Les Cahiers du Centre de Morphologie Mathématique
DE FONTAINEBLEAU

No 5

THE THEORY OF REGIONALIZED VARIABLES AND ITS APPLICATIONS

By

G. MATHERON

1971

Published by the École Nationale Supérieure des Mines de Paris

THE THEORY OF REGIONALIZED VARIABLES, AND ITS APPLICATIONS

Table of Contents

FOREWORD

I have written down here the lectures I gave in 1970 at the Summer School of Fontainebleau. This text is the outcome of several years of reflection on the ambiguous character of the interpretation in probabilistic terms of a unique but partly unknown natural phenomenon, and about the difficult problems raised when seeking the conditions which allow statistical inference starting from a single realization of a non-stationary random function. Thus I have laid more emphasis on methodological problems than on the mathematical aspects of the theory. From a practical point of view, the useful results have already been given in [9] (except for universal kriging) but which are now presented in a new light. The solution to the above-mentioned methodological problem is to introduce a hypothesis of quasi-stationarity, weak anough to be always physically conceivable. The combination of chapter 1 (transitive methods) and chapter 2 (intrinsic random functions theory) shows that the problem of statistical inference is solvable (and therefore the operative theory) insofar as it concerns the global estimation of a regionalized variable. Chapter 4, devoted to universal kriging, leads to an analogous conclusion concerning the problem of local estimation. I have not presented universal kriging in terms of Hilbertian space in order to avoid any mathematical difficulty and to concentrate more on methodological problems. The reader is referred to [6] for the proofs of certain results (particularly the theorems of existence and uniqueness). On the contrary I have studied from various points of view the prime concept of drift (functional drift, random drift, maximum likelihood, theory of interpolators) which emphasize its physically ambiguous character, or at least show that the drift is closely related to considerations of scale, and also show that these apparently contrary points of view lead to convergent conclusions : this will at least appear as reassuring from the methodological point of view.

Finally, a word about the exercises given at the end of each chapter, some of which were actually discussed during the Summer School in 1970. Some of them are simply application exercises meant to make the actual use of a rather abstract theory more familiar to the reader. Others are complementary to the chapter itself and in some cases produce results which have fundamental methodological consequences. In each case a solution is given, as well as the procedure to follow to find it, and the methodological conclusions to be drawn therefrom. That is enough, I think, to convince the reader of the necessity to do these exercises conscientiously.

G. MATHERON

0 - INTRODUCTION

0-1 NOTATIONS

The symbols x, y, z refer to points of the n-dimensional space (n = 1, 2 or 3) ; dx, dy, dz to elements of length (n = 1), surface (n = 2) or volume (n = 3) centred on these points, and f(x), g(y).... to functions of these points.

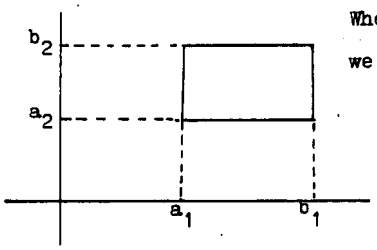
The integral extended over the whole space of a function f(x) will be denoted by :

$$\int f(x)\, dx$$

For example, when n = 3, and calling (x_1, x_2, x_3) the three coordinates of the point x, this notation produces the explicit form :

$$f(x)\, dx = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \int_{-\infty}^{+\infty} f(x_1,x_2,x_3)\, dx_3$$

In the same way $\int_A f(x)\, dx$ will denote the integral of a function f(x) over a domain A of the n-dimensional space.



When n = 2, for example, and when A is the rectangle shown opposite, we have explicitly :

$$\int_A f(x)\, dx = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} f(x_1,x_2)\, dx_2$$

These notations are both very condensed and very expressive : initially, they can be interpreted in the case of the one-dimensional space where their significance is generally clear. The extension to two or three dimensions follows easily from this.

Example : Let V be a volume and $g(h) = g(h_1,h_2,h_3)$ a function of the vector h with coordinates h_1, h_2, h_3. Let $x = (x_1,x_2,x_3)$ and $y = (y_1,y_2,y_3)$ be the origin and the end of this vector h, that is to say h = y - x. When the extremities x and y of the vector h sweep the volume V, each one on its own account, the mean value of the function g(h) will be denoted by :

$$\frac{1}{v^2} \int_v dx \int_v g(y-x)\, dy$$

In explicit notation, this can be written as :

$$\frac{1}{v^2} \iiint_v dx_1\, dx_2\, dx_3 \quad \iiint_v g(y_1-x_1\ ;\ y_2-x_2\ ;\ y_3-x_3)\, dy_1\, dy_2\, dy_3$$

The first notation is a symbolic one, showing directly the concept of mean value ; the second one is an algorithm showing the course to follow in calculating this quantity, but its significance is not so apparent at first sight as in the first case.

0-2 CONVOLUTION (Moving Average)

The convolution of two functions $f_1(x)$ and $f_2(x)$ is the function $g(x)$ defined by :

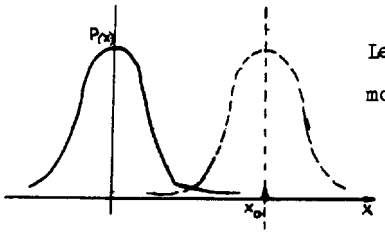
$$g(x) = \int f_1(y)\ f_2(x-y)\ dy = \int f_2(y)\ f_1(x-y)\ dy$$

This is denoted by :

$$g = f_1 * f_2$$

This operation of convolution plays a fundamental rôle in geostatistics as in probability theory and theoretical physics. It can be connected with the intuitive notion of "moving average" :

Regularization of a function f(weighted moving average of this function)



Let p(y) be a weighting function. The value at the point x_o of the moving average of the function f weighted by p is :

$$f_p(x_o) = \int p(y)\ f(x_o+y)\ dy = \int p(-y)\ f(x_o-y)\ dy$$

(the weight $p(y)\,dy$ is given to the value taken by f at $x_o + y$, and the second expression is obtained by changing y into -y).

Let p be the reflection of p at the origin (definition : $p(x) = p(-x)$). $f_p(x_o)$ is the value at x_o of $f * \check{p}$:

$$f_p = f * \check{p}$$

This moving average f_p of the function f (weighted by p) is called the regularization (of f by p).

Example 1 (taking a sample v located at the point x).

Let v be the sample located at the origin of coordinates, and k(x) its indicator ($k(x) = 1$ if $x \in v$ and $k(x) = 0$ if $x \notin v$). Let us take the weighting function $p(x) = \frac{1}{v}\,k(x)$. The corresponding moving average

$$f_v = \frac{1}{v}\ f * \check{k}$$

represents the average grade of the sample v taken at the point x. Explicitly, we have :

$$f_v(x) = \frac{1}{v} \int_v f(x+y)\,dy$$

Example 2 (Radioactivity) .

If a unit mass of a radioactive substance is placed at the origin O of the coordinates, the radioactivity at a distance d is given by $A\ e^{-\lambda d}/d$. (A is a constant, λ the absorbtion coefficient of the medium). Let f(x) be the grade of this radioactive substance at the point x, and let $d(x-x_o)$ be the distance between two points x and x_o. The mass $f(x)\,dx$ placed at x induces at x_o the radioactivity $A\ e^{-\lambda d(x_o-x)}/d(x_o-x)\ .\ f(x)\,dx$. Thus, the radioactivity observed at x is :

$$A \int f(x)\ \frac{e^{-\lambda d(x_o-x)}}{d(x_o-x)}\ dx$$

This is none other than the convolution $Af * \check{p}$ (with a weighting function $p(x) = e^{-\lambda d}/d$, which is symmetric : $p = \check{p}$).

0-3 GEOSTATISTICS AND THE THEORY OF REGIONALIZED VARIABLES.

Geostatistics are the application of the theory of the regionalized variables to the estimation of mineral deposits (with all that this implies). More generally, when a phenomenon spreads in space and exhibits a certain spatial structure, we shall say that it is regionalized. Earth sciences, among others, offer numerous examples. If $f(x)$ denotes the value at the point x of a characteristic f of this phenomenon, we shall say that $f(x)$ is a regionalized variable (abv. Re. V.). This is a neutral term, purely descriptive, and prior, in particular, to all probabilistic interpretation.

From the mathematical point of view, a Re. V. then is simply a function $f(x)$ of the point x but is, generally, a very irregular function ; ex : a grade in a mineral deposit.

It shows two contradictory (or complementary) aspects :

- a random aspect (marked irregularity and unpredictable variations from one point to another)
- a structured aspect (it must reflect in some extent the structural characteristics of the regionalized phenomenon).

The theory of regionalized variables has therefore two main purposes :

- on theoretical grounds, to express these structural properties in an adequate form.
- on practical grounds, to solve the problem of the estimation of a Re. V. from fragmentary sampling data.

These two purposes are related : for the same network of samples, the error of estimation depends on the structural characteristics ; for example, it becomes greater when the Re. V. is more irregular and more discontinuous in its spatial variations.

Field and support of a Re. V.

The field V of a Re. V. is the region where it differs from zero. A panel is a subset V' of V.

Support - Often the function $f(x)$ in itself is unknown , but instead only its mean value $f_v(x)$ in a sample v drawn at the point x. This regularization $f_v(x)$ is actually more regular than the Re. V. $f(x)$. The volume v is called the support of the Re. V. $f_v(x)$, regularization of f. Another important task of the theory of the regionalized variables will be to determine the characteristics

of f_o, knowing those of f. Ex : to predict in an orebody the characteristics of panels V' (variable $f_{V'}$), knowing those of f or f_v (samples). More generally, we will try to relate the characteristics of f to those of a regularization f_p through a given function p (see previous example about radioactivity).

0-4 TRANSITIVE METHODS AND INTRINSIC THEORY.

To achieve these objectives, two groups of methods are available :

- transitive methods : absolutely general, and in particular not necessitating any hypothesis of a probabilistic nature and, a fortiori, any hypothesis of stationarity.

- intrinsic theory : an application of the random functions theory ; probabilistic interpretations are introduced, as well as a certain hypothesis of stationarity (the intrinsic hypothesis).

From the theoretical point of view, these two groups of methods lead to equivalent results : this is important for the methodology, for it shows that the results of the intrinsic theory do not depend upon the hypothesis of stationarity (besides, it is possible to build a probabilistic theory free from this hypothesis, which still produces the principal results of geostatistics).

Here we come up against a methodological problem of prime importance, as well for the theory itself as for the critical study of the value of the results to which it leads. It is clear that the ambiguous, locally erratic character of a Re. V. calls for a probabilistic interpretation and as a matter of fact, our second group of methods makes use explicitly of the theory of random functions. But then, there arise two fundamental questions, the first of which is :

a/ What is the true epistomological meaning of the interpretation of a unique natural phenomenon (the Re. V.) as being a realization of a random function (i.e. as the result of drawing at random from an infinite population of regionalized variables considered as "possible")?

In fact, such a probabilistic interpretation is in itself a conceptualization of reality (a constitutive model), rather than an assumption capable of being proved or disproved by experiment. It is justified only insofar as it creates a better picture of reality, and allows to solve effectively practical problems which would otherwise be unsolvable. Thus it seems wise to try,

first of all, to see how far it is possible to go without appealing to this interpretation, and we will see that in fact it is possible to go quite far. Hence this first group of methods called transitive methods in which, seemingly, there appears no probabilistic concept. Here, the Re. V. is characterized by its transitive covariogram g(h), not probabilistic, which sums up its essential structural features and allows - if it is known - the complete solution of certain practical problems like estimation.

If we could effectively determine the covariogram g(h) starting from fragmentary sampling data, it would be possible to leave it at that and to avoid any probabilistic interpretation. But such is not the case, and a more precise analysis will show that it is necessary (in order, for example, to compute an estimation variance) to introduce a special kind of assumption concerning the behaviour of g(h) near the origin : assumption whose epistemological significance is exactly a disguised passage to expectations. So, as it is not really possible to avoid a probabilistic interpretation, it is better to introduce it explicitly. The second fundamental question arises from this :

b/Once this probabilistic interpretation is admitted, is statistical inference from a single realization possible ? In other words, starting from the only material at our disposal (the Re. V. itself or a fragmentary sample of it), is it really possible to reconstruct, at least in part, the probability distribution of the hypothetical random function of which our Re. V. would be a realization ?

In order to give a positive answer to this second question, hypotheses are often introduced, like stationarity and ergodicity, far stronger than is really necessary : in many applications these hypotheses are glaringly false (examples : an ore-body where assays decrease more or less regularly outwards from a rich core ; submarine topography where depth increases seawards, etc..., these are obviously non-stationary phenomena). Even more frequently these hypotheses appear to be unverifiable : therefore we will constantly try to weaken them, and to reduce them to the indispensable, permissible minimum. More precisely, question b/ above may be replaced by the following one :

b'/ Which minimal probabilistic characteristic is it necessary to know in order to solve a given practical problem (for example the computation of an estimation variance) and which minimal assumption is it necessary to introduce in order to make the estimation of this characteristic possible, starting from a single realization ?

Generally, this minimal assumption, weak enough to be physically conceivable, will be the as-

sumption that the random function is quasi-intrinsic (i.e. locally equivalent to an intrinsic R.F.) while the minimum characteristic which is necessary and possible to estimate, will be essentially associated with the behaviour near the origin of a variogram or a quasi-stationarity covariance function.

This reduction of the problem to its minimum probabilistic characteristics explains the success of the transitive methods (which, once their probabilistic content is explicitly stated, obviously do not require any hypothesis of stationarity), and also the equivalence of their results with those given by the intrinsic theory. Moreover, it emphasizes the prime importance of the behaviour near the origin of a variogram or of a transitive covariogram. Finally, from a practical point of view, it shows that we are very often correct in applying to outwardly non-stationary phenomena, computational methods which were initially thought to be justifiable only under some stationary hypothesis.

In practice, the intrinsic theory is easier to use, and will almost always be employed, except in the particular and very important case of the estimation of a surface or a volume (geometrical problem).

Bibliography : (see references at the end of this booklet)

[4] is the complete theory of the regionalized variables. In [5] there is an abstract of this theory followed by a complete treatise on applied geostatistics, including the study of economical optimisation problems (the original French text can be consulted at the library of the School of Mines of Paris). The thesis of J. Serra [7] gives a very complete general account (without any mathematical difficulties) and also a detailed study of the spherical scheme. Treatise [3] is mostly obsolete, except for the part dealing with the De Wijsian scheme : based on this same scheme, the Carlier's thesis studies the special problems of estimation of radioactive substances. Finally the problems of economical optimisation are discussed in two contributions to the French Annales des Mines [8], but the essential parts of these are to be found in [5]. The complete theory of universal kriging is given in [6].

CHAPTER 1

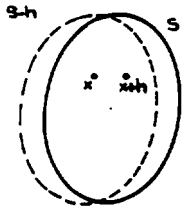
TRANSITIVE METHODS

1-1 INTRODUCTORY EXAMPLE.

Let us consider the simplest "transition" phenomenon which can be imagined : the presence or absence of a characteristic. Consider, for example, a geological formation S of limited extent : a bore hole drilled at the point x either encounters it or does not. Let k(x) be the indicator of S, i.e. the function defined by :

$$k(x) = \begin{cases} 0 & \text{if } x \notin S \\ 1 & \text{if } x \in S \end{cases}$$

We are concerned here with a unique phenomenon for which no probabilistic formulation is possible : to speak of the probability that a given point x may belong to S would not make much sense. It can be noted that in this case all interest will be concentrated on the boundary of S. Indeed, k(x) is constant inside S and also outside S, and it is only while crossing this boundary that k(x) varies, changing from 0 to 1 and from 1 to 0. Hence the names "transition phenomenon" and "transitive methods".



The area S of the formation is obviously given by :

$$S = \int k(x)\, dx$$

The value of S is of great interest from a practical point of view : most often, this is what we want to estimate from a grid of drill-holes. However, this scalar parameter tells us nothing of a structural nature. Indeed, the structure of a set can be defined as the system of relations existing between the elements or subsets of this set. Hence we will get information of a structural nature about our surface S only by using two points simultaneously.

Let then x and x+h be two points (i.e. the smallest structuring set can can be imagined). Consider then the expression k(x) k(x+h) : it is equal to 1 if both x and x+h belong to S, and

to 0 otherwise. But to say that x+h belongs to S is equivalent to saying that x belongs to the translate S_{-h} of S by the translation by a vector -h. Then

$$k(x)\ k(x+h) = \begin{cases} 1 & \text{if } x \in S \cap S_{-h} \\ 0 & \text{if } x \notin S \cap S_{-h} \end{cases}$$

By integrating this expression with respect to x, we get a function of h

$$K(h) = \int k(x)\ k(x+h)\ dx = \text{Meas}\ (S \cap S_{-h})$$

which represents the measure (the area) of the intersection of S with its translation S_{-h} by -h. This function is symmetric since the two intersections $S \cap S_{-h}$ and $S \cap S_h$ are deduced from one another by translation. The function K(h) is the <u>geometrical covariogram</u> associated with S. It gives a certain image of the shape of the set S :

<u>Properties of the geometrical covariogram K(h)</u>.

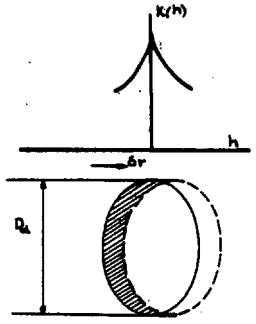
a/ <u>Symmetry</u> : $K(h) = K(-h)$; inequalities $0 \leq K(h) \leq K(0)$

Relations : $S = K(0)$

$$S^2 = \int K(h)\ dh$$

b/ <u>Ranges</u> : The range $a(\alpha)$ in direction α is the distance beyond which K(h) becomes equal to zero in this direction. It is thus the largest dimension of S in this direction.

c/ <u>Slope at the origin</u> (derivatives on left and right sides) :



If the modulus δr of h is small, we have

$$K(h) = K(0) - \delta r\ D_\alpha$$

$\delta r\ D_\alpha$ is one half of the small surface swept by the vector δr, the origin of which describes the boundary of S. D_α is the <u>diametral variation</u> of S in the direction α (if S is <u>convex</u>, this is the appa-

rent diameter in this direction).

Although a derivative $K'_\alpha(0) = - D_\alpha$ does exist in each direction, it should be noted that the function K(h), in itself, is not differentiable at h = 0. Exercise 6 (at the end of this chapter) shows how Minkowski's formula relates the derivatives $K'_\alpha(0)$ in different directions to the parameter of S (or to its surface in the three-dimensional case). These are very useful relationships in mathematical morphology.

1-2 THE TRANSITIVE COVARIOGRAM.

