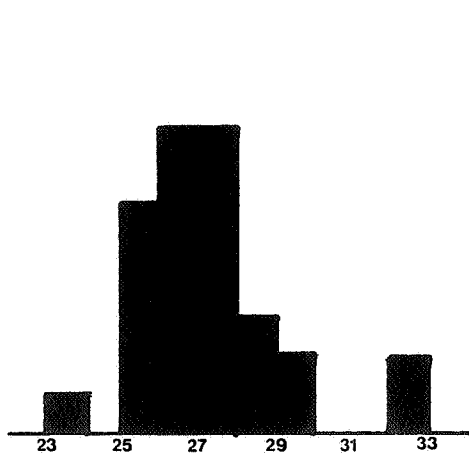
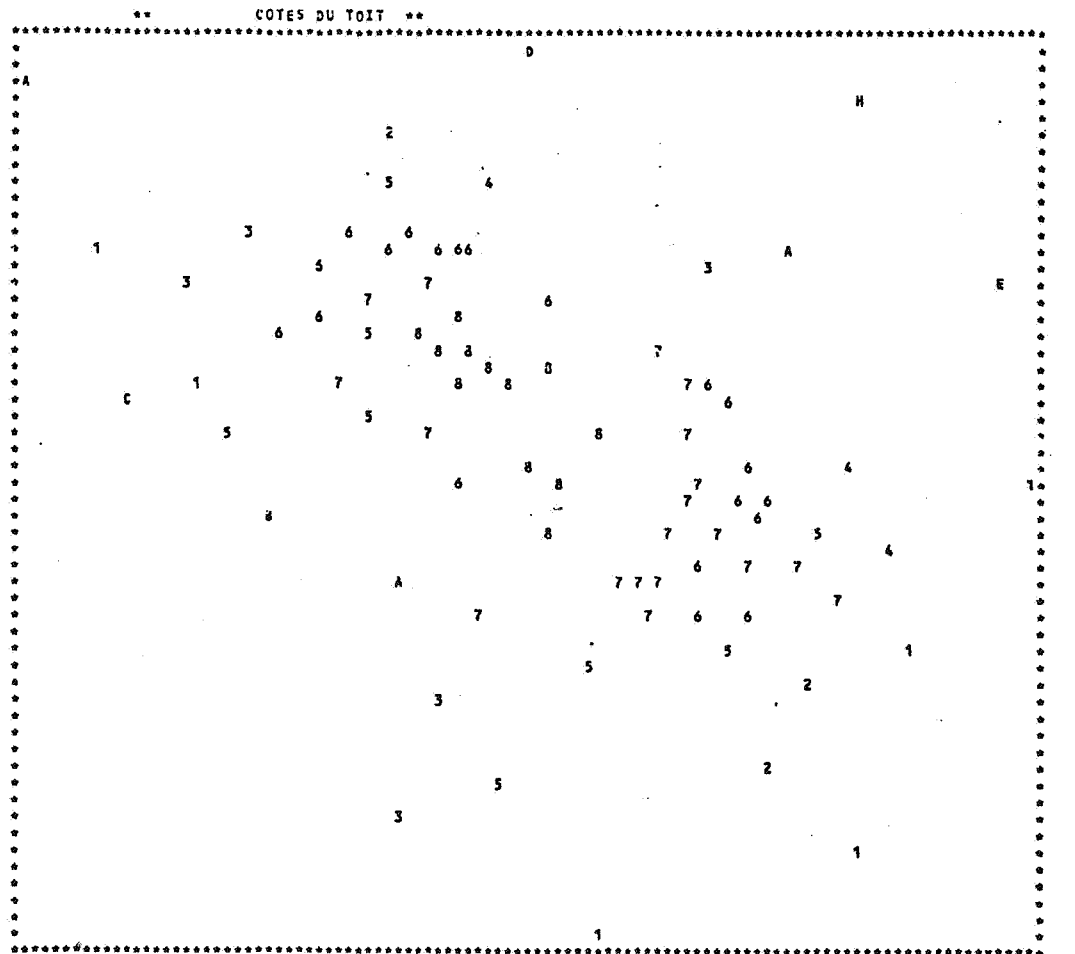