Let f(x) be a Re. V. which vanishes identically outside a bounded field V. The transitive covariogram of this Re. V. is the function g(h) defined by

$$g(h) = \int f(x)\, f(x+h)\, dx \qquad (1\text{-}1)$$

Properties of the transitive covariogram g(h).

a/ Symmetry : $g(h) = g(-h)$ inequality $|g(h) \leq g(0) = \int [f(x)]^2\, dx$

Let $Q = f(x)\, dx$ be the quantity of metal. Then

$$Q^2 = \int g(h)\, dh \qquad (1\text{-}2)$$

(To demonstrate this, first substitute for g the expression (1-1) and integrate first with respect to h).

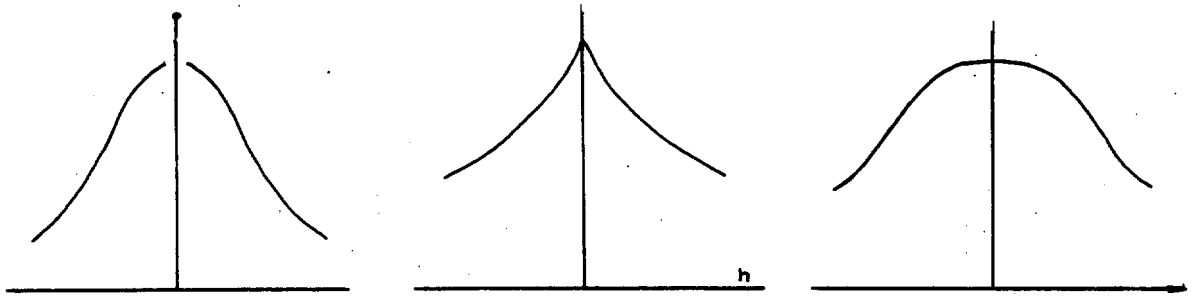
b/ Range : a(α) is defined by the condition g(h) = 0 when $|h| > a(\alpha)$ for the vector h in direction α : it is a property of the field of the Re. V.

c/ Behaviour of g(h) near the origin : the regularity of g(h) near the origin reflects the properties of continuity of the Re. V. in its variations in space. This follows from :

$$g(0) - g(h) = \frac{1}{2} \int [f(x+h) - f(x)]^2\, dx$$

If f(x) is piecewise continuous, g(h) has a linear behaviour near the origin. If f(x) is differen-

tiable, g(h) has a parabolic behaviour.



Nugget effect | Piecewise continuous Re. V. | Differentiable Re. V.

It may happen that g(h) is not continuous at h = 0 : in other words, g(0) is greater than the limit of g(h) when h tends to 0. This is called the nugget effect. Rather than representing an actual discontinuity, it is more often a reflection of a very rapid transition zone which appears on the experimental covariogram as a discontinuity. In the probabilistic part of the theory, we shall deal in greater detail with this nugget effect and with its interpretation as a scale phenomenon (see also exercise 7 of this chapter, and for the interpretation as a Dirac measure, see exercises 16 to 20).

Isotropic case - In practical applications we try (by means, for example, of a linear transformation) to make g(h) depend only on the modulus

$$r = |h| = \sqrt{h_1^2 + h_2^2 + \ldots + h_n^2}$$

of the vector h so that it can be expressed in the form of a function g(r) of the single parameter r. As functions of the coordinates h_1, h_2, ..., h_n, only the even powers r^{2k} of r are indefinitely differentiable at h = 0. The odd powers r^{2k+1}, the powers r^{λ} (λ real, non integer) and also logarithmic terms such as $r^{2n} \log r$ are irregular at h = 0. So in the isotropic case we can distinguish two parts in the limited expansion of g(r) near the origin : a regular part containing only terms of even integer degree and an irregular part (terms in r^{λ}, $\lambda \in \mathbb{R}$ different from an even integer, or else, terms in $r^{2n} \log r$)

$$g(r) = g(0) + a_2 r^2 + \ldots\ldots + \sum_{\lambda} r^{\lambda} + \sum_{k} r^{2k} \log r$$

regular part | irregular part

The term of lowest degree in the irregular part characterizes very precisely the irregularity of the Re. V. in its variations in space. Example : the geometric Re. V. associated with a surface S has a covariogram $K(h) = S - D|h|$. The term of lowest degree of the irregular part is of degree 1 in this case.

d/ The covariogram $g(h)$ is a positive-definite function. A function $g(h)$ is said to be positive-definite if, for any integer $k > 0$, any set $x_1, x_2, \ldots, x_k$ of points of the n-dimentional space and any system $\lambda_1, \ldots, \lambda_k$ of real numbers, we have :

$$\sum_{i,j} \lambda_i \lambda_j g(x_i - x_j) \geq 0 \tag{1-3}$$

These positive-definite functions play a great part in physics (where they have in general an energetical meaning)and in probability theory (characteristic functions of probability distributions are of this type). We will see in Chapter 2 that covariance functions of R.F are also of this type.

Let us show that a transitive covariogram is a positive-definite function. From (1-1) we can write :

$$\sum_{i,j} \lambda_i \lambda_j g(x_i - x_j) = \sum_{i,j} \lambda_i \lambda_j \int f(y)\, f(y+x_i-x_j)\, dy =$$

$$= \sum_{i,j} \lambda_i \lambda_j \int f(y+x_j)\, f(y+x_i)\, dy = \int [\sum_i \lambda_i f(y+x_i)]^2\, dy \geq 0$$

Thus, when choosing a model for a transitive covariogram, one must restrict oneself to positive-definite functions (we will see in paragraph 1-4 that this condition expresses the fact that the estimation variances are necessarily positive).

According to Bochner's classical theorem, a continuous function is positive definite if and only if it is the Fourier transform of a summable positive measure. This gives a second proof of this important property of the transitive covariogram. Let Φ be the Fourier transform of the Re. V. $f(x)$ and G that of its covariogram $g = f * \check{f}$. The Fourier transform of a convolution is the ordinary product of the Fourier transforms of the two factors. As f $\left(f(x) = f(-x)\right)$ has as a Fourier transform the conjugate $\overline{\Phi}$ of Φ, we get here

$$G = |\, \Phi \,|^2 \tag{1-4}$$

whence $G \geq 0$.

Remark - From the mathematical point of view, the covariogram g and its Fourier transform are perfectly equivalent to each other. As relationship (1-4) is very simple, it is sometimes more convenient (in some theoretical demonstrations) to use G instead of g itself. But from the practical point of view and in the applications, the covariogram g is in general a far better tool than G. Indeed the behaviour of g(h) near h = 0 is almost always a decisive criterion : in the Fourier transform, the image of properties near h = 0 are properties of G near infinity. In general it is not very easy to appreciate the behaviour of an experimental curve at infinity. Therefore, harmonic analysis is not a very useful tool for the practical applications of the Re. V. theory. The same remark will apply to the probabilistic version of the theory where the covariance or the variogram will be used directly instead of their Fourier transforms.

1-3 REGULARIZATION AND GRADING.

1-3-1 Regularization of a Re. V.

Let f(x) be a Re. V, p(x) a weight function, $f_p = f * \check{p}$ the regularization of f by p. Already, relation (1-1) can be written as :

$$g = f * \check{f}$$

which shows that the transitive covariogram is the auto-regularization of the Re. V. The covariogram $g_p = f * \check{p} * \check{f} * p = f * \check{f} * p * \check{p} = f_p * \check{f}_p$ can also be put in the form :

$$g_p = g * P$$

with $P = p * \check{p}$: we get the regularized covariogram by regularizing g by the transitive covariogram P of the weight function p : it is a more regular function than g in the same way as f_p is more regular than the initial Re. V. f .

1-3-2 Grading.

The rest of this paragraph will be devoted to what is called grading, an operation which plays a great part in the Re. V. theory (for example, it enables one to reduce the computation of an estimation variance in the n-dimensional space to a succession of much simpler operations executed in the one-dimensional space). In mining terms, grading is simply the abstract transposition of placing a bore-hole at the point (x_1,x_2) of the topographical surface.

If $f_3(x) = f_3(x_1,x_2,x_3)$ is a Re. V. in the three-dimensional space, it is said that the Re. V.

$$f_2(x_1,x_2) = \int f_3(x_1,x_2,x_3)\, dx_3$$

defined in the two-dimensional space is deduced from f_3 by grading (grading of order 1, parallel to the x_3 axis). For example, if $f_3(x)$ is the punctual assay at x, then $f_3(x_1,x_2)$ is the accumulation (quantity of metal per square meter) of the bore-hole located at the point (x_1,x_2).

Gradings of higher order can be defined without difficulty. For example, grading of order 2 (parallel to the x_1,x_2 plane) leads to the Re. V.

$$f_1(x_3) = \iint f_3(x_1,x_2,x_3)\, dx_1\, dx_2$$

defined in the one-dimensional space (the x_3 axis), and which represents, in mining terms, the quantity of metal per meter of dip at the level x_3. The thing to note is that the covariogram $g(x_1,x_2)$ of f_2 deduced from f_3 by grading, is deduced from the covariogram $g_3(h_1,h_2,h_3)$ of f_3 by this same grading procedure. In short : grading has the same effect at the same time on the Re. V. and on its associated covariogram.

This result can be demonstrated with the help of relation (1-4). Let $f_n(x_1,\ldots,x_n)$ be a Re. V. in $\mathbb{R}^n$, $\Phi_n(u_1,..,u_n)$ being its Fourier transform. The Fourier transform of $f_{n-1}(x_1,\ldots,x_{n-1})$ deduced from f_n by grading of order 1 parallel to the x_n axis is :

$$\Phi_{n-1}(u_1,\ldots,u_{n-1}) = \Phi_n(u_1,\ldots,u_{n-1},0) \tag{1-5}$$

In other words, in accordance with a well-known process found in probability theory when passing to a marginal distribution, Φ_{n-1} is obtained by putting $u_n = 0$ in the expression of Φ_n. Calling

g_n and g_{n-1} the covariograms of f_n and f_{n-1}, G_n and G_{n-1} their Fourier transforms taken in their respective spaces $\mathbb{R}^n$ and $\mathbb{R}^{n-1}$, we get, from (1-4) :

$$G_n = |\Phi_n|^2 \quad , \quad G_{n-1} = |\Phi_{n-1}|^2$$

From (1-5) follows :

$$G_{n-1}(u_1,\ldots,u_{n-1}) = |\Phi_n(u_1,\ldots,u_{n-1},0|^2 = G_n(u_1,\ldots,u_{n-1},0)$$

So G_{n-1} is obtained by putting $u_n = 0$ in the expression of G_n. According to the reciprocity of the Fourier transform, and to the expression (1-5), it follows therefore that g_{n-1} is deduced from g_n by grading.

1-3-3 Grading in the isotropic case.

Examine now the isotropic case, i.e. when the Re. V. $f_n(x)$, $x \in \mathbb{R}^n$ has a covariogram $g_n(r)$, which is a function of the vector radius $r = |h|$ only. It is no longer necessary to specify the direction in which grading will be performed. The gradings of order 1, 2,... lead to the Re. V. f_{n-1}, f_{n-2},..., with covariograms $g_{n-1}(r)$, $g_{n-2}(r)$ which are functions of the vector radius only (in their respective spaces with n-1, n-2,... dimensions). From the preceding remarks, these covariograms are deduced from g_n by grading and we get without any difficulty :

$$\begin{cases} g_{n-1}(r) = 2\int_0^\infty g_n(\sqrt{h^2+r^2})\, dh \\ \\ g_{n-2}(r) = 2\pi\int_0^\infty g_n(\sqrt{h^2+r^2})\, hdh \end{cases}$$

Putting $u^2 = h^2 + r^2$, we get :

$$\begin{cases} g_{n-1}(r) = 2\int_r^\infty g_n(u)\, \dfrac{udu}{\sqrt{u^2-r^2}} \\ \\ g_{n-2}(r) = 2\pi\int_r^\infty g_n(u)\, u\, du \end{cases} \qquad (1\text{-}6)$$

Thus, there is a remarkable difference between gradings of order 1 and 2. A grading of order 2 (and more generally, of even order) is an elementary operation. It can be easily inverted by differentiation :

$$g_n(r) = \frac{1}{2\pi r} \; g'_{n-2}(r) \tag{1-7}$$

On the contrary, a grading of order 1 (and more generally, of odd order) is a more difficult operation because of the root under the integration symbol, and thus cannot be easily inverted.

But it can be shown that gradings, and their inverse, or degradings, considered as operators acting upon isotropic functions of the $g_n(r)$ type, form a group (cf[4]). In order to get an expression for the degrading of order 1 ($g_{n-1} \rightarrow g_n$), it is possible, first, to carry out a degrading of order 2 which leads from (1-7) to :

$$g_{n+1}(r) = \frac{1}{2\pi r} \; g'_{n-1}(r)$$

next, to carry out a grading of order 1 which gives, from (1-6) :

$$g_n(r) = 2\int_r^\infty g_{n+1}(u) \, \frac{u \, du}{\sqrt{u^2-r^2}} = -\frac{1}{\pi}\int_r^\infty g'_{n-1}(u) \, \frac{du}{\sqrt{u^2-r^2}}$$

In other words, we have established in this way the converse expressions of the grading and degrading of order 1 :

$$\begin{cases} g_{n-1}(r) = 2\displaystyle\int_r^\infty g_n(u) \, \frac{u \, du}{\sqrt{u^2-r^2}} \\ \\ g_n(r) = -\dfrac{1}{\pi}\displaystyle\int_r^\infty g'_{n-1}(u) \, \dfrac{du}{\sqrt{u^2-r^2}} \end{cases} \tag{1-8}$$

An interesting application of these converse formulae will be found in Exercise 11 bis : it shows how the size distribution of a population of spheres in the three-dimensional space can be reconstituted from the size distributions of circles or intersections (chords) induced on planes or lines.

Now considering the problem from the Fourier transform point of view, let $G = \mathfrak{F}_n \, g_n$ be the

trandform of the covariogram g_n (taken in a n-dimensional space). As the function g_n is isotropic, its transform depends only on the vector radius

$$p = \sqrt{u_1^2 + \ldots + u_n^2}$$

in the n-dimensional space where the transforms $\mathcal{J}_n$ are defined. Thus G_n is itself an isotropic function $G(\rho)$. From relationship (1-5), we get $\mathcal{J}_{n-1}\, g_{n-1}$ by putting $u_n = 0$ in the expression for $\mathcal{J}_n\, g_n$. Now, putting $u_n = 0$ leads to the expression $\sqrt{u_1^2 + \ldots + u_{n-1}^2}$, i.e. the vector radius in the n-1 dimensional space, and which will still be designated by ρ. Accordingly, as a function $G(\rho)$ of the vector radius ρ in the Fourier space, the transform G_n of the isotropic covariogram g_n is unchanged by grading (and by degrading as well)

$$G(\rho) = \mathcal{J}_n\, g_n = \mathcal{J}_{n-1}\, g_{n-1} = \mathcal{J}_{n+1}\, g_{n+1} = \ldots . \tag{1-9}$$

This is often a convenient analytical method of expressing a grading or degrading operation.

Correspondence principle.

Let us examine now the behaviour during this grading operation of what is the most interesting part of an isotropic covariogram ; the irregular part of its limited expansion near the origin (paragraph 1-4). The essential result (the proof of which can be found in [4]) is : the grading operation can be extended term by term on the irregular part of $g_n(\underline{r})$, by the correspondence rule :

$$r^\lambda \rightarrow A_\lambda\, r^{\lambda+1} \qquad (\lambda \text{ a non integer}) \tag{1-10}$$

This correspondence principle gives the limited expansion of $g_{n-1}(r)$, except for an even integer series. There is no analogous rule for the regular part (but this is of no importance for we will see that estimation variances only depend on the irregular part). The power of the irregular term of lowest degree has thus increased of 1 and this emphasizes the regularizing effect of grading.

Concerning the terms of odd integer degree and the logarithmic terms of the irregular part, the term by term correspondence principle gives :

$$\begin{cases} r^{2k} \log r \rightarrow A_{2k}\, r^{2k+1} \\ r^{2k+1} \rightarrow A_{2k+1}\, r^{2k+2} \log r \end{cases} \tag{1-11}$$

Thus by successive gradings a singular sequence in which odd terms and logarithmic terms alternate is obtained. For example

$$\log r \to \pi r$$

$$r \to - r^2 \log r$$

Computation of the coefficients A_λ - For an exact demonstration, the reader is referred to [4]. Here only a symbolic justification is given, the principle of which will also be used for the computation of estimation variances. As a distribution (and not as a function, and this is why this justification is only symbolic) r^λ has, in the n-dimensional space, for a Fourier transform :

$$\mathfrak{F}_n\, r^\lambda = \frac{1}{\pi^{\lambda+\frac{n}{2}}} \quad \frac{\Gamma(\frac{\lambda+n}{2})}{\Gamma(-\frac{\lambda}{2})} \quad \frac{1}{\rho^{\lambda+n}}$$

(λ non integer). According to (1-9), the grading of order 1 carried out on r^λ leads to $\mathfrak{F}_{n-1}\, \mathfrak{F}_n\, r^\lambda$. Computing $\mathfrak{F}_{n-1}\, \rho^{-\lambda-n}$ from the formula above, we get :

$$\mathfrak{F}_{n-1}\, \rho^{-\lambda-n} = \pi^{\lambda+\frac{n+1}{2}} \quad \frac{\Gamma(-\frac{\lambda+1}{2})}{\Gamma(\frac{\lambda+n}{2})}\, r^{\lambda+1}$$

from which the coefficient A_λ appearing in (1-10) may be deduced with the help of the functions Γ (Euler functions).

(1-12)
$$A_\lambda = \pi \frac{\Gamma(-\frac{1+\lambda}{2})}{\Gamma(-\frac{\lambda}{2})} = \pi \operatorname{tg} \lambda \frac{\pi}{2} \frac{\Gamma(1+\frac{\lambda}{2})}{\Gamma(1+\frac{1+\lambda}{2})}$$

Again this justification is only symbolic. In particular, it does not show that the correspondence rule only gives the result of the grading operation except for an even integer series. But this result (1-12) is nevertheless correct. In the singular sequence (1-11) we get (by passages to the limit which we do not need to stress here)

(1-13)
$$\begin{cases} A_{2k} = \pi \dfrac{2^k\, k!}{1.3\ldots(2k+1)} = \pi \dfrac{2^{2k}(k!)^2}{(2k+1)^2} \\ \\ A_{2k+1} = - \dfrac{2^{-2k}\,(2k+1)!}{k!\,(k+1)!} \end{cases}$$

1-4 ESTIMATION OF A RE. V.

In this fundamental paragraph, the estimation variance will be defined in the case of a regular grid and two equivalent expressions (rigorous) which are valid in the n-dimensional space will be given. In paragraph (1-4-2) the case of a stratified random network will be analysed. Then we will look for approximation methods, first in the 1-dimensional space (1-4-3), next in the two or three-dimensional spaces (1-4-4), and these results will be applied to the geometrical problem (1-4-5). Finally, the critical study of the conditions according to which these results might be put to use will lead us to prepare the passage to the probabilistic version of the theory (1-4-6).

1-4-1 Exact expression of the estimation variance (regular grid)

For brevity, the argument is developped for the one-dimensional space only, but the results may be generalized without difficulty to the n-dimensional space. Let f(x) be a Re. V. The quantity of metal

$$Q = \int_{-\infty}^{+\infty} f(x)\, dx$$

is estimated by means of a regular sampling grid of spacing a. In other words, if x_0 is the location of (any) one of the samples, we know the numerical values $f(x_0+pa)$ of the Re. V. f(x) for p a positive or negative integer. In fact, only a finite number of these values are different from 0 for f(x) has a bounded field ; in what follows, the problem of the convergence of series will not arise for only a finite number of terms is involved. In practice, it is sufficient for the grid to extend slightly outside the field.

We take as an estimator of the quantity of metal Q :

$$Q^*(x_0) = a \sum_{p=-\infty}^{+\infty} f(x_0+pa)$$

This is a periodic function (of period a) of x_0, since it obviously does not matter whether we take as origin of the grid one or another of the samples. The estimation error is $[Q - Q^*(x_0)]$: it is also a periodic function of x_0, and its possible amplitude will be characterized by an estimation variance denoted by $\sigma^2(a)$.

In accordance with our methodological line of approach (paragraph 0-4), no probabilistic interpretation of the Re. V. itself will be introduced. It represents a unique physical reality, perfectly determined (even if it is not perfectly known). But as we know nothing about it to begin with, we will set our sampling grid "anywhere" or "at random" with regard to this unknown phenomenon. It is easy enough to give the expression "anywhere" a more precise epistemological meaning : we assume that everything follows from having the origin x_o of the grid placed at random on a segment of length a in accordance with a uniform probability distribution. This assumption simply means that initially, we do not know the exact location of a deterministic phenomenon, and does not imply that we are, as yet, very seriously engaged epistemologically.

Following the random setting of the origin x_o within the segment (0,a), the estimator $Q^*(x_o)$ becomes a random variable. Let us calculate its expectation and variance : its variance will be, by definition, the estimation variance $\sigma^2(a)$. We have for the expectation :

$$E(Q^*) = \int_o^a \frac{dx_o}{a} \; a \sum_p f(x_o+pa) = \int_{-\infty}^{+\infty} f(x)\, dx = Q$$

Our estimator is thus without bias. The estimation variance is then

$$\sigma^2(a) = A(Q^{*2}) - Q^2 = \frac{1}{a} \int_o^a [Q^*(x_o)]^2 \, dx_o - Q^2$$

Let us compute the integral of $[Q^*(x_o)]^2$. We have

$$[Q^*(x_o)] = a^2 \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} f(x_o+pa)\, f(x_o+qa) = a^2 \sum_k \sum_p f(x_o+pa)\, f(x_o+pa+ka)$$

By integration from 0 to a we get :

$$\int_o^a [Q^*(x_o)]^2 \, dx_o = a^2 \sum_k \int_o^a \sum_p f(x_o+pa)\, f(x_o+pa+ka)\, dx_o =$$

$$= a^2 \sum_k \int_{-\infty}^{+\infty} f(x_o)\, f(x_o+ka)\, dx_o = a^2 \sum_k g(ka)$$

Taking into account the relationship (1-2), we get the formula :

(1-14) $$\sigma^2(a) = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h)\, dh$$

For the n-dimensional space we get a quite analogous result. For example, where n = 3, and with a parallelepipedic grid $a_1\ a_2\ a_3$, we get :

$$\sigma^2(a_1,a_2,a_3) = (a_1\ a_2\ a_3) \sum_{k_1} \sum_{k_2} \sum_{k_3} g(k_1a_1, k_2a_2, k_3a_3) - \int g(h)\, dh$$

REMARK - The estimation variance (1-14) appears as the difference between an approximate value and the exact value of the integral $\int g(h)\, dh$. Hence it becomes smaller as :

- the grid spacing decreases,

- the function g, and therefore the Re. V. itself, becomes more regular. If the grid spacing a is small compared to the range, formula (1-14) contains a great many terms. So we shall try to find approximation formulae (see paragraphs 1-4-3 and 1-4-4). We can develop a second equivalent expression for the estimation variance by using the Fourier transform of the covariogram g. The estimator $Q^*(x_o)$, as a periodic function, has the following expansion as a Fourier series :

$$\begin{cases} Q^*(x_o) = \sum_{p=-\infty}^{+\infty} C_p\, e^{2i\pi p \frac{x_o}{a}} \\ C_p = \frac{1}{a} \int_o^a Q^*(x_o)\, e^{-2i\pi p \frac{x}{a}}\, dx \end{cases}$$

Replacing $Q^*(x_o)$ by its explicit expression, we notice that the Fourier coefficient C_p is :

$$C_p = \sum_k \int_o^a f(x+ka)\, e^{-2i\pi p \frac{x}{a}}\, dx = \int_{-\infty}^{+\infty} f(y)\, e^{-2i\pi p \frac{y}{a}}\, dy$$

Consequently, by writing the Fourier transform of f(x) as $\Phi(u)$, we get :

$$C_p = \Phi\left(\frac{p}{a}\right)$$

In particular, $C_o = \Phi(o) = Q$. From the orthogonality properties of the trigonometric functions, it follows that :

$$\sigma^2(a) = \sum_p |C_p|^2 - Q^2$$

and, taking (1-4) into account :

(1-15) $$\sigma^2(a) = \sum_p G\left(\frac{p}{a}\right) - G(o)$$

This formula, equivalent to (1-14), shows that $\sigma^2(a)$ is positive as soon as G is a positive function (thus as soon as g is positive definite). It can be extended to the n-dimensional case. For $\mathbb{R}^2$, for example, with the rectangular grid (a_1, a_2), we get :

(1-16) $$\sigma^2(a_1, a_2) = \sum_{p,q} G\left(\frac{p}{a_1}, \frac{q}{a_2}\right) - G(o,o)$$

NOTE : Thus, we have above all the important following conclusion : the estimation variance $\sigma^2(a)$ only depends on the transitive covariogram $g(h)$ and on the grid spacing a. Formula (1-14) also shows that it depends linearly on g. Thus if we knew the real covariogram of a Re. V., we would be able to solve the problem of estimation in an entirely satisfactory way, and without probabilizing the true phenomenon. To prepare the reader for the change-over to the point of view of paragraph (1-4-6), let us show that in fact there is something illusory in the above arguments.

If the only available data are the $f(x_o+pa)$, for a given x_o, the true covariogram g(h) will not be known. Only the numerical values of the g(ka) can be estimated by means of the estimators :

$$g^*(ka) = a \sum_p f(x_o+pa)\ f(x_o+pa+ka)$$

In the same way, we can only use the square $[Q^*(x_o)]^2$ of the estimator $Q^*(x_o)$ itself to estimate $Q^2 = \int g(h)\, dh$. But if we substitute these estimators in formula (1-14), we get identically 0, as it can easily be shown by :

$$a \sum_k g^*(ka) - [Q^*(x_o)]^2 = 0$$

This result is easy to understand : by replacing the true values by their estimates in (1-14), we have implicitly substituted for the true Re. V. f(x), continuously defined on the line, its restric-

tion to the discrete set of the sampling points x_o+pa themselves. Thus the result obtained simply means that $Q^*(x_o)$ is an excellent estimator of $Q^*(x_o)$ itself, as could be suspected. In other words, this means that it is illusory to hope, by means of purely empirical procedures, to get both an estimation of Q and the accuracy of this estimation from the same experimental material (the $f(x_o+pa)$.

To get out of this methodological deadlock, it is necessary to turn to a theoretical model. Thus, an appropriate mathematical expression $g(h;\lambda,\mu..)$ will be chosen, and the one or more parameters λ, μ,... will be chosen so as to fit best with the experimental $g^*(ka)$ (it is useless, and illusory, to try to fit the theoretical curve exactly to the experimental points) : formula (1-14) will be applied to this theoretical $g(h)$. The result so obtained will have exactly as much value as the choice of the theoretical model of covariogram.

But this choice is by no means an arbitrary one. Firstly, the function $g(h;\lambda,\mu)$ must be positive definite (paragraph 1-2). Next, there is an extreme importance to be attached to the analytical behaviour of the function near the origin, i.e. its irregular part : the order of the irregular term of lowest degree, and eventually the existence of a nugget effect, have the significance of physical laws and are thus objective characteristics of the actual phenomenon. With the help of experience, especially from those occasions where particularly close sampling grids were available, certain types of phenomena can be quickly recognized as having a typical analytical covariogram behaviour. As it will be seen later (paragraph 1-4-3) that the estimation variance depends mainly on the behaviour of the covariogram, the problem is reduced on the whole to determining the coefficient of the irregular term of lowest degree : in general this can be done experimentally and satisfactorily.

1-4-2 Case of a stratified random grid.

Let $f(x)$ be a Re. V. in $\mathbb{R}^n$, $g(h)$ its covariogram, and V the parallelepiped of sides $a_1, a_2, \ldots, a_n$ centred at the origin. Let h_i be the vectors with components $p_1a_1, \ldots, p_na_n$ which are integer multiples of the sides of V, so that the translates V_{h_i} of V by the vectors h_i cover the whole space. The stratified random grid defined by the parallelepiped V is constructed in the following way :

~ an origin $x_o \in V$ is chosen

~ for any index i, a point $X_i \in V$ is chosen at random, so that the X_i are inde-

pendent and follow the same uniform probability distribution within V.

~ a sample is taken at each of the points $x_o + h_i + X_i$.

The estimator $Q^*(x_o)$ of the quantity of metal for a given $x_o \in V$, is then :

$$Q^*(x_o) = \sum_i v\, f(x_o+h_i+X_i)$$

For a given x_o, the $v\, f(x_o+h_i+X_i)$ are independent random variables. Their expectations and their variances can be expressed with the help of the regularization :

$$f_v(x) = \frac{1}{v} \int_v f(x+y)\, dy$$

(which is the mean of f within the parallelepiped v_x centred at x). We get :

$$\begin{cases} E(f(x_o+h_i+X_i)|x_o) = \frac{1}{v} \int_v f(x_o+h_i+y)\, dy = f_v(x_o+h_i) \\ \\ D^2(f(x_o+h_i+X_i|x_o) = \frac{1}{v} \int_v f^2(x_o+h_i+y)\, dy - f_v^2(x_o+h_i) \end{cases}$$

These variables being independent for a given x_o, we deduce the conditional expectation and variance of $Q^*(x_o)$:

$$\begin{cases} E(Q^*|x_o) = \sum_i v\, f_v(x_o+h_i) = Q \\ \\ D^2(Q^*|x_o) = v^2 \sum_i D^2(f(x_o+h_i+X_i)|x_o) = v\, g(o) - v^2 \sum_i f_v^2(x_o+h_i) \end{cases}$$

For a given x_o, $Q^*(x_o)$ is thus an unbiased estimator. But - just as in the previous paragraph - our ignorance at the start about the true location of the real phenomenon can be expressed by considering x_o as a random variable, independent of the X_i, and following a uniform probability distribution within v. As $E(Q^*|x_o) = Q$ does not depend on x_o, the a priori variance of $Q^*(x_o)$ - that is the <u>estimation variance</u> of the stratified random grid - is obtained directly : taking the expectation of the conditional variance $D^2(Q^*|x_o)$ with respect to x_o, we get :

$$D^2(Q^*) = \frac{1}{v} \int_v D^2(Q^*|x_o)\, dx_o = v\, g(o) - v \sum_i \int_v f_v^2(x_o+h_i)\, dx_o =$$

$$= v\, g(o) - v \int f_v^2(x)\, dx$$

Let g_v be the covariogram of the regularization f_v. From paragraph (1-3-1), it can be deduced from the geometrical covariogram K(h) of the parallelepiped v by the relationship :

$$g_v = \frac{1}{v^2}\, g * K$$

In particular

$$g_v(o) = \frac{1}{v^2} \int g(h)\, K(h)\, dh = \int f_v^2(x)\, dx$$

According to Cauchy's algorithm, $g_v(o)$ represents the mean value of g(x-y) when x and y sweep separately the volume v (cf. exercise 2), and the estimation variance we are seeking can be written as :

$$D^2(Q^*) = v[g(o) - g_v(o)]$$

It depends only upon the behaviour of g(h) within a neighbourhood of the origin which is precisely defined by the parallelepiped v itself.

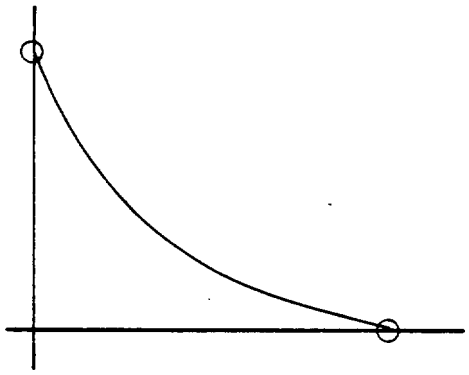
1-4-3 Approximation formulae in the one-dimensional space.

Let f(x) be a Re. V. in a one-dimensional space, g(h) its covariogram, and a the sampling interval. When the spacing a is small, formula (1-14) contains a large number of terms, and it becomes desirable to obtain approximation formulae to speed up calculations. As a matter of fact, the interest of the approximation formulae about to be established goes far beyond reasons of simple convenience, and has a fundamental epistemological significance regarding the problem with which we are concerned. Indeed, it will be seen that the estimation variance $\sigma^2(a)$ depends only on the analytical behaviour of g(h) near h = 0 when a is small (except for a fluctuating term whose meaning will be examined).

The estimation variance $\sigma^2(a)$ in formula (1-14) appears as the numerical error which would be made in computing the integral $\int g(h)\,dh$ with the help of the discrete sums $a \sum_k g(ka)$. Thus, if $g(h)$ were a sufficiently regular function (that is differentiable enough times everywhere), it would be possible to compute $\sigma^2(a)$ directly with the help of classical methods based upon the Euler-Mac-Laurin formula. But in general, $g(h)$ is not sufficiently regular for any h. The theoretical models of covariograms that are used in practice locate analytical irregularities at two critical spots :

~ near the origin, as already seen

~ but also near h = b, where b is the range, as the intersection at b of $g(h)$ with the axis can be more or less acute. Anywhere else, the theoretical models of $g(h)$ never present any analytical irregularity. Obviously this does not mean that this applies also to actual covariograms. But the true ones are never determined by experiment, and as previously seen at the end of paragraph (1-4-1), we always have to substitute a theoretical model for them. In paragraph (1-4-6) we shall examine the epistemological meaning of the equivocal operation of replacing the true (unknown) covariogram by an analytically simple theoretical model whose irregularities are concentrated near the origin and near the range.



For the moment, we begin with the assumption that the irregularities are effectively located at these two spots. Without going into the calculations at this stage (cf. [4]), it is possible to show that when a is small, the estimation variance is the sum of two terms :

$$\sigma^2(a) = T(a) + Z(a)$$

The first one, T(a), the extension term, depends on the behaviour of $g(h)$ near the origin ; the second one, Z(a), the fluctuating term or Zitterbewegung, depends on the behaviour of $g(h)$ near

the range b.

Concerning the fluctuating term Z(a), it can be shown that this depends essentially on the quantity $\varepsilon = b/a$ modulo 1 (more precisely, we may find an integer n such that $na \leq b < (n+1)a$, and put $\varepsilon = \frac{b-na}{a}$). Z(a) is a periodic function of ε, of period 1, and can be expanded in a Fourier series which has no constant term, and so has a mean value equal to 0 when ε varies between 0 and 1 :

$$\int_0^1 Z_\varepsilon(a)\, d\varepsilon = 0$$

It is possible to determine the theoretical expression for this fluctuating term. Its amplitude is not negligible, and it can even be very great (see fig. 1 and Ex. 14). However, in practice, the exact value of this term can never be calculated as the true value of the range b is only known up to $\pm$ a, thus up to ε, which is $\frac{b}{a}$ modulo 1, and therefore completely indeterminate. But as the mean value of this fluctuating term with respect to $\varepsilon \in [0,1]$ is always equal to 0, it seems justifiable to neglect it, and this will be done in what follows. The disguised probabilistic character of this approximation will be noticed and we shall refer again to it in paragraph 1-4-6.

Let us now study the principal term T(a), or extension term, linked to the analytical behaviour of g(h) near h = 0. A detailed analytical study (cf. [4]) shows that it depends only on the irregular part of the limited expansion of g(h) near the origin. More precisely, the validity of a principle of term by term correspondence can be established, similar to that encountered when studying grading in the isotropic case : each term in $|h|^\lambda$ in the expansion of g(h) makes a contribution of $T_\lambda\, a^{1+\lambda}$ to $\sigma^2(a)$, with a coefficient T_λ which is only a function of λ :

$$|h|^\lambda \rightarrow T_\lambda\, a^{1+\lambda} \tag{1-18}$$

When λ is an even integer, we get $T_\lambda = 0$, and this corroborates the fact that the regular part of the covariogram makes no contribution to the expansion of the estimation variance. When λ is an odd integer, the rule (1-18) still holds good. For a logarithmic term in $h^{2n} \log h$, we get :

$$h^{2n} \log h \rightarrow T'_{2n}\, a^{1+2n}$$

with T'_{2n} equal to the value at $\lambda = 2n$ of the derivative $\frac{d}{d\lambda} T_\lambda$.

Computation of the coefficient T_λ : A rigorous calculation of T_λ will be found in [4]. Here we will be content with a symbolic justification (similar to that already given in the computation of the coefficient A_λ of the grading rule). If G(u) is the Fourier transform of the covariogram g, we get, according to (1-15) :

$$\sigma^2(a) = 2 \sum_{p=1}^{\infty} G\left(\frac{p}{a}\right)$$

If $|h|^\lambda$ had been taken to be the covariogram, the transform G (in the sense of distributions) would be :

$$G(u) = \frac{1}{\pi^{\lambda+\frac{1}{2}}} \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(-\frac{\lambda}{2})} \frac{1}{|u|^{1+\lambda}}$$

and if formula (1-15) remains valid for this Fourier transform taken in the sense of distributions, we get :

$$\sigma^2(a) = \frac{2\, a^{1+\lambda}}{\pi^{\lambda+\frac{1}{2}}} \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(-\frac{\lambda}{2})} \sum_{p=1}^{\infty} \frac{1}{p^{1+\lambda}}$$

and hence the expression of T_λ :

(1-19)
$$T_\lambda = \frac{2}{\pi^{\lambda+\frac{1}{2}}} \frac{\Gamma(\frac{\lambda+1}{2})}{\Gamma(-\frac{\lambda}{2})} \sum_{p=1}^{\infty} \frac{1}{p^{1+\lambda}}$$

In fact, this value as found by a somewhat approximate reasoning is the true value. In particular :

$$\begin{cases} T_1 = -\frac{1}{6} \\ \\ T_2' = 0{,}0609 \ldots . \end{cases}$$

The correspondence principle gives a rapid and easy calculation of the estimation variance. But it must be remembered that the Zitterbewegung has been neglected, the amplitude of which can be very great, and also that the result has been obtained from only a limited expansion of $\sigma^2(a)$ near $a = 0$, and so is only true for small grid spacings. When a is large, the general formula (1-14)

contains only a few terms, and thus can be used directly. When the spacing a becomes greater than the range, only the term corresponding to k = 0 remains in (1-14) and we obtain the following simple formula :

$$(1\text{-}20) \qquad \sigma^2(a) = a\, g(o) - \int g(h)\, dh \qquad (a \geq b)$$

the probabilistic significance of which is given in Exercise 15.