Table II

HISTOGRAM AND EXPECTATION CURVES OF OIL IN PLACE VOLUMES



ÉCOLE DES MINES
DE PARIS

101 CARACTÈRES/LIGNE
X HORIZONTAL VARIE DE GAUCHE À DROITE, ENTRE .5701E 05 ET .6179E 05
Y VERTICAL VARIE DE BAS EN HAUT, ENTRE .2555E 05 ET .2980E 05
POUR LES 84 VALEURS PRÉSENTES DANS LE DOMAINE CONSIDÉRÉ, IL Y A 54 LIGNES
ECH : 47,82 UNITÉS/CARACTÈRE
ECH : 79,49 UNITÉS/LIGNE
84 POINTS REPRÉSENTATIFS DISTINCTS

1	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1700.	ET	-1675.
2	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1675.	ET	-1650.
3	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1650.	ET	-1625.
4	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1625.	ET	-1600.
5	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1600.	ET	-1575.
6	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1575.	ET	-1550.
7	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1550.	ET	-1525.
8	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1525.	ET	-1500.
9	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1500.	ET	-1475.
+	REPRÉSENTE LES VALEURS SUPPLÉMENTAIRES À	-1475.		
A	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1725.	ET	-1700.
B	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1750.	ET	-1725.
C	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1775.	ET	-1750.
D	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1800.	ET	-1775.
E	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1825.	ET	-1800.
F	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1850.	ET	-1825.
G	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1875.	ET	-1850.
H	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1900.	ET	-1875.
I	REPRÉSENTE LES VALEURS COMPRISSES ENTRE	-1925.	ET	-1900.

ÉCOLE DES MINES
DE PARIS

Fig. 5

[illegible]

RECAPITULANT DES YSOLIGNES= 20.00 YSOLIGNE DE REFERENCE= 1640.