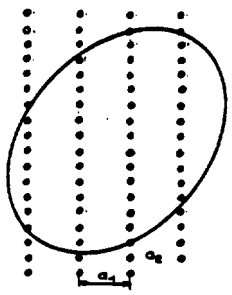
REMARK - If L = na is the mineralized length, and n the number of positive samples, the estimation variance corresponding to the covariogram in r^λ is :

$$\sigma^2(a) = A_\lambda\, a^{1+\lambda} = (A_\lambda\, L^{1+\lambda})\, \frac{1}{n^{1+\lambda}}$$

It is in inverse ratio to $n^{1+\lambda}$ (and not proportional to 1/n as would have been suggested by a maladroit application of classical statistics).

1-4-4 Approximation formulae in $\mathbb{R}^n$.

The two-dimensional case will be studied thoroughly, the generalization to the n-dimensional case following immediately. Our aim is to find a second approximation principle, which enables the computation of an estimation variance $\sigma^2(a_1,a_2)$ to be reduced to the sum of two estimation variances in a single dimension case, each obeying the correspondence rule (1-18) - (1-19) and so being easy to compute. This principle will be called the principle of composition of line terms and slice terms. Let us first examine the intuitive significance of this principle. Let (a_1,a_2) with $a_1 > a_2$ be a rectangular grid, f(x,y) a Re. V. in the plane, $g(h_1,h_2)$ its covariogram, and $g_1(h)$ the covariogram deduced by grading along the y axis. Our principle states that the estimation variance $\sigma^2(a_1,a_2)$ with a rectangular grid is the sum



$$(1\text{-}20) \qquad \sigma^2(a_1,a_2) = \sigma_1^2(a_1) + a_1\, \sigma_2^2(a_2)$$

of two terms : the first one, $\sigma_1^2(a_1)$ is computed from relationship (1-14) applied to the covario-

gram g_1 deduced by grading. It represents the error made in estimating the quantity of metal Q from the lines of greatest sampling density (parallel to the y axis) : it is the slice term. The second one, $a_1\ \sigma_2^2(a_2)$ is called the line term. It represents the error made when estimating the accumulations of the lines from the punctual samples. $\sigma_2^2(a_2)$ is computed with the help of formula (1-14) applied to the spacing a_2 and to the one-dimensional covariogram g(o,h). The principle (1-20) states thus that : when computing estimation variances, it may be assumed that the errors made in estimating the line accumulations from the samples, and in estimating the quantity of metal from the accumulations (supposedly known) are without correlation.

Let us give a more precise statement of this principle, before giving an approximate justification. In order to condense notations, δ_a will designate the linear operator associated, according to (1-14), with the estimation variance with a spacing of a :

$$\delta_a\, g = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h)\, dh$$

Let (x_o, y_o) be the coordinates of one of the samples. Let

$$L_i(x_o) = \int f(x_o + i\ a_1\ ;\ y\ )\ dy$$

be the accumulation of the line i at the abscissa $x_o + i\ a_1$, and :

$$L_i(x_o, y_o) = a_2 \sum_{p=-\infty}^{+\infty} f(x_o + i\ a_1\ ;\ y_o + p\ a_2)$$

the estimator of L_i which we can form. As seen in paragraph 1-4-1, x_o and y_o have to be considered as two independant random variables having uniform probability densities on (o, a_1) and (o, a_2) respectively. Thus the accumulation $L_i(x_o)$ is a random variable and clearly, for a given x_o :

$$E(L_i^* | x_o) = L_i(x_o)$$

In order to estimate $Q = \iint f(x,y)\ dx\ dy$, we form the following estimator :

$$Q^*(x_o, y_o) = a_1 \sum_i L_i^*(x_o, y_o)$$

and the estimation error can be written as :

$$(1\text{-}21) \qquad Q - Q^* = (Q - a_1 \sum_i L_i) + a_1 \sum_i (L_i - L_i^*)$$

The first term is the error made when estimating Q from the accumulations. From the definition of grading, the variance of this term (slice term) is obtained when applying (1-14) to the covariogram g_1 deduced from $g_2(h_1,h_2)$ by grading along the y axis, that is, with our notation :

$$\sigma_1^2(a_1) = \delta_{a_1} g_1$$

The other term is the sum of the errors $L_i - L_i^*$ made when estimating each one of the accumulations with the help of the punctual samples. For each line i, the expectation of this error is equal to 0, as seen previously when x_o is given :

$$E(L_i - L_i^* | x_o) = 0$$

For a given x_o, the Re. V. $f(x_o+i\,a_1,y)$ defined on the line i has for its covariogram g_i in one dimension :

$$(1\text{-}22) \qquad g_i(h) = \int_{-\infty}^{+\infty} f(x_o+i\,a_1,y)\, f(x_o+i\,a_1,y+h)\, dy$$

and the conditional variance of $L_i - L_i^*$ is then

$$D^2(L_i - L_i^* | x_o) = \delta_{a_2} g_i$$

We now assume (justified further on) that the $L_i - L_i^*$ can be considered as being <u>without correlation</u> at a given x_o. The conditional variance of the line-term is then :

$$a_1^2 \sum_i D^2(L_i - L_i^* | x_o) = a_1^2 \sum_i \delta_{a_2} g_i$$

As $E(L_i - L_i^*) = 0$, the a priori variance of the line term is then obtained by taking the expectation with respect to x_o :

$$a_1 \int_0^{a_1} dx_0 \sum_i \delta_{a_2} g_i$$

However, it can be seen that the linear operator δ_{a_2} allows the discrete sum to be changed for the integral. Using (1-22), we find for the line term :

$$a_1 \delta_{a_2} \int_0^{a_1} dx_0 \sum_i \int_{-\infty}^{+\infty} f(x_0 + i\, a_1 ; y)\, f(x_0 + i\, a_1 ; y+h)\, dy =$$

$$= a_1 \delta_{a_2} \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} f(x,y)\, f(x,y+h)\, dy = a_1 \delta_{a_2} g(o,h)$$

The line term is thus :

$$a_1 \delta_{a_2} g(o,h) = a_1 \sigma_2^2(a_2)$$

according to (1-20). To obtain the relationship (1-20) itself, we still have to admit there is no correlation between the two errors $Q - a_1 \Sigma L_1$ and $a_1 \Sigma(L_i - L_i^*)$ of (1-21).

This principle of composition of line-term and slice-term has a very great importance, first for practical reasons for it allows the easy computation of an approximate value of the estimation variance, but also for methodological reasons : for, taking into account the correspondence principle, it shows that in two dimensions as well, the estimation variance depends mostly on the behaviour of g(h) near the origin. We will find this composition principle again in a similar form in the probabilistic version of the theory. Thus it is not a useless exercise to give a justification of it, even a symbolic one.

To do that, we shall use the correspondence principle. To any term in r^{λ} of the covariogram's irregular part, there corresponds by grading the term $A_\lambda r^{1+\lambda}$ in the covariogram g_1 deduced by grading (relationships (1-10) and (1-12)). Thus we get the term $A_\lambda T_{1+\lambda} a_1^{2+\lambda}$ in the expression of the slice-term (relationships (1-18) and (1-19)),,and similarly the contribution to the line-term of r^{λ} is $T_\lambda a_1 a_2^{1+\lambda}$. Our correspondence principle is thus expressed by the correspondence rule :

(1-23) $$r^{\lambda} \rightarrow A_\lambda T_{1+\lambda} a_1^{2+\lambda} + T_\lambda a_1 a_2^{1+\lambda}$$

Conversely, we can find again the principle (1-20) by applying rule (1-23) (provided that the covariogram g is isotropic or can be reduced to the isotropic form by a linear transformation). This rule will now be justified :

Justification of rule (1-23) :

Start from the general formula (1-16) where G(u,v) is the Fourier transform of the covariogram g.

$$\sigma^2(a_1,a_2) = \sum_{p,q} G\left(\frac{p}{a_1} , \frac{q}{a_2}\right) - G(o,o)$$

In the double sum $\sum_{p,q}$ the terms corresponding to q = 0 give the slice-term :

$$\sigma_1^2(a_1) = \sum_{p} G\left(\frac{p}{a_1} , 0\right) - G(o,o)$$

Indeed, G(u,o) is the transform of the covariogram deduced by grading $g_1(h)$: formula (1-15) shows then that this term is in fact the variance (in one dimension) $\sigma_1^2(a_1) = \varepsilon_{a_1} g_1$ computed from the covariogram g_1 . The second term, or line term, has thus the following exact expression :

$$2 \sum_{q=1}^{\infty} \sum_{p=-\infty}^{+\infty} G\left(\frac{p}{a_1} , \frac{q}{a_2}\right)$$

In order to justify (1-23), we have to show that the contribution of a term in r^λ (in the expansion of g) is $T_\lambda\ a_1\ a_2^{1+\lambda}$. In two dimensions, the transform of r^λ is

$$\mathcal{F}_2\ r^\lambda = \frac{1}{\pi^{1+\lambda}} \frac{\Gamma(1 + \frac{\lambda}{2})}{\Gamma(- \frac{\lambda}{2})} \frac{1}{(u^2 + v^2)^{1 + \frac{\lambda}{2}}}$$

So we have to establish the following relationship :

$$\frac{2}{\pi^{1+\lambda}} \frac{\Gamma(1 + \frac{\lambda}{2})}{\Gamma(- \frac{\lambda}{2})} \sum_{q=1}^{\infty} \sum_{p=-\infty}^{+\infty} \frac{1}{\left(\frac{p^2}{a_1^2} + \frac{q^2}{a_2^2}\right)^{1+ \frac{\lambda}{2}}} = T_\lambda\ a_1\ a_2^{1+\lambda} \tag{1-24}$$

For a given integer q, we have :

$$(1\text{-}25) \qquad \sum_{p=-\infty}^{+\infty} \frac{1}{\left(\frac{q^2}{a_2^2} + \frac{p^2}{a_1^2}\right)^{1+\frac{\lambda}{2}}} = \left(\frac{a_2}{q}\right)^{\lambda+2} \sum_{p=-\infty}^{+\infty} \frac{1}{\left(1+\frac{p^2 a_2^2}{q^2 a_1^2}\right)^{1+\frac{\lambda}{2}}}$$

But a_2 is supposed to be small compared with a_1, and the quantity :

$$b = \frac{a_2}{q\, a_1}$$

is small. The Riemann sum $\sum_p \frac{b}{(1+b^2p^2)^{1+\frac{\lambda}{2}}}$ gives therefore a good estimate of the integral

$$\int_{-\infty}^{+\infty} \frac{dx}{(1+x^2)^{1+\frac{\lambda}{2}}}$$

On the other hand, it can be shown that the error made in replacing this integral by the discrete sum is of a high order in b. Then with excellent precision, we have :

$$\sum_p \frac{1}{(1+b^2p^2)^{1+\frac{\lambda}{2}}} \neq \frac{1}{b} \int_{-\infty}^{+\infty} \frac{dx}{(1+x^2)^{1+\frac{\lambda}{2}}} = \frac{1}{b} \sqrt{\pi}\, \frac{\Gamma(\frac{1+\lambda}{2})}{\Gamma(1+\frac{\lambda}{2})}$$

Replacing b by its value $\frac{a_2}{q\, a_1}$ in (1-25), we get :

$$\sum_p \frac{1}{\left(\frac{q^2}{a_2^2} + \frac{p^2}{a_1^2}\right)^{1+\frac{\lambda}{2}}} = a_1\, a_2^{1+\lambda}\, \sqrt{\pi}\, \frac{\Gamma(\frac{1+\lambda}{2})}{\Gamma(1+\frac{\lambda}{2})}\, \frac{1}{q^{1+\lambda}}$$

Substituting this result in the first member of (1-24), it follows that :

$$a_1\, a_2^{1+\lambda}\, \frac{2}{\pi^{\lambda+\frac{1}{2}}}\, \frac{\Gamma(\frac{1+\lambda}{2})}{\Gamma(-\frac{\lambda}{2})} \sum_{q=1}^{\infty} \frac{1}{q^{1+\lambda}}$$

But, according to (1-19), this is none other than $T_\lambda\, a_1\, a_2^{1+\lambda}$, so that relationship (1-24) is proved, as well as the correspondence rule (1-23), and the composition principle is justified.

1-4-5 Application to the estimation of a surface.

We can apply the composition principle of the previous paragraph to the estimation of a surface. The Re. V. to be estimated here is the indicator $k(x)$ of the set S. As already seen, the geometric covariogram $K(h)$ associated with S is linear near the origin (paragraph 1-1), and we have

$$K(h) = S - |h| D_\alpha + \ldots$$

in which D_α represents half the diametral variation of S in the direction α of the vector h.

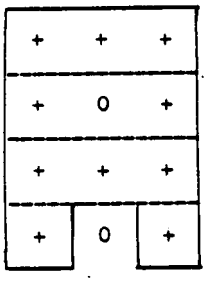
a/ We shall consider the isotropic case first, i.e. the case where $D_\alpha = D$ is almost independent of the direction α. With $T_1 = -\frac{1}{6}$ and $A_1 T_2 = -0{,}0609$, the rule (1-23) gives the estimation variance of S for a rectangular grid (a_1, a_2) in the form

$$\sigma_S^2 = D \; \frac{1}{6} \left(a_2 \, a_1^2 + 0{,}0609 \, a_2^3 \right) \qquad (a_1 \leq a_2) \tag{1-26}$$

We can also determine the relative variance $\frac{\sigma_S^2}{S^2}$. We do not, in fact, know the true value of S, but only an estimation of it, $n \, a_1 \, a_2$, where n denotes the number of boreholes having positive value. Dividing by $(n \, a_1 \, a_2)^{3/2}$ S, we obtain the following interesting expression :

$$\frac{\sigma_S^2}{S^2} = \frac{1}{n^{3/2}} \; \frac{D}{\sqrt{S}} \left(\frac{1}{6} \sqrt{\frac{a_1}{a_2}} + 0{,}0609 \left(\frac{a_2}{a_1} \right)^{3/2} \right) \tag{1-27}$$

The true value of the diametral variation is also unknown. To estimate it, we replace the set S by the union of positive bores' influence zones (rectangles a_1, a_2) We next count the numbers $2N_1$ and $2N_2$ respectively parallel to a_1 and a_2, which constitute the perimeter of the contour obtained. The corresponding diametral variations are $D_1 = N_1 \, a_1$ and $D_2 = N_2 \, a_2$. In the isotropic case, we have



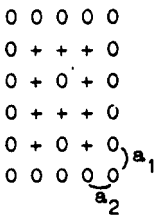
$$D = N_1 \, a_1 = N_2 \, a_2$$

Replace D in (1-26) and divide by $S^2 = (N a_1 \, a_2)^2$. This gives :

$$\frac{\sigma_S^2}{S^2} = \frac{1}{N^2} \left[\frac{1}{6} N_2 + 0{,}061 \; \frac{(N_1)^2}{N_2} \right] \qquad (N_2 \leq N_1) \tag{1-28}$$

b/ Nevertheless, the contour will not in general be sufficiently isotropic for D_α to be regarded as a constant D. It will present, for example, a principal axis of elongation. If one of the sides of the grid pattern is parallel to this principal axis - which will often be so - the above formula (1-28) still remains applicable : indeed, by taking this direction as the x axis and by multiplying the ordinates by a suitable factor, we get a new figure - isotropic this time - (at least to a first approximation) for which (1-28) is valid. But this linear transformation has changed neither N_1 or N_2, nor the relative variance σ^2/s^2, so that (1-28) is still valid in the case of the original anisotropic figure.

Exemple - On the figure below, the mineralized area is estimated as being 10 times the area of the grid rectangle $a_1\ a_2$. It contains a gap (or lacuna). N_1 and N_2 must include the interior elements as well as the exterior ones : hence we read on the figure



$$2\,D_1 = 12\,a_1 \quad \text{i.e.} \quad N_1 = 6$$

$$2\,D_2 = 8\,a_2 \quad \text{i.e.} \quad N_2 = 4$$

Therefore :

$$\frac{\sigma_S^2}{s^2} = \frac{1}{100}\left[\frac{4}{6} + 0{,}061\ \frac{36}{4}\right] = \frac{1{,}21}{100}$$

leading to a relative standard deviation $\sigma_S/S = \frac{11}{100}$ and a relative error range of $\pm$ 22% .

However, we must remember that this calculation disregards the fluctuating term or Zitterbewegung whose amplitude, as we know, can be very great.

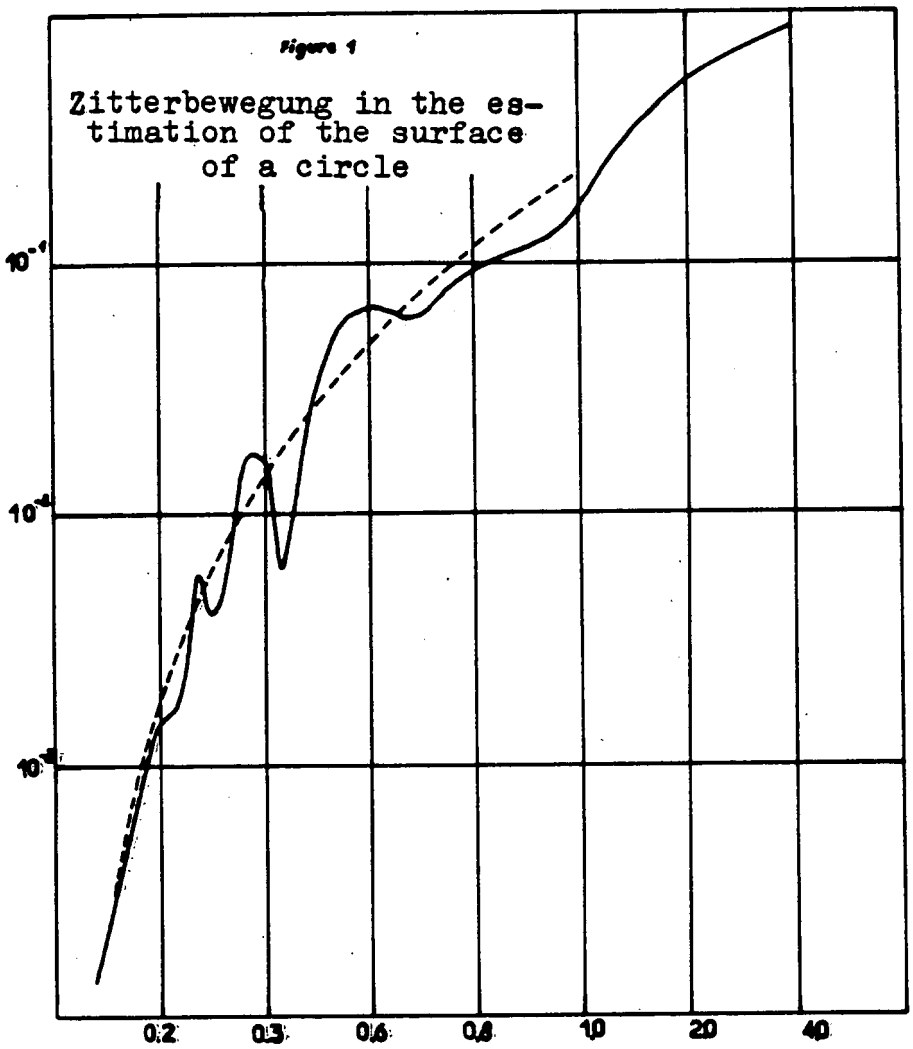


Fig. 1. – Zitterbewegung in the estimation of the surface of the circle of unit diameter, by means of a square grid network. Abscissae, the grid spacing a. Ordinates, the corresponding estimation variance $\sigma^2(a)$. The curve shown as an unbroken line represents the exact value, calculated on the basis of the exact formula (IV,1,14). The curve shown as a dashed line represents the formula $\sigma^2(a) = 0.2276a^2 + 0.00477a^3$ obtained by neglecting the Zitterbewegung and retaining the first two terms of the limited expansion given by the principle of correspondence.