Fig. 6

ECA RTEMENT DES SOLIGNES 10.00 TOTAL DES PREP. 100.00

Fig. 7

SIMULATION CONDITIONNELLE NO 16 PLAN D'EAU 1620



Fig. 9



ÉCOLE DES MINES
DE PARIS



ÉCOLE DES MINES
DE PARIS

SIMULATION CONDITIONNELLE NO 9 PLAN D'EAU 1640

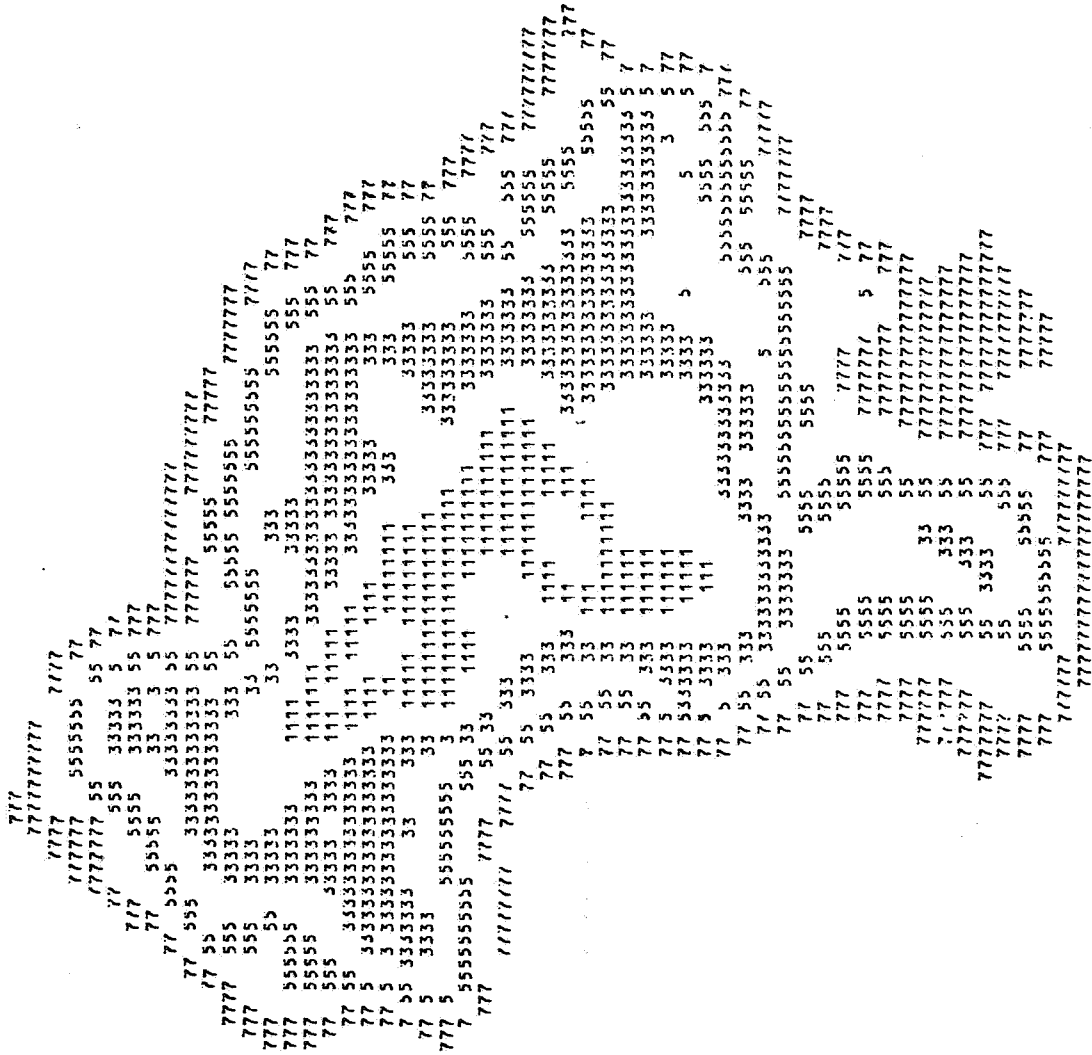


Fig. 10



ÉCOLE DES MINES
DE PARIS



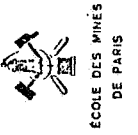
ECOLE DES MINES
DE PARIS

[illegible]

SIMULATION CONDITIONNELLE NO 27 PLAN D'EAU 1640

[illegible]

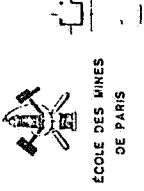
Fig. 13



SIMULATION CONDITIONNELLE NO 28 PLAN D'EAU 1640



Fig. 14



SIMULATION CONDITIONNELLE NO 29 PLAN D'EAU 1640



Fig. 15

MOYENNE DES SIMULATIONS CONDITIONNELLES

[illegible]

Fig. 16

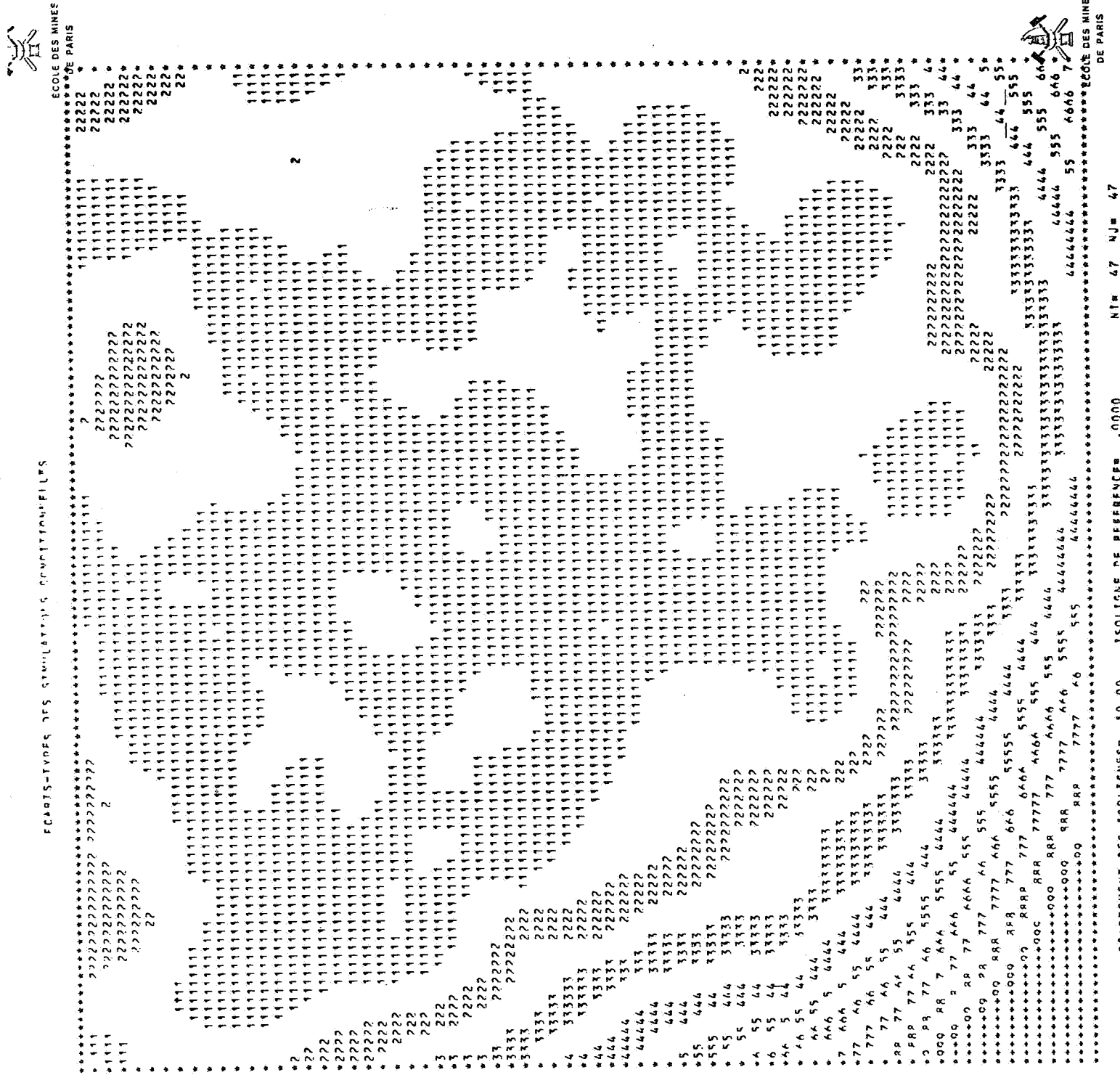


Fig. 17

REFERENCES

The following list includes only the references which have actually been used to write this course.

- *DAVID, M. (1974) : "A course on Geostatistical Ore Reserve Estimation".- Département de Génie Minéral, Ecole Polytechnique de Montréal - 304 p.
- *DELFINER, P., ETIENNE, J. and FONCK, J.M. (1972) : Application de l'analyseur de texture a l'étude morphologique des réseaux poreux en lames minces". Rev. Inst. Franç. du Pétrole, Vol. XXVII, n° 4, pp. 535-588.
- *DELFINER, P. and DELHOMME, J.P. (1973) : "Application du krigage à l'optimisation d'une campagne pluviométrique en zone aride". Symposium on the Design of Water Resources Projects with Inadequate Data, UNESCO - WMO - IAHS - Madrid, June 1973, pp. 191-210.
- *DELFINER, P. and DELHOMME, J.P. (1973) : "Optimum Interpolation by Kriging". Display and Analysis of Spatial Data.- Ed. Wiley and Sons, London, p. 96-114.
- *DELFINER, P. (1975) : "Linear Estimation of Non-Stationary Spatial Phenomena". Proceedings of NATO A.S.I., Geostat 75, Tome. Ed.: Reidel, Dordrecht, 20 p.
- *DELHOMME, J.P. (1974) : "La Cartographie d'une Grandeur Physique à partir de Données de différentes Qualités". Proceedings Congrès A.I.M., Montpellier 1974.
- *HAAS, A. and JOUSSELIN, C. (1975) : "Geostatistics in Petroleum Industry". Proceedings of NATO A.S.I., Geostat 75, Rome, Ed. : Reidel, Dordrecht, 15 p.
- *HUIJBREGTS, Ch. (1971) : "Courbes d'Isovariance en cartographie automatique". Sciences de la Terre, Tome VXi (1971), n° 3-4. Journées d'Etude C.N.R.S., p. 291-301.
- *HUIJBREGTS, C. (1975) : "Three-day Short Course on Practical Mining Geostatistics". Clausthal, 2-4 Oct. 1975 - 120 p.

- *JOURNEL, A. (1975) : "Geological Reconnaissance to Exploitation" A decade of Applied Geostatistics" - The Canadian Mining and Metallurgical Bulletin, June 1975, p. 1-10.
- *MATHERON, G. (1965) : "Les Variables Régionalisées et leur estimation". Ed. : Masson et Cie, Paris, 305 p.
- *MATHERON, G. (1967) : "Eléments pour une Théorie des Milieux Poreux. Ed. : Masson et Cie, Paris, 166 p.
- *MATHERON, G. (1971) : "The Theory of Regionalized Variables and its Applications". Les Cahiers du Centre de Morphologie Mathématique, N° 5. Ed. : Ecole Nat. Sup. des Mines Paris, 212 p.
- *MATHERON, G. (1973) : "The Intrinsic Random Functions and their Applications". Advances in Applied Probability, N° 5, Dec. 1973, p. 439-468.
- *MOOD A.M. and GRAYBILL, F.A. (1963) : "Introduction to the Theory of Statistics". McGraw Hill, 443 p.
- *SERRA, J. (1967) : "Echantillonnage et Estimation locale des Phénomènes de Transition Miniers". Thesis - University of Nancy, 670 p.