1-4-6 Passage to the probabilistic version of the theory.

In this first chapter we have followed the spontaneous and in a way naïve point of view which consists in taking the phenomenon as it comes without making any particular assumptions about it. However, though transitive methods pretended at the start to build up the theory of Re. V. without any probabilistic interpretation, this ambition soon appeared to be unjustified. First, an unexpected phenomenon such as the Zitterbewegung (paragraph 1-4-3) is characteristic of a prerandom situation well known to physicists : where a slight change in the initial conditions, experimentally unnoticeable (here the ratio range upon spacing modulo 1) entails, after some time, a radical modification of the observed phenomenon. Moreover, should the covariogram g be known perfectly, it would show small undulations, singular points, a host of structural details which would give almost as much information as the Re. V. itself. But this is never experimentally obtainable. Starting from a discontinuous set of experimental points, we fit the covariogram to a theoretical model, which is a regular and continuous curve, except in the neighbourhood of the origin and the range (ending note of paragraph 1-4-1). More precisely, we make an assumption about the analytical behaviour (in h^{λ} for example) of g(h) near the origin. In fact, it is absolutely not certain that the true g(h) should show such a simple type of analytical behaviour. From the mathematical point of view, the smoothing or regularizing operations involved by such an assumption are very difficult to analyze with accuracy. Obviously, they obliterate the detailed structure of g(h) and all the potential information it contains. Though the mathematical character of these operations is somewhat obscure, their epistemological significance is obvious : they constitute a disguised passage to expectations. Thus transitive methods, meant at the start to be purely geometrical, emerge, when the conditions of their effective use are analyzed, as being rich in implicit probabilistic contents.

Hence we will bring into the open the probabilistic assumptions which the transitive methods concealed, and this will be our theory of intrinsic random functions. But the natural phenomena we are interested in must in general be considered as unique, so statistical inference will not be (theoretically) possible, unless a special hypothesis, such as the stationary or intrinsic type, is used. Sometimes this hypothesis will appear as likely, sometimes as being contrary to the experimental data, and in any case as arbitrary. Now it may seem illusory to interpret a natural phenomenon as a realization of a random function, unless we are sure we can reconstitute, without ambiguity, the proper probability law of this R.F. from the real phenomenon. Hence there appears to be a new contradiction. It has just appeared that the transitive representations contained an unacknowledged probabilistic content: but they did not involve any stationarity hypothesis (hence their

methodological interest). Let us now bring fully into the open their previously concealed probabilistic assumptions, but it will now appear that they only have significance when the hypothesis of stationarity is used, which the transitive methods did not need.

Bringing together these two contrary points of view clearly reveals the solution : the probabilistic interpretation is inevitable, but the hypothesis of stationarity is not in fact necessary. The Re. V, in general, has to be considered as a realization of a non-stationary R.F., and the basic tool will be a non-stationary covariance function C(x,y) depending separately on the two hold points x and y (and not only on the difference x-y). Statistical inference will still remain impossible, that is to say that it will not be possible to determine C(x,y) from the available data. But we do not need to know the actual value of this covariance. To see this, we only need to transpose the results of the transitive methods in probabilistic terms. For example, to estimate a volume V with indicator k(x) and geometric covariogram K(h), the transitive methods show that it is necessary to know the function :

$$g(h) = \int k(x)\ f(x)\ k(x+h)\ f(x+h)\ f(x)\ dx$$

In the probabilistic version, the covariogram g(h) becomes a R.F. itself, and it is necessary and sufficient to know its expectation E[g(h)] in order to solve the problem of estimation. But this expectation is deduced from the non-stationary covariance by the relationship :

$$E[g(h)] = \int k(x)\ k(x+h)\ C(x,x+h)\ dx \qquad (1\text{-}29)$$

In fact it is this very relationship which was already implicitly applied when substituting a theoretical model for the true unknown g(h). This expectation, apart from a factor, is not different from the pseudo-covariance $\overline{C}(h)$, the mean value of C(x,y) within the volume V to be estimated and which would be attributed to the random function if it were considered as stationary (but which may not be so). We have in fact :

$$\overline{C}(h) = \frac{1}{K(h)} \int k(x)\ k(x+h)\ C(x,x+h)\ dx = \frac{E[g(h)]}{K(h)}$$

Hence it is sufficient to know the expectation E[g(h)] or the mean covariance $\overline{C}(h)$ in order to solve the problem of estimation. Moreover, the approximation methods which were found show that it is even sufficient to know only the behaviour near the origin of this function E[g(h)] or $\overline{C}(h)$. In what follows, it will be shown that statistical inference from a single realization is possible for what concerns us (the behaviour near the origin), and that only a hypothesis of quasi-stationarity

needs to be specified : but this time the hypothesis will be weak enough to be physically conceivable in all the cases which concern us. In other words, the problem of estimation can be solved within the bounds of our probabilistic interpretation, and using only the available experimental data.

We can now appreciate the methodological importance of this first chapter, which may have looked rather theoretical at first sight: once probabilised, the results of the transitive method show the possibility of getting rid of the famous hypothesis of stationarity ; therefore, we have given quite a detailed account of it. Let us pass on now to the probabilistic theory.

1-5 EXERCISES ON TRANSITIVE METHODS.

1-5-1 Exercises on paragraphs 1-1 and 1-2.

Exercise 1 (Triangular covariogram) - Consider in a space of one dimension a Re. V. f(x) equal to 1 in the interval (0,b) and zero elsewhere (indicator of the segment of length b).

a) Find its geometric covariogram.

(Solution : $K(h) = b - |h|$ when $|h| \leq b$, $= 0$ when $|h| > b$. (Notice the linear behaviour near $h = 0$ and the straight intersection at $h = b$).

b) Same question in $\mathbb{R}^2$ for the rectangle $a \times b$. (Sol. : $K(h_1,h_2) = (a - |h_1|)(b - |h_2|)$ when $|h_1| \leq a$ and $|h_2| \leq b$) and in $\mathbb{R}^3$ for the parallelepiped $a \times b \times c$.

Exercise 2 (Cauchy's Formula) - Let V be a set belonging to $\mathbb{R}^n$, K(h) its geometric covariogram, and g(h) any function of the vector h. Show that the mean value of g(x-y) when the two extremities x and y of the argument of g sweep V separately, can be deduced from K(h) by the formula :

$$\frac{1}{V^2} \int_V dx \int_V g(x-y)\, dy = \frac{1}{V^2} \int g(h)\, K(h)\, dh$$

(Solution : write down explicitly the double integral with the help of the indicator of V, and change the variables and the order of integrations. This algorithm is used

constantly in the Re. V. theory).

Application : (cf. Ex. 1) If $g = g(r)$ only depends on the modulus r of h, $r = |h|$, the mean value of g within the rectangle a × b is :

$$\frac{4}{a^2b^2} \int_0^a dx \int_0^b (a-x)(b-y)\, g(\sqrt{x^2+y^2})\, dx\, dy$$

Exercise 3 - Consider in $\mathbb{R}^1$ the function $f(x) = e^{-ax}$ when $x \geq 0$ and $f(x) = 0$ when $x < 0$. Find its covariogram.
(Solution : $g(h) = \frac{1}{2a}\, e^{-a|h|}$ (linear behaviour near the origin)).

Exercise 4 - Consider the function $f(x) = x$ within the interval $(-b,+b)$ and $= 0$ otherwise. Show that $g(h) = \frac{2}{3} b^3 - b^2|h| + \frac{1}{6}|h^3|$ (linear at the origin). Prove that $\int g(h)dh = 0$ and interpret.

Exercise 5 - In $\mathbb{R}^1$, let $f(x)$ be the function $f(x) = (b^2-x^2)$ when $-b \leq x \leq b$ and $f(x) = 0$ otherwise. Show that the associated covariogram is :

$$g(h) = b^5\left(\frac{16}{15} - \frac{4}{3}\frac{h^2}{b^2} + \frac{2}{3}\frac{|h|^3}{b^3} - \frac{1}{30}\frac{|h|^5}{b^5}\right)$$

(Note the principal irregular term in h^3 : it is a continuous function. Rather than making a direct computation, prove that if f is differentiable, then its covariogram $g = f * \check{f}$ has for a second derivative $g'' = - f * (\check{f}')$, and use the results of Ex. 3. Integrate g" twice, noticing that $g'(o) = 0$ and compute $g(o)$ directly).

Exercice 6 - (Minkowski's formula) -: Let S be a sufficiently regular set in $\mathbb{R}^2$, K(h) its geometric covariogram, $2\mathcal{G}$ its perimeter. Establish the relationship :

$$2\mathcal{G} = - \int_0^\pi K'_\alpha(o)\, d\alpha$$

(put $- K'_\alpha(o)$ in the form of a curvilinear integral $\frac{1}{2}\int|\cos(\theta-2)|\, ds$ on the contour of S, and integrate first with respect to α).

Likewise, if V is a set in $\mathbb{R}^3$, K(h) its geometric covariogram, and S its surface, show that :

$$S = -\frac{1}{\pi}\int K'_{\omega}(o)\, d\omega$$

($d\omega$ is the elementary solid angle around direction ω, integrate over 4 π steradians). (These formulae are useful in mathematical morphology : they allow the perimeter or specific surface to be reconstituted from the derivative of the covariance function or of the covariogram, i.e. from purely linear observations).

Exercise 7 (Nugget effect) - In $\mathbb{R}^1$, the interval (0,b) is divided into n equal intervals of length $a = \frac{b}{n}$. Consider n independent random variables X_i with the same expectation m and the same variance σ^2 . Put $f(x) = X_i$ when $(i-1)a \leq ia$, $i = 1,2,\ldots n$, and $f(x) = 0$ outside (0,b). Find the expectation g(h) of the covariogram of f(x).

(Solution : $\sigma^2(b-n|h|) + m^2(b-|h|)$ when $|h| \leq a$, $m^2(b-|h|)$ when $a < |h| \leq b$, 0 otherwise. When a is small, or n great, this scheme corresponds to a nugget effect).

Exercise 8 (Size distribution of a convex grain) - a) Let S be a convex set of $\mathbb{R}^2$, g its geometric covariogram expressed in the form $g(r,\alpha)$, r, α being the polar coordinates. When α is given, the derivative with respect to r, $-g'(r,\alpha)$ is the measure of the projection of the intersection $S \cap S_h$ onto a line perpendicular to the direction α. Deduce that the size distribution of the intersections of S in the direction α follows the probability distribution function $F_{\alpha}(r)$ defined by :

$$1 - F_{\alpha}(r) = \frac{g'(r,\alpha)}{g'(o,\alpha)}$$

Deduce that, for a given α and for $r \geq 0$, $g(r,\alpha)$ is a convex function of r.

b) The direction α is now drawn at random with density :

$$\frac{D_{\alpha}}{2\mathcal{B}}\, d\alpha = -\frac{g'(o,\alpha)}{2\mathcal{B}}\, d\alpha \qquad (0 \leq \alpha \leq \pi)$$

(cf. Ex. 6) to verify that this is indeed a probability distribution. This distribution gives to each direction α a probability proportional to the apparent contour of S in that direction, and corresponds exactly to the (conditional) distribution of the inter-

sections of S with a previous layout of random straight lines with uniformly distributed directions on o,π). Show that the corresponding size distribution is :

$$1 - F(h) = -\frac{1}{2\mathscr{B}} \int_0^{\pi} g'(h,\alpha)\, d\alpha$$

with the mean value $E(h) = \pi \frac{S}{2\mathscr{B}}$

Exercise 9 (Geometric covariogram of a sphere) - Calculate the geometric covariogram of a sphere of diameter D in R^3 .

($K(h) = \frac{\pi D^3}{6} \left(1 - \frac{3}{2} \frac{|h|}{D} + \frac{1}{2} \frac{|h|^3}{D^3}\right)$ when $|h| \leq D$. Start from $-K'(h) = \frac{\pi}{4}(D^2-h^2)$ (cf. previous exercise) and integrate, remembering that $g(o) = \frac{\pi D^3}{6}$).

1-5-2 Exercises on grading.

Exercise 10 (grading of order 2) - Perform directly the isotropic grading of order 2 on the function $g_n(r) = r^{\lambda}$ $(0 \leq r \leq 1)$ $g_n(r) = 0$ $(r > 1)$. Deduce that the coefficients A_{λ} of relationship (1-10) verify :

$$A_{\lambda}\, A_{\lambda+1} = -\frac{2\pi}{\lambda+2}$$

(Solution : $g_{n-2}(r) = \frac{2\pi}{\lambda+2}(1 - r^{\lambda+2})$ for $r \leq 1$, and 0 when $r > 1$. The grading of order 2 follows the rule $r^{\lambda} \rightarrow A_{\lambda}\, A_{\lambda+1}\, r^{\lambda+2}$ in the irregular part ; the rule gives the result only except for a series of even integral powers, which is reduced to a constant in this case.

Exercise 11 - Covariogram of the function $f(x) = \pi \left(\frac{D^2}{4} - x^2\right)$ when $-D \leq x \leq D$, or 0 otherwise on R^1 (intersection of the sphere of diameter D by a plane at the reading x).

(Solution : $g(h) = \frac{\pi^2}{6} \left(\frac{1}{5} - \frac{h^2}{D^2} + \frac{|h|^3}{D^3} - \frac{1}{5}\frac{|h|^5}{D^5}\right) D^5$. Perform the grading of order 2 directly on the geometric covariogram of Ex. 9. Compare with Ex. 5).

Exercise 11 bis (Reconstitution of the size distribution of spheres) - A population of spheres in $\mathbb{R}^3$ is given. Let θ_3 be the number of spheres per unit of volume, and $F_3(D)$ be their size distribution. In other words, $\theta_3[1-F_3(D)]$ is the number of spheres with diameters $\geq D$ per unit volume. These spheres induce respectively circles and chords on the planes and lines of $\mathbb{R}^3$. Let θ_2 be the number of circles per unit surface, and $1-F_2(D)$ be their size distribution (size distribution induced by F_3 on the planes) ; in the same way, let θ_1 be the number of intersections per unit length and $1-F_1(D)$ be the size distribution of these intersections (induced on the lines by F_3 or by F_2) .

a/ Establish :

$$\theta_1[1-F_1(h)] = \theta_2 \int_h^\infty \frac{D}{\sqrt{D^2-h^2}} \left(1-F_2(D)\right) dD = \frac{\pi}{4} \theta_3 \int_h^\infty \left(1-F_3(D)\right) 2D\, dD$$

$$\theta_2[1-F_2(h)] = \theta_3 \int_h^\infty \frac{D}{\sqrt{D^2-h^2}} \left(1-F_3(D)\right) dD$$

(for example, establish first $\theta_1[1-F_1(h)] = \theta_2 \int_h^\infty \sqrt{D^2-h^2}\, F_2(dD)$, and integrate by parts).

b/ Invert these relationships : using (1-8), show that (for example) :

$$\begin{cases} \theta_1 F_1'(h) = \frac{\pi}{2} \theta_3 h[1-F_3(h)] \\ \theta_2[1-F_2(h)] = \frac{2}{\pi} \theta_1 \int_h^\infty F_1'(u) \frac{du}{\sqrt{u^2-h^2}} \end{cases}$$

Deduce $\theta_3 = \frac{2}{\pi} F_1''(o)\, \theta_1$, $\theta_2 = \frac{2}{\pi} \int_o^\infty \frac{1}{u} F_1(u)\, du$, and reconstitute the size distribution of the circles and the spheres from that of the chords.

Exercise 12 (Grading on the Gaussian exponential) - Show that a grading of order 1 transforms e^{-ar^2} into $\sqrt{\frac{\pi}{a}}\, e^{-ar^2}$.

1-5-3 Exercises on estimation.

Exercise 13 (Exponential covariogram) - Consider a Re. V. $f(x)$ whose transitive covariogram in one dimension is $g(h) = e^{-\lambda|h|}$. Form the exact expression of the estimation variance $\sigma^2(a)$, find its principal part, and draw conclusions concerning the absence of a Zitterbewegung.

(Solution : $\sigma^2(a) = a\left(\frac{2}{1-\lambda a} - 1 - \frac{2}{\lambda a}\right) \# \frac{1}{6}\ \lambda\ a^2$

By taking the expansion further, we could identify the terms T_1, T_3, T_5... of relationships (1-18). In particular, $T_1 = -\frac{1}{6}$)

Exercise 14 (Triangular covariogram) - Consider in $\mathbb{R}^1$ $f(x) = 1$ within $(0,b)$.

a/ Find the covariogram of $f(x)$.

(Solution : $g(h) = b - |h|$ when $|h| \leq b$ and 0 elsewhere. Establish this result geometrically).

b/ Calculate the estimation variance for a grid having a small spacing a, by means of the approximation formula. (Sol.: $1/6\ a^2$).

c/ Consider a layout of n+2 boreholes S_0, S_1,...,S_n, S_{n+1} on a regular grid, the first and last of which are negative, the others positive. We try to estimate b, the length of an interval, each extremity of which can fall anywhere on the segments(S_0 , S_1) and (S_n , S_{n+1}) respectively,independently of each other,and with uniform probability. The length b is then a random variable. Show that $E(b) = na$ and $D^2(b) = a^2/6$: the estimation variance as given by the approximation formula is thus found again, but we cannot obtain the fluctuating term.

d/ Calculate the exact value of $\sigma^2(a)$. To do this, let $b = na + \varepsilon a$ with $0 \leq \varepsilon < 1$. As ε is unknown in practice, consider it to be a Re. V. with uniform probability density on (0,1) and re-establish the approximate formula of b.

(Solution : If the first positive sample is at x ($0 \leq x < a$), we get the estimation $(n+1)a$ when $x \leq \varepsilon a$ and na when $\varepsilon a < x < a$. As x is set at random on $(0,a)$ with uniform density, this estimation is a R.V. of expectation b and variance $(\varepsilon - \varepsilon^2)a^2$. This expression includes the Zitterbewegung).

Exercise 15 (Estimation variance in the random case) - Let $f(x)$ be a Re. V. in two dimensions, having a covariogram $g(h)$ and a field S. If the rectangular grid pattern (a_1, a_2) is large with respect to S, at the mos one sample is positive. We may consider the sample as falling at random within a rectangle (a_1, a_2) containing S. The estimator is $Q^* = a_1 a_2 f(x)$ if this sample falls at a point $x \in S$, and 0 otherwise. $Q^*(x)$ is thus a R.V. Show directly that $E(Q^*) = \int f(x)\,dx = Q$, $E[(Q^*)^2] = [\int [f(x)]^2\,dx]\, a_1 a_2$ and $D^2(Q^*) = a_1 a_2\, g(0) - \int g(h)\,dh$.

1-5-4 Transitive theory for measures.

Exercise 16 (Generalization of the transitive theory for measures) - Let $f(x)$ be a summable positive function in $\mathbb{R}^n$. Suppose that $\int f(x)dx = 1$ (which is permissible) and consider $f(x)$ as the probability density associated with a vectorial variable $X = (X_1, X_2, \ldots X_n)$. Let X and X' be two independent variables obeying the same law f. Show that the covariogram $g = f * \check{f}$ is the probability density of the variable X-X'.

Exercise 17 (Continuation) - Let $X = (X_1, X_2, \ldots X_n)$ be a vectorial variable, μ its probability distribution (positive measure with sum equal to 1 defined on $\mathbb{R}^n$) not necessarily having a density. The covariogram $g = \mu * \check{\mu}$ will be defined as the measure giving the distribution of X-X', X and X' designating two random variables of same distribution μ.

a/ Study the case where μ is a discrete distribution concentrated at N points x_1, $x_2, \ldots x_N$ with $P(X=x_i) = \frac{1}{N}$. Show that g is the discrete distribution comprising the mass $\frac{1}{N}$ at the origin and the mass $\frac{1}{N^2}$ at each of the points $x_i - x_j$, $i \neq j$.

b/ The N points x_i of a/ above are now considered as random, mutually independent, set in $\mathbb{R}^n$ according to the same distribution with density f. The distribution g is then considered as a conditional distribution for a given x_i. Show that the corresponding a priori (deconditioned) distribution is of the form :

$$g = \frac{1}{N}\delta + (1 - \frac{1}{N})\, f * \check{f}$$

(δ Dirac measure : $\frac{1}{N}\delta$ thus designates a mass $\frac{1}{N}$ set at 0, and signifies a nugget effect: in paragraph 1-2, the nugget effect is described as a very sharp and highly localised

transition zone around the origin. It is then natural, at the limit, to idealize the nugget effect by symbolizing it with a Dirac measure placed at 0).

Exercise 18 (Continuation) - p is a positive function of sum one. Show that the regularization $\mu * \check{p}$ is the density of X-Y, X and Y designating two independent vectorial variables with respective distribution μ and p. Interpret the covariogram $\mu * \check{\mu} * p * \check{p}$ of the regularization $\mu * \check{p}$ (distribution of X-X' - Y+Y'). Study the case where μ is the discrete distribution of Ex. 17 a/. When the points x_i are random (as in Ex. 17 b/) the above covariogram $\mu * \check{\mu} * p * \check{p}$ represents a conditional distribution. By deconditioning with respect to x_i, show that the a priori corresponding distribution is :

$$\frac{1}{N} p * \check{p} + (1 - \frac{1}{N}) p * \check{p} * f * \check{f}$$

From these exercises, conclude that it is natural to symbolize a nugget effect with a Dirac measure set at 0. A covariogram with a nugget effect will have the form $g = C\delta + g_1$, where g_1 is continuous near the origin. If we regularize the measure whose covariogram is g by the function p, we get a function with covariogram $g * P = C P + g_1 * P$ ($P = p * \check{p}$) so that the rule of paragraph 1-3-1 remains valid.

Exercise 19 (Continuation) - Let μ_n be the distribution of the vectorial variable $(X_1, X_2, \ldots X_n)$ having n components. Grading along the x_n axis makes us pass from μ_n to the marginal distribution μ_{n-1} of the n-1 variables $(X_1, X_2, \ldots X_{n-1})$. Show that the covariogram $\mu_{n-1} * \check{\mu}_{n-1}$ is deduced from the covariogram $\mu_n * \check{\mu}_n$ by the same grading process.

More generally, if p_n is a function (or a measure) on $\mathbb{R}^n$, and p_{n-1} the function (or the measure) deduced from p_n by grading, the regularization $\mu_n * \check{p}_n$ becomes $\mu_{n-1} * \check{p}_{n-1}$ through this same grading process. How does a nugget effect behave when it is graded ?

Exercise 20 (Continuation)- (Estimation in the presence of a nugget effect) - Let f be a measure with covariogram $g = C\delta + g_1$ comprising a nugget effect and a continuous component g_1. Let k(x) be the indicator of a small volume v, K its covariogram. f is regularized by 1/v k(x), that is $f_v = \frac{1}{v} f * \check{k}$: $f_v(x)$ is the grade of a sample v set at x. Samples v are taken on a regular grid $a = (a_1, a_2, \ldots a_n)$. Show that the estimation variance is of

the type $\sigma^2(a) = \sigma_p^2(a) + \sigma_t^2(a)$, $\sigma_t^2(a)$ being calculated from g_1 by means of the formulae of paragraph 1-4-1 . The supplementary term, or nugget term is $\sigma_p^2(a) = C\left(\frac{a_1 a_2 \cdots a_n}{V} - 1\right)$ or else $C[V/V' - 1]$: V is the volume of the field, V' the total volume of the samples (i.e. the supplementary term produced by the nugget effect is in inverse ratio to the total volume of the samples).

CHAPTER 2

THEORY OF INTRINSIC RANDOM FUNCTIONS

2-1 GENERAL DEFINITIONS

Notion of a Random Function.

In probability theory, a sequence of k random variables (r.v.) Y_1, Y_2,...Y_k (generally non-independent) defines a vectorial random variable (or random vector) $Y = (Y_1, Y_2, \ldots Y_k)$ with k components. When the number k of these components becomes infinite, we get an infinite family of r.v.'s: this is a random function. In particular, if x is a point describing the space of n dimensions R^n, an infinite family $(Y_x)_{x \in R^n}$ can be defined. Thus, to every point x_0 of the space, there corresponds an ordinary r.v. Y_{x_0}, also denoted by $Y(x_0)$. Y(x) is then a function of the point x, whose "value" at x_0 is not a number but a r.v. (namely $Y(x_0)$). We say that Y(x) is a random function (for short R.F.). It will be noted that generally the r.v.'s corresponding to two support points x_1 and x_2, i.e. $Y(x_1)$ and $Y(x_2)$ are not independent.

If Y is an ordinary r.v., the result of drawing Y at random according to its probability distribution is a particular numerical value y. In the same way, if Y is a vectorial r.v. $(Y_1, Y_2 \ldots Y_k)$, a random drawing of Y according to its distribution (of k variables) gives a vector $y = (y_1, y_2 \ldots y_k)$ i.e. k particular numerical values. Finally, if Y(x) is a R.F. - i.e. a vectorial r.v. with an infinite number of components - the drawing at random of Y(x) according to its distribution (of an infinite number of variables) gives a particular numerical function y(x), generally a highly irregular one. The function y(x) is said to be a realization of the R.F. Y(x).

A realization y(x) of a R.F. Y(x) can always be considered as a regionalized variable. Conversely, a given Re. V. y(x) can be regarded as the realization of a certain R.F. Y(x) : this interpretation makes the application of the probabilistic theory of R.F.'s to the Re. V.'s possible.

Remarks - 1/ It can never be said that a given Re.V. y(x) is a R.F.. This would make no more sense than to say that the number 98 is a r.v. The correct statement of the basic probabilistic hypothesis we wish to introduce is : y(x) is a realization of a random function Y(x).

2/ For this probabilistic hypothesis to have real significance, we should be able to reconstitute, at least in part, the law of the R.F. Y(x), of which y(x) is supposed to be a realization, and that implies that statistical inference is possible. But in general, statistical inference is not possible if we have only one realization y(x) of Y(x) at our disposal. (In the same way, the distribution of a r.v. Y cannot be reconstituted on the basis of the numerical result y = = 98 of a single trial). To make statistical inference possible, the introduction of supplementary hypotheses about the R.F. Y(x) is necessary in order to reduce the number of "parameters" on which its law depends. Such is the aim of the stationarity hypothesis we are going to define : a stationary R.F. is, in a way, repeating itself in space, and this repetition gives a new opportunity for statistical inference, starting from a single realization. Let us now specify this hypothesis :

Stationary R.F. - A R.F. Y(x) is called stationary when its law is invariant under translation: in other words if $x_1, x_2, \ldots, x_k$ are k arbitrary support points (k being any integer) and h any vector, the k r.v. $Y(x_1)$, $Y(x_2), \ldots, Y(x_k)$ have the same distribution (of k variables) as the k r.v. $Y(x_1+h)$, $Y(x_2+h), \ldots, Y(x_k+h)$. In what follows, Y(x) will represent a stationary R.F.

Expectation. - Consider a support point x_0. If the ordinary r.v. $Y(x_0)$ has an expectation, this is a function $m(x_0) = E[Y(x_0)]$ of the support point x_0 . But Y(x) is stationary, and consequently $m(x_0+h) = m(x_0)$ for any vector h ; hence $m(x_0)$ is a constant m independent of x_0 :

$$m = E[Y(x)]$$

Even if we have to replace Y(x) by [Y(x)-m], we will very often assume that m = 0, (provided that this expectation does actually exist).

The covariance K(h). - Consider now the two points x_0 and x_0+h. If the two r.v.'s $Y(x_0)$ and $Y(x_0+h)$ have finite variances, (and hence also an expectation m which we shall suppose to be equal to 0) they also have a covariance $K(x_0;h)$ which in principle depends on the support point x_0 and on the vector h. But as Y(x) is stationary, we have : $K(x_0+a;h) = K(x_0;h)$ for any vector a : hence $K(x_0;h)$ is independent of x_0, and we will simply write K(h) as :

$$K(h) = E[Y(x)\ Y(x+h)] \tag{2-1}$$

Compare this definition of the covariance with that of the transitive covariogram : K(h) is the probabilistic transposition of g(h). When h = 0, we have $K(o) = E\left([Y(x)]^2\right)$: it is the variance of the

r.v. $Y(x_o)$. The stationary R.F. $Y(x)$ has a covariance function $K(h)$ if and only if it has a finite variance $K(o)$.

Second order stationary hypothesis. - A R.F. $Y(x)$ is said to be stationary of the second order if the r.v. $Y(x_o)$ has an expectation independent of the support point x_o, and if, for any vector h, the covariance

$$K(h) = E[Y(x_o+h)\ Y(x_o)] - m_o^2$$

exists and is independent of x_o . This hypothesis (which does not imply stationarity in the strict sense, as we have defined it above) is sufficient for the theory of the Re. V. But it does suppose a finite a priori variance $K(o)$.

Infinite a priori variance. - Many phenomena have an almost unlimited capacity of dispersion, and cannot be properly described by attributing to them a finite a priori variance : this assertion is perhaps surprising, but we have to realize that nature is setting a sort of trap here. When samples v are taken from a field V, we get a histogram from which it is always possible to calculate a variance which takes thus a perfectly definite value. But this experimental variance is in fact a function $\sigma^2(v/V)$ of the support v and the field V. In particular, it increases as the field increases. If the samples of size v have a finite a priori variance, this should appear as the limit of the experimental variance $\sigma^2(v/V)$ in an infinite V.

It was in this way that the South African School (D.G. Krige, etc...), starting from hundreds of thousand of samples drawn from the large gold ore-body of the Rand, have been able to calculate the variance of these samples in larger and larger panels, then in an entire concession, and then in the whole Rand ore-body : thus they have obtained an experimental relationship of the form

$$\sigma^2(v/V) = \alpha \log (V/v)$$

The variance increases in accordance with this logarithmic rule (the De Wijs formula) right up to the last experimental point, for which the ratio V/v is about ten thousand million. It can be concluded, beyond any doubt, that in this case, an a priori finite variance does not exist.

Hence we must replace the stationary of second order hypothesis by a weaker hypothesis of the same meaning :

Intrinsic Hypothesis.- Even when the a priori variance K(o) does not exist (i.e. is infinite) it may happen that the increments $Y(x_o+h) - Y(x_o)$ have a finite variance. Hence we will say that the R.F. Y(xp) obeys the intrinsic hypothesis if, for any vector h, the increment $Y(x_o)h - Y(x_o)$ has an expectation and a variance which are independent of the support point x (but depend on h), i.e. :

$$E[Y(x+h) - Y(x)] = m(h)$$

$$D^2[\big(Y(x+h) - Y(x)\big)] = 2\,\gamma(h)$$

The function m(h) is the linear drift. To show that it is linear in h, we start from the obvious relationship:

$$Y(x+h'' + h') - Y(x) = [Y(x+h'' + h') - Y(x+h')] + [Y(x+h') - Y(x)]$$

and pass on to expectations, whence : $m(h'+h'') = m(h') + m(h'')$. Anyway, we can assume that this linear drift m(h) is equal to zero, even if we have to replace Y(x) by Y(x) - m(x).

The function $\gamma(h)$

$$\gamma(h) = \tfrac{1}{2} D^2[\big(Y(x+h) - Y(x)\big)^2] \tag{2-2}$$

is called the semi-variogram, or intrinsic function. A R.F. satisfying the intrinsic hypothesis (or I.R.F.) constitutes what is called an intrinsic scheme, characterized by its semi-variogram.

Remark. If Y(x) obeys the stationary hypothesis of second order, it also obeys the intrinsic hypothesis, and we have in that case

$$\gamma(h) = K(o) - K(h) \tag{2-3}$$

As $|K(h)| \leq K(o)$, we have $\gamma(h) \leq 2\,K(o)$ so that the semi-variogram of a stationary R.F. of second order is necessarily bounded. There are intrinsic schemes in general use, the semi-variograms of which are not bounded, and which consequently cannot satisfy the stationary hypothesis of second order (i.e. they have an infinite a priori variance). Ex. : De Wijs scheme ($\gamma(h) = 3\,\alpha \log |h|$), linear scheme ($\gamma(h) = A\,|h|$).

2-2 PROPERTIES OF THE COVARIANCE AND OF THE SEMI-VARIOGRAM.

A covariance or a semi-variogram are symmetric functions :

$$K(h) = K(-h) \ , \ \gamma(h) = \gamma(-h)$$

In addition, a covariance verifies Schwarz's inequality :

$$|K(h)| \leq K(o)$$

In the case of a variogram , we find only $\gamma(h) \geq 0$ and $\gamma(o) = 0$. However, it can be shown that the increase of a variogram at infinity is necessarily less rapid than that of $|h|^2$, that is :

$$\lim_{|h| \to \infty} \frac{1}{|h|^2} \gamma(h) = 0$$

These conditions are necessary, but even if a function K or γ satisfies them, it does not follow that there is a stationary or intrinsic R.F. having this particular covariance or that particular semi-variogram. In fact, it is so if and only if K belongs to the class of "positive-definite" functions, and $-\gamma$ to that of the "conditional positive-definite" functions. For example, the functions log r and r^λ with $\lambda < 2$ can be used as a semi-variogram but nor r^λ for $\lambda \geq 2$. These conditions express, amongst others, that the formulae which will be established about extension variances or estimation variances necessarily lead to positive values (which would not always be the case, had we taken any ordinary function as a semi-variogram).

These conditions deserve looking into more carefully, even if only to define more precisely the class of linear combinations we can use when there is no finite a priori variance.

2-2-1 Authorized linear combinations.

Consider first the stationary case of order two ; take m = 0, and take N points $x_1, x_2, \ldots x_N$ of $\mathbb{R}^n$ and N coefficients $\lambda_1, \lambda_2, \ldots \lambda_N$. The expression

$$Y = \sum_i \lambda_i \, Y(x_i) \tag{2-4}$$

is a random variable (being a finite linear combination of r.v.). As a covariance function K(h) does exist by hypothesis, Y has a finite variance, and an elementary calculation gives :

$$D^2(Y) = \sum_{i,j} \lambda_i \lambda_j K(x_i - x_j) \tag{2-5}$$

The obvious condition $D^2(Y) \geq 0$ expresses the requirement that any covariance function is positive-definite (paragraph 1-3). Conversely, if K(h) is a positive-definite type function, the possibility of constructing a R.F. with precisely this covariance K(h) can be demonstrated. Formula (2-5) appears in slightly different forms, in all parts of the theory. For the moment, note that in the stationary case of order two, any finite linear combination has a finite variance, and hence using them is always legitimate.

This is not the case when only a variogram exists, without a covariance. In this case, only linear combinations such as (2-4) obeying the condition

$$\sum_i \lambda_i = 0 \tag{2-6}$$

have a finite variance. Thus, in the case of an I.R.F. without a covariance, only linear combinations with the sum of coefficients equal to zero can be used. This plays a fundamental part in all the intrinsic theory. It can be shown that this condition (2-6) is sufficient. If it is verified, it can be written as :

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

Now the (non-stationary) R.F. defined by :

$$Z(x) = Y(x) - Y(o)$$

has a covariance function $C(x,y) = E[Z(x) Z(y)]$, as Z(x), as an increment of the I.R.F. Y(x), has a finite variance. Let us determine this (non-stationary) covariance function C(x,y) : it will be useful later on. The simplest way to do this is to start from the definition of the variogram :

$$2\gamma(x,y) = E[(Y_x - Y_o - Y_y + Y_o)]^2 = 2\gamma(x) + 2\gamma(y) - 2C(x,y)$$

Whence :

$$C(x,y) = \gamma(x) + \gamma(y) - \gamma(x,y) \tag{2-7}$$

When (2-6) is true, the linear combination $\sum \lambda_i Y_{x_i} = \sum \lambda_i (Y_{x_i} - Y_o)$ has the finite variance :

$$D^2(Y) = \sum_{i,j} \lambda_i \lambda_j C(x_i, x_j)$$

that is, taking into account (2-7) :

$$D^2(Y) = -\sum_{i,j} \lambda_i \lambda_j \gamma(x_i - x_j) + \sum_i \lambda_i \sum_j \lambda_j \gamma(x_j) + \sum_j \lambda_j \sum_i \lambda_i \gamma(x_i)$$

whence, from (2-6) :

$$D^2(Y) = -\sum_{i,j} \lambda_i \lambda_j \gamma(x_i - x_j) \tag{2-8}$$

This is another result which will be very useful in what follows : as soon as the sum of coefficients of a linear combination is equal to 0, we can compute its variance from formula (2-5), as if a covariance K(h) existed, on condition that K is replaced by $-\gamma$. This device is of very general application and often has a great simplifying effect on calculations.

Conversely, if there is no a priori finite variance, a linear combination with a finite variance is necessarily a linear combination of increments of the R.F. Y(x) for only these increments have a finite a priori variance. Consequently, this combination obeys (2-6), which is thus necessary and sufficient, as previously stated.

Thus, the condition that $-\gamma$ must be conditionnally positive-definite can be introduced very naturally. By definition, a function g(h) will be said to be conditionnally positive-definite if for any integer N, any system of N points $x_1, x_2, \ldots x_N$ in $\mathbb{R}^n$, and any system of coefficients $\lambda_1, \lambda_2, \lambda_3 \ldots \lambda_N$ obeying the condition $\sum \lambda_i = 0$, we have :

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

As the allowed linear combinationa have a positive or zero variance, relationship (2-8) indeed means that $-\gamma$ is conditionnally positive-definite. Conversely, if $-\gamma$ is conditionnally positive-

definite, it can be shown that it is possible to construct an I.R.F. with γ as a semi-variogram. This is thus a necessary and sufficient condition.

2-2-2 Continuity q.m. and other properties.

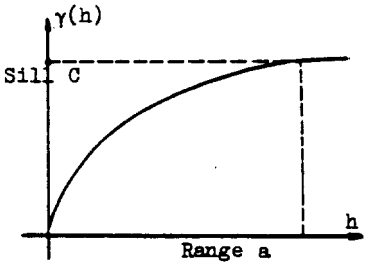
We shall next consider further properties of a covariance or of a covariogram. As in the transitive methods, our attention will be concentrated on the behaviour near the origin.

a/ It will be seen that the variogram gives a precise significance to the traditional notion of the zone of influence of a sample : its more or less rapid increase in fact reflects the greater or lesser rate of deterioration of the influence of the sample over the more and more distant parts of the ore-body.

b/ Anisotropies become apparent with the different behaviour of the variogram in different directions in space. In the absence of any anisotropy, $\gamma(h) = \gamma(r)$ only depends on the modulus r of h and not on its direction. It is said that there is a geometric anisotropy when a simple linear transformation of the coordinates is sufficient to restore isotropy.

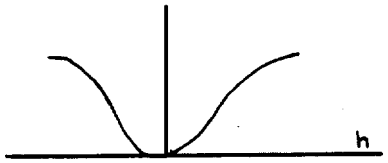
There are more complex kinds of anisotropy. For example, in a three-dimensional space, it may happen that Y(x) only depends on the third coordinate x_3 and consequently is constant in the planes parallel to the two other axes of coordinates. The semi-variogram thus only depends on the third component of h, $\gamma(h) = \gamma(h_3)$. Again, more often, Y(x) will not be actually constant in the horizontal plane, but will vary there less rapidly or more regularly than in the vertical direction. In this case, we will take as a semi-variogram $\gamma(h) = \gamma_0(h_1,h_2,h_3) + \gamma_1(h_3)$ (zonal anisotropy).

c/ Range. In the stationary case of second order, the range $a(\alpha)$ in the direction α is the value of the distance beyond which, in that direction, Y(x) and Y(x+h) are without correlation (or have a negligible one): $K(h) = 0$ (or $\neq 0$) for $|h| \geq a(\alpha)$. From c/, $K(h) = 0$ is equivalent to $\gamma(h) = K(o) = \gamma(\infty)$. The range is also the distance beyond which the semi-variogram has reached its limiting value $\gamma(\infty)$ or sill. Hence an intrinsic R.F. with an unbounded variogram cannot have a finite range.

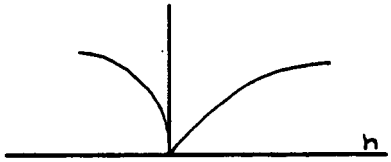


d/ Behaviour near the origin. The continuity and the regularity in the space of the R.F. Y(x) are expressed by the behaviour of the $\gamma(h)$ near the origin. In order of

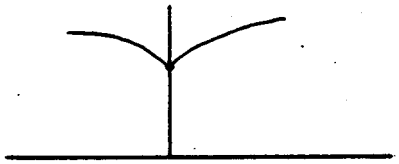
decreasing regularity, there are roughly four types to be found :



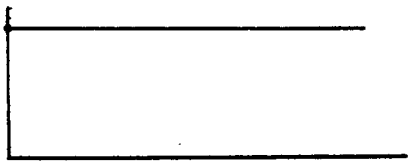
Parabolic trend : $\gamma(h)$ is twice differentiable at $h = 0$. $Y(x)$ is then itself differentiable (mean square convergence) and thus presents a high degree of regularity.



Linear behaviour (oblique tangent at the origin) : $\gamma(h)$ is continuous at $h = 0$ but non-differentiable. $Y(x)$ is continuous (in the mean square) but not differentiable and thus less regular.



Nugget effect : $\gamma(h)$ does not approach 0 when h approaches 0 (discontinuity at the origin). $Y(x)$ is not even continuous in the mean square, and is thus highly irregular.



Limiting case - completely random : $Y(x)$ and $Y(x')$ are independent for any two distinct points, regardless of how close they are (white noise of the physicists, or pure nugget effect).

Irregular part : In the isotropic case ($\gamma(h) = \gamma(r)$) and without a nugget effect, we shall characterize the behaviour of $\gamma(h)$ by a limited expansion such as :

$$\gamma(r) = \Sigma\, a_{2n}\, r^{2n} + \Sigma\, c_{\lambda}\, r^{\lambda} + \Sigma\, c_{2n}\, r^{2n} \log r$$

In the same way as with the transitive covariograms, we will distinguish a regular part (terms of even integer degree) and an irregular part (terms in r^{λ} with λ different from an even integer, and also logarithmic terms such as $r^{2n} \log r$). In the absence of an irregular part, the R.F. would be indefinitely differentiable, and thus perfectly regular. Hence it is only the irregular part which represents the degree of irregularity of the random function, and in the irregular part it is the term of lowest degree which plays the principal rôle : we can define the degree of irregularity of the R.F. as the degree λ of the principal irregular term. We shall now consider some complementary mathematical indications (but without going into details, as these are classical results).

e/ Continuity and differentiation in the mean square. The R.F. $Y(x)$ is said to be continuous in the mean square (continuous q.m.) if we have :

$$E\left([Y(x+h) - Y(x)]^2\right) \to 0 \qquad \text{when } |h| \to 0$$

This condition is satisfied (by definition) if and only if $\gamma(h)$ is continuous at $h = 0$, i.e. if there is no nugget effect.

In a one-dimensional space, we define the R.F. $Y'(x)$ as the derivative in the mean square sense (derivative q.m.) of the R.F. $Y(x)$ if

$$E\left(\left[\frac{Y(x+h) - Y(x)}{h} - Y'(x)\right]^2\right) \to 0 \quad \text{when } |h| \to 0$$

There are similar definitions for n-dimensional spaces ($n \neq 1$).

It can be shown that an intrinsic R.F. $Y(x)$ has a derivative q.m. $Y'(x)$ if and only if $\gamma(h)$ is twice differentiable at $h = 0$. The second derivative $\gamma''(h)$ then exists for any h, and $Y'(x)$ is stationary of second order (even if $Y(x)$ is only intrinsic) and has as a covariance the function $-\gamma''(h)$. In the same way, $Y(x)$ is differentiable n times q.m. if and only if the derivative $\gamma^{(2n)}$ exists at $h = 0$ (and then it exists for any h).

If λ is the degree of the principal irregular term of $\gamma(h)$, then $Y(x)$ has a derivative q.m. of order n if and only if $\lambda > 2n$ (if this irregular term is $r^{2n} \log r$, then $Y(x)$ is (n-1) times differentiable q.m. (but not n times).

2-3 REGULARIZATION OF AN INTRINSIC R.F.

In order to simplify this account, we will give the argument in the case of a stationary R.F. of second order which has a covariance function $K(h)$, but all the results obtained with the aid of the intrinsic function $\gamma(h) = K(o) - K(h)$ still remain valid in the case of an intrinsic R.F. (and hence even if $\gamma(h)$ is not bounded, and if consequently the covariance $K(h)$ does not exist) : this follows simply from the calculation mechanism seen in paragraph 2-2-1. (The following results are still classical ones).

2-3-1 Stochastic Integral. $I = \int_V Y(x)\, p(x)\, dx$

The integral I is defined as the limit in the mean square (l.q.m.) (if it does

exist) of the discrete sums :

$$I_n = \sum_{i=1}^{n} Y(x_i)\ p(x_i)\ \Delta\ x_i$$

(the $\Delta\ x_i$ are small separate elements of volume, the reunion of which is v, and x_i is a point belonging to $\Delta\ x_i$). The stochastic integral I is thus a random variable, like the I_n .

It can be demonstrated that this R.V. exists if and only if the integral

$$D^2(I) = \int_v p(x)\ dx \int_v K(x-y)p(y)\ dy \tag{2-9}$$

is finite, and $D^2(I)$ is then the (finite) variance of this r.v. (2-9) is easily shown again by the following calculation (not rigourous) :

$$I^2 = \int_v Y(x)\ p(x)\ dx \int_v Y(y)\ p(y)\ dy$$

$$E(I^2) = \int_v p(x)\ dx \int_v E[Y(x)\ Y(y)]\ p(y)\ dy$$

$$= \int_v p(x)\ dx \int_v K(x-y)\ p(y)\ dy$$

(this calculation is not rigourous as it is not quite evident at the start whether the exchange of the symbols $\int$ and E is justified : in fact, it can be shown that this exchange is legitimate).

2-3-2 Stochastic Convolution.

The convolution Y * f of the R.F. Y(x) by an ordinary function Y(x) is the stochastic integral (if it does exist)

$$\int Y(x-x')\ f(x')\ dx'$$

In this way the regularization $Y_p = Y * \check{p}$ of the R.F. Y(x) by the weighting function p(x) can be defined. It is the"weighted moving average" :

$$Y_p(x) = \int Y(x+x')\ p(x')\ dx'$$

Caution : the regularization $Y_p(x)$ is <u>still a random function</u> (though more regular than $Y(x)$, it is still random) : it is not enough to smooth or regularize a R.F. by a moving average process to make the random character of this function disappear as if by magic.

The <u>variance</u> of the regularization Y_p is given by formula (2-9) above, and Y_p exists in the sense of q.m. integration if and only if $D^2(I) < \infty$. Let us compute the covariance :

$$\mathcal{L}[Y(x_o)\ Y(x_o+h)] \text{ of } Y(x_o) \text{ and } Y(x_o+h)$$

We get :

$$Y_p(x_o)\ Y_p(x_o+h) = \int\int p(x')\ p(x'')\ Y(x_o+x')\ Y(x_o+h+x'')\ dx'\ dx''$$

Passing on to expectations, by exchanging E and $\int$, we get :

$$E[Y_p(x_o)\ Y_p(x_o+h)] = \int\int K(h+x''-x')\ p(x')\ p(x'')\ dx'\ dx''$$

This covariance does not depend on x_o, but only on h. Hence the <u>regularization</u> Y_p is <u>stationary of second order and has the covariance</u> :

$$K_p(h) = \int p(x)\ dx \int K(h+x-y)\ p(y)\ dy \tag{2-10}$$

This formula is a generalization of (2-9). To present it in a more synthetic form, let us make the change of variables $x = y + z$. We get :

$$K_p(h) = \int K(x+z)\ dz \int p(y+z)\ p(y)\ dy$$

Let P be the transitive covariogram ($P = p * \check{p}$) of the weighting function p. By definition we have $P(z) = \int p(y+z)\ p(y)\ d(y)$, hence :

$$K_p(h) = \int K(h+z)\ P(z)\ dz$$

i.e., in the form of a convolution (for $P = \check{P}$)

$$K_p = K * P = K * \check{P} \tag{2-11}$$

We thus obtain the covariance K_p of the regularization Y_p by regularizing the covariance K(h) of Y by the transitive covariogram of the weighting function. Compare this result with that obtained for the transitive methods.

The <u>semi-variogram</u> γ_p of the regularization Y_p is $K_p(0) - K_p(h)$. By replacing K(h) by K(0) - $\gamma(h)$ in (2-9) and (2-10), we notice that K(0) is eliminated and we get :

$$\gamma_p(h) = \int \gamma(h+z)\ P(z)\ dz - \int \gamma(z)\ P(z)\ dz \qquad (2\text{-}12)$$

which can also be written as :

$$\gamma_p = \gamma * P - A$$

with a constant A determined by the condition $\gamma_p(o) = 0$. These relations remain valid for any intrinsic R.F. (even if the covariance does not exist) (see note above).

<u>2-3-3 Grading</u>.

Grading is a particular case of regularization. In the transitive methods we could integrate the Re. V. from $-\infty$ to $+\infty$ with respect to one of the coordinates, x_3 for example, as this Re. V. vanished outside a bounded field. In this case, we can only integrate over a finite length ℓ. Hence, we shall call <u>grading over a constant thickness ℓ</u> the operation which allows us to pass from the R.F. $Y_3(x_1,x_2,x_3)$ in three dimensions to the R.F. $Y_2(x_1,x_2)$ defined in two dimensions, i.e. :

$$Y_2(x_1,x_2) = \frac{1}{\ell} \int_0^{\ell} Y(x_1,x_2,x_3)\ dx_3$$

If $Y_3(x)$ is a punctual assay in a stratiform formation of thickness ℓ, $Y_2(x_1,x_2)$ is the mean grade of the borehole at the geographical point (x_1,x_2) .

It is not very difficult to establish the formula giving the variogram $\gamma_2(h_1,h_2)$ of the R.F. $Y_2(x_1,x_2)$ - which is obviously intrinsic like $Y_3(x)$ - as a function of the variogram γ_3 of Y_3 (this can be done by the reader as an exercise). In the case when $\gamma_3 = \gamma_3(r)$ is an isotropic function, the following algorithm (straight grading over constant thickness in the isotropic case) is

obtained :

$$\gamma_{n-1}(r) = \frac{2}{\ell^2} \int_0^{\ell} (\ell - x)\ \gamma_n\left(\sqrt{r^2+x^2}\right) dx - \frac{2}{\ell^2} \int_0^{\ell} (\ell - x)\ \gamma_n(x)\ dx$$

which is in fact a simple particular case of (2-12).

As in the transitive methods, there is here also a <u>term by term correspondence rule</u> between the <u>irregular parts</u> of $\gamma_n(r)$ and $\gamma_{n-1}(r)$. These rules, somewhat less simple than in the transitive case, are written as :

$$(2\text{-}13) \qquad \begin{cases} r^{\lambda} \rightarrow A_{\lambda} \dfrac{r^{1+\lambda}}{\ell} + A'_{\lambda} \dfrac{r^{2+\lambda}}{\ell^2} \\ r^{2n} \log r \rightarrow A_{2n} \dfrac{r^{2n+1}}{\ell} + A'_{2n} \dfrac{r^{2n+2}}{\ell^2} \log r \\ r^{2n+1} \rightarrow A_{2n+1} \dfrac{r^{2n+2}}{\ell} \log r + A'_{2n+2} \dfrac{r^{2n+3}}{\ell^2} \end{cases}$$

The coefficients A_{λ}, A_{2n}, A_{2n+1} are <u>the same</u> as those which appear in the transitive methods (formulae (1-10) to (1-13)) and this is the first demonstration of the close affinity between the two aspects (transitive and probabilistic) of the theory. However, notice that the rules (2-13) are less simple than (1-10). Instead of $r^{1+\lambda}$, we get $\frac{r^{1+\lambda}}{\ell}$, but the coefficient $\frac{1}{\ell}$ is due to obvious dimensional reasons (in the intrinsic theory, we deal with mean grades rather than with accumulations as in the transitive case). But above all, besides the principal term $A_{\lambda} \frac{r^{1+\lambda}}{\ell}$ - which the exact equivalent of the transitive term - there is a complementary term $A'_{\lambda} \frac{r^{2+\lambda}}{\ell^2}$, whose coefficient A'_{λ} needs not be given explicitly (cf. [4]). This complementary term comes from the finite character (grading under finite thickness) of the integration domain or, better, from the presence of a term in $x\ \gamma(\sqrt{r^2+x^2})$ under the integration symbol of the straight grading algorithm. However, this complementary term is in $r^{2+\lambda}$, and thus of higher order than the principal one, and the <u>principal part</u> of $\gamma_{n-1}(r)$ near $r = 0$ is thus the same as would be given by the transitive rule.

Note also that the domain of validity of the expansion of $\gamma_{n-1}(r)$ obtained by rule (2-13) is necessarily within the sphere of radius ℓ . <u>At large distances</u>, it follows quite simply, from the straight grading algorithm that

$$\gamma_{n-1}(r) \neq \gamma_n(r) - \frac{2}{\ell^2} \int_0^{\ell} (\ell - x)\ \gamma_n(x) \qquad (r \gg \ell)$$

so that the variogram deduced by grading is the same as the initial one, apart from a constant.

2-4 EXTENSION AND ESTIMATION VARIANCES

2-4-1 Extension variance.

First of all we shall define the fundamental idea of extension variance. Let $Y(x)$ be a R.F. which we will for the moment suppose to be stationary of order 2, and let $K(h)$ be its covariance. Let us call $Z(v)$ and $Z(v')$ the "average grades" of the two domains v and v' in an n-dimensional space, i.e. the stochastic integrals :

$$Z(v) = \frac{1}{v}\int_v Y(x)\,dx \quad ; \quad Z(v') = \frac{1}{v'}\int_{v'} Y(x)\,dx$$

Formula (2-9) shows that, if v and v' are bounded, $Z(v)$ and $Z(v')$ have finite variances. The first one, for instance, can be expressed as :

$$\sigma^2(v) = \frac{1}{v^2}\int_v dx \int_v K(x-y)\,dy$$

In the same way, we can calculate the covariance $\sigma(v,v')$ of $Z(v)$ and $Z(v')$. From

$$Z(v)\,Z(v') = \frac{1}{vv'}\int_v Y(x)\,dx \int_{v'} Y(y)\,dy$$

we deduce by passing on to expectations that

$$\sigma(v,v') = \frac{1}{vv'}\int_v dx \int_{v'} E[Y(x)\,Y(y)]dy = \frac{1}{vv'}\int_v dx \int_{v'} K(x-y)\,dy$$

We will call the extension variance of v to v' (or of v' to v) the variance of the error $Z(v') - Z(v)$ that we make by attributing to v' the average grade $Z(v)$ of v. This extension variance σ_E^2 is equal to :

$$\sigma_E^2 = \sigma^2(v) + \sigma^2(v') - 2\,\sigma(v,v')$$

Taking into account the values calculated above, we get :

$$\sigma_E^2 = \frac{1}{v^2} \int_v dx \int_v K(x-y)\, dy + \frac{1}{v'^2} \int_{v'} dx \int_{v'} K(x-y)\, dy .$$

$$- \frac{2}{vv'} \int_v dx \int_{v'} K(x-y)\, dy$$

Replacing K(h) by K(o) - γ(h), we see that the constant K(o) disappears from the expression for σ_E^2 , and we get the following fundamental formula

$$(2\text{-}14) \qquad \sigma_E^2 = \frac{2}{vv'} \int_v dx \int_{v'} \gamma(x-y)\, dy - \frac{1}{v^2} \int_v dx \int_v \gamma(x-y)\, dy - \frac{1}{v'^2} \int_{v'} dx \int_{v'} \gamma(x-y)\, dy$$

It can be shown that (2-14) remains valid for any intrinsic R.F., even if the covariance K(h) does not exist (as a consequence of paragraph 2-2-1).

2-4-2 Estimation variance.

Suppose now that instead of knowing the "average grade" Z(v') of Y(x) in a volume v', we know the average grade

$$Z' = \frac{1}{N} \sum_{i=1}^{N} Y(x_i)$$

of N samples drawn at the N points x_i. Z' is a random variable, the characteristics of which are easily deduced from K(h) or γ(h). We shall call the estimation variance σ_N^2 (of v by the N samples drawn at the N points x_i) the variance of the difference [Z(v) - Z'] . To get the expression for σ_N^2 , we have, at each step of the argument which led us to (2-14), to replace the integrals over v' by discrete sums on the N points x. Thus we get the second fundamental formula :

$$(2\text{-}15) \qquad \sigma_N^2 = \frac{2}{Nv} \sum_i \int_v \gamma(x_i-x)\, dx - \frac{1}{} \int_v dx \int_v \gamma(x-y)\, dy - \frac{1}{N^2} \sum_i \sum_j \gamma(x_i-x_j)$$

This formula, in which exact and approximate expressions for the same integrals alternate, has a remarkable structure, analogous, although more complex, to that of formula (1-14) in the transitive methods. Notice, in particular, that the estimation variance decreases as :

- the sample grid is more closely spaced and more representative of the geometry of the volume v we wish to estimate,

- the function $\gamma(h)$ is more regular (because the R.F. Y(x) itself is more regular in its variations in space).

In practice however, if N is great formula (2-15) leads to tiring calculations. We shall give later on simpler approximation formulae which will also allow a methodologically instructive comparison with the similar formulae of the transitive methods.

Remark - There is no conceptual difference between the notions of extension variances and estimation variances. Formula (2-15) is a particular case of (2-14), v' being replaced by the union of the N points x_i. Practice has reserved the term of extension variance to the extension of a single sample in its "influence zone", and that of estimation variance to the extension of a greater number of samples in the whole ore-body, or in a large panel.

2-4-3 Variance of v within V.

The notion of variance $\sigma^2(v/V)$ of a sample v within a field V seems at first sight experimentally evident. Nevertheless it has a precise sense only when the volume V appears as the reunion $V = \bigcup v_i$ of separate volumes v_i, each equal to v, and deduced from each other by translation. When n samples v_i have been taken in any volume V' and when their assays Y_i are known, what is numerically computed

$$\bar{Y} = \frac{1}{n} \sum_i Y_i$$

$$s^2 = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2$$

represents an estimation of the variance of v within $V = \bigcup v_i$ (reunion of the n volumes v_i of the actual samples) and not in the domain V' as we could be inclined to believe. These experimental results lead to the two following definitions :

a/ First, examine the case when the samples v_i are reduced to points (and when of course the relationship $V = \bigcup v_i$ is true as V is clearly the reunion of the punctual sets it contains) ; the variance $\sigma^2(0/V)$ of the punctual samples within V is then defined. Let

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

be the average grade of V. For a given realization Y(x), Z(V) is an ordinary integral, and the experimental variance of the punctual sample assays is

$$s^2(0/V) = \frac{1}{V} \int_V [Y(x) - Z(V)]^2 \, dx \tag{2-16}$$

This relationship can also be interpreted by saying that $s^2(0/V)$ is the conditional variance (for a given realization) of the random variable Y(x) obtained when setting a point x at random within V with a uniform probability density. We will call variance $\sigma^2(v/V)$ of v within V the variance of the random variable obtained when deconditioning the random variable Y(x) with respect to the realization, that is to say the variance of the random variable Y = Y(x) where this time Y(x) is the R.F. itself (and no longer a realization of it), and x any point of V. For a given realization, the conditional expectation of Y(x) is evidently Z(V). We get thus

$$\sigma^2(0/V) = E[s^2(0/V)] = \frac{1}{V} \int_V E[Y(x) - Z(V)]^2 \, dx$$

But $E[Y(x) - Z(V)]^2$ is also the extension variance to V of the punctual sample set at x. Thus, the variance $\sigma^2(0/V)$ is the mean value when $x \in V$ of the extension variance to V of the punctual sample set at x. The expression of this variance is then easily deduced from (2-14) :

$$\sigma^2(0/V) = \frac{1}{V^2} \int_V dx \int_V \gamma(x-y) \, dy \tag{2-17}$$

It is the mean value of $\gamma(x-y)$ when x and y sweep V separately.

b/ When V is the reunion of N separate volumes v_i, with assays Z_i, the experimental variance

$$s^2(v/V) = \frac{1}{N} \sum_{i=1}^{N} [Z_i - Z(V)]^2$$

is the variance of the finite population composed of the N values Z_i (which are numbers, for a given realization) or, otherwise, the variance of the random variable Z_i obtained when drawing one of the N indices i at random. Deconditioning with respect to the realization, we obtain a random variable Z_i, where the index is drawn at random as above, and where Z_i is now a stochastic integral (and no longer a numerical one)· The variance of this deconditioned variable is (by definition) the variance $\sigma^2(v/V)$ of v within V. As previously, we find that it is equal to the mean value with respect to i

$$\sigma^2(v/V) = \frac{1}{N} \sum_i E[Z_i - Z(V)]^2$$

of the extension variance of v_i to V. Formula (2-14) leads easily to :

$$\sigma^2(v/V) = \frac{1}{V^2} \int_V dx \int_V \gamma(x-y)\, dy - \frac{1}{v^2} \int_v dx \int_v \gamma(x-y)\, dy \tag{2-18}$$

In particular

$$\sigma^2(v/V) = \sigma^2(0/V) - \sigma^2(0/v) \tag{2-19}$$

c/ In the general case when v and V are any volumes at all, not necessarily geometrically compatible, we define the variance by the same formula (2-19) : this quantity is thus only a simple artifice of calculation. It can even take negative values : in particular, we always have $\sigma^2(v/V) = -\sigma^2(V/v)$.

d/ Additivity relationship . Let v, V and V' be three volumes with for instance $v \subset V \subset V'$. From (2-19) we get

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

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

and by difference

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

or

$$\sigma^2(v/V') = \sigma^2(v/V) + \sigma^2(V/V') \tag{2-20}$$

The variance of the sample v within the field V' is equal to the sum of the variances of v within the panel V and of the panel V in the field V'.

e/ Covariance of v and v' within V. We can define in an analogous manner the covariance $\sigma(v, v'/V)$ of two samples v and v' (the distance and mutual disposition of which remain fixed) within the field V. We get :

$$\sigma(v,v'/V) = \frac{1}{V^2} \int_V dx \int_V \gamma(x-y)\, dy - \frac{1}{vv'} \int_v dx \int_{v'} \gamma(x-y)\, dy$$

f/ It is often convenient to express the estimation variance (2-15) or the extension variance (2-14) by means of the variances and covariances of the samples within an arbitrary field V. For example we get identically

$$\sigma_E^2 = \sigma^2(v/V) + \sigma^2(v'/V) - 2\,\sigma(v,v'/V)$$

and as can be seen by substituting the expressions for the variances and covariances within V : the term $\frac{1}{V^2} \int_V \int_V$ disappears and the three remaining terms give the second member of (2-14). Hence we get a perhaps more intuitive expression of the variance.

2-4-4 Application : Random or stratified random networks.

The case of the regular network, which is by far the most difficult, will be dealt with later on and we will analyze first the case of two usual types of random networks.

a/ Purely random network. To estimate the grade Z(v) of a volume V, we have the values $Y(x_i)$ of the R.F. at N points x_i located "anywhere" within V. We will assume that each x_i has been placed at random within V with a uniform probability density, and independently from the other samples. We can get the estimation variance σ_N^2 by integrating (2-15) over V with respect to each x_i. This is an easy calculation, but it is even simpler to notice that the partial errors $Y(x_i) - Z(V)$ are independent of each other for a given realization and have the same variance $s^2(0/V)$: the resulting error, which is $\frac{1}{N} \sum [Y(x_i) - Z(V)]^2$, has thus the variance :

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

By deconditioning with respect to the realization, we find that the estimation variance in the case of a purely random network, is equal to the variance of a sample within the field V divided by the number of samples N, that is :

$$\sigma_N^2 = \frac{1}{N}\, \sigma^2(0/V)$$

b/ Stratified random network. The volume V to be estimated is in this case divided into N equal and separate zones of influence v_i . Within each v_i a sample is drawn at a point x_i chosen at random within the influence zone v_i with a uniform probability density, and independently of the other samples. We can calculate σ_N^2 by integrating (2-15) with respect to x_i within v_i for each of the v_i's, but it is easier to notice that the total error is

$$\frac{1}{N} \sum_i [Y(x_i) - Z(v_i)]$$

and that each partial error $[Y(x_i) - Z(v_i)]$ is independent of the others and has a variance $\sigma^2(0/v)$. Reasoning as above, first on a given realization and then by deconditioning, it follows that :

$$\sigma_N^2 = \frac{1}{N} \sigma^2(0/v)$$

In the case of a stratified random network, the estimation variance is thus equal to the variance of a sample within its influence zone, divided by the number N of samples.

Remark - It is obvious that the stratified random network always gives better results that the purely random network. Indeed from (2-20) we get :

$$\frac{1}{N} [\sigma^2(0/V) - \sigma^2(0/v)] = \frac{1}{N} \sigma^2(v/V) \geq 0$$

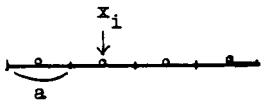
2-5 APPROXIMATION METHODS IN ONE DIMENSION.

Since the general formula (2-15) is rather cumbersome to manipulate numerically once the number N of samples becomes great, it will be better if we derive some approximation methods which will simplify the calculation of estimation variances. Moreover, we will see when we do this that the close relationship with the corresponding formulae obtained for the transitive methods will reappear, relationship whose methodological importance was underlined in the section on grading. We shall consider initially a one-dimensional space.

2-5-1 The correspondence principle.

Let L = Na be a segment composed of n contiguous segments of the same length a,

at each centre x_i of which a punctual sample has been taken. In order to estimate the value



$$Z(L) = \frac{1}{L}\int_L Y(x)\,dx$$

we construct the estimator :

$$Z^*(L) = \frac{1}{N}\sum_i Y(x_i)$$

The corresponding estimation variance is given by the general formula (2-15). Performing computations similar to those already presented in the transitive methods, we come to a <u>term by term correspondence principle</u> between <u>the irregular part of the semi-variogram</u> $\gamma(h)$ and the expansion of <u>the estimation variance near a = 0</u>. This principle may be expressed by the following rules :

$$(2\text{-}21)\qquad \begin{cases} |h|^\lambda \rightarrow -T_\lambda \dfrac{a^\lambda}{N} + T'_\lambda \dfrac{a^\lambda}{N^2} & (\lambda > 0 \text{ and not an even integer}) \\ \\ |h|^{2n} \log |h| \rightarrow -T_{2n} \dfrac{a^{2n}}{N} \end{cases}$$

When λ is an even integer, T_λ is equal to 0 (but not T'_λ). The coefficients T_λ have the <u>same value</u> as those appearing in the rule (1-18) for transitive methods.

As in the case of grading, a supplementary term of higher order (in $1/N^2$) appears, which here also is linked with the <u>finite character</u> of the operations in the intrinsic case (mean value of a finite number N of samples, and not, as in the transitive case, an integral theoretically ranging from $-\infty$ to $+\infty$). In practice, this supplementary term is negligible when N is not very small.

The rules (2-21) allow us to obtain an approximate value of the estimation variance for very small spacings of a. This expression is valid only provided that a is inferior to the limit r_o, beyond which the expansion of $\gamma(h)$ itself can no longer be used. For any value of a greater than r_o, we have to introduce a second principle of approximation :

<u>2-5-2 Principle of composition of elementary extension variances</u>.

a/ <u>Auxiliary intrinsic functions</u>. In addition to $\gamma(h)$, the following functions are constantly used in practice :

$$X(h) = \frac{1}{h} \int_0^h \gamma(x)\, dx$$

$$F(h) = \frac{2}{h^2} \int_0^h x\, X(x)\, dx = \frac{2}{h^2} \int_0^h (h-x)\, \gamma(x)\, dx$$

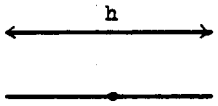
F(h) is the mean value of $\gamma(h')$ when the two extremities of h' sweep the segment (0,h). Accordingly, the variance of the segment h within the segment L is

$$\sigma^2(h/L) = F(L) - F(h)$$

Likewise, the covariances are calculated from X(h). The covariance within L of the segment h with one of its (point) extremities is

$$\sigma(0,h/L) = F(L) - X(h)$$

b/ Elementary extension variance. This is the extension variance to the segment h of the sample set at the center of this segment. It is given by :



(2-22) $$\sigma_E^2 = 2\, X\, (\tfrac{h}{2}) - F(h)$$

Equation (2-22) is deduced from (2-15) and the two relations established in a/ .

c/ Elementary extension variances combination principle. Consider the problem of estimating the grade of the segment L = na divided into n zones of influence of length a, at the centre of each of which a sample has been taken. Let Y_i be the grade of the sample i, Z_i that of its zone of influence, and $Z = \frac{1}{n} \sum Z_i$ that of L. The total error is the mean

$$\frac{1}{N} \sum_i (Y_i - Z_i)$$

of the partial errors $Y_i - Z_i$. The apprcximation principle used is to assume that these partial errors are independent of each other (this principle is verified to a very reasonable degree of approximation for the usual types of $\gamma(h)$). Since the variance of $Y_i - Z_i$ is exactly the same as the elementary extension variance calculated in (2-22), it follows that

(2-23) $$\sigma_n^2 = \frac{1}{n}\sigma_E^2 = \frac{1}{n}\left[2\,\bar{X}\left(\frac{a}{2}\right) - F(a)\right]$$

The estimation variance is thus obtained by dividing the elementary extension variance of each sample within its zone of influence by the number of samples n.

d/ The case of a closed arrangement. Starting from n+1 samples taken along a regular grid, the segment L = na made up of the portion between the first and last sample is to be estimated. It can be shown that this closed arrangement is equivalent to that analysed in c), and the estimation variance is the same as that given in (2-23) (provided that n is greater than 1). Hence the closed arrangement of n+1 samples is equivalent to the centralized arrangement of n samples.

Remark - For n = 1, the closed arrangement of two samples leads to the estimation variance

$$\sigma_{E_1}^2 = 2\,\bar{X}(h) - F(h) - \frac{1}{2}\gamma(h)$$

which differs from the elementary extension variance (2-23). We are not justified in dividing $\sigma_{E_1}^2$ by n, as the partial errors arising in the estimation of the two consecutive closed segments

are not at all independent (these two figures, in particular, have a common central sample).

e/ Comparison of the two approximation principles. When a is small, a correspondence principle similar to (2-21) can be deduced from (2-23), whose coefficients are not numerically very different from the T_λ (at least if λ is not too great) so that the rules (2-21) and (2-23) are practically equivalent. For simplicity, the rule (2-23) is always used in practice, as it remains valid in almost all cases, i.e. for large and small sample intervals.

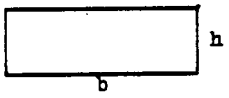
2-6 APPROXIMATION METHOD IN $\mathbb{R}^n$.

The case of a two-dimensional space will now be studied in some detail. It may be generalized, by analogy, without difficulty. The approximation principle used will be that of the principle of composition of the line and slice terms studied in detail in paragraph 1-4-4. An equiva-

lent justification to that given for the transitive methods could be presented, but the calculations are far more complicated.

We shall suppose that the semi-variogram is isotropic (i.e. it only depends on r). In the case of a geometric anisotropy, it is easy to reduce the situation to this case.

a/ The auxiliary functions. As in the case above, it is convenient to introduce the following functions :



$\gamma_b(h)$: mean of $\gamma(x-y)$ when x and y describe the two parallel sides of length b of the rectangle b × h (except for a constant, $\gamma_b(h)$ can be derived from $\gamma(h)$ by grading of order 1).

$X_b(h)$: mean of $\gamma(x-y)$, x describing one of the sides b and y the rectangle itself. Then

$$X_b(h) = \frac{1}{h} \int_0^h \gamma_b(x)\, dx$$

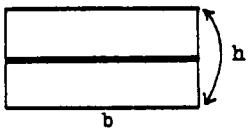
F(b,h) : mean of $\gamma(x-y)$ when x and y sweep the rectangle. This function is symmetrical in b and h, and proves to be

$$F(b,h) = \frac{2}{h^2} \int_0^h x\, X_b(x)\, dx = \frac{2}{h^2} \int_0^h (h-x)\, \gamma_b(x)\, dx$$

Q(b,h) : mean of $\gamma(x-y)$ when x describes b and y describes h or, instead, x sweeps the rectangle while y remains at one of its corners.

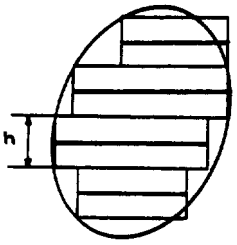
b/ Extension of a median segment to its influence zone.

(2-24) $$\sigma_E^2 = 2\, X_b\left(\frac{h}{2}\right) - F(h,b) - \gamma_b(0)$$



c/ Estimation of S by equidistant parallel drives.

Let $b_1, b_2, \ldots, b_n$ be the lengths of the n drives, h being their equidistance. The surface S to be estimated is similar to that of the reunion of the influence rectangles of the n drives. Formula (2-24) gives the extension variance of b_i to its rectangle of influence, $\sigma_{E_i}^2$. If Y is the



grade of b_i, Z_i that of its zone of influence, then it is assumed that the $(Y_i - Z_i)$ are independent, and that the total error

$$\frac{\Sigma\, b_i\, (Y_i - Z_i)}{\Sigma\, b_i}$$

leads to a variance which is calculated by weighting the extension variances $\sigma^2_{E_i}$ by the squares of the b_i . In this way, the estimation variance is

(2-25) $$\sigma^2_E = \frac{\Sigma\, b_i^2\, \sigma^2_{E_i}}{(\Sigma\, b_i)^2}$$

NOTE : If the b values are equal, this leads simply to $\sigma^2_E = \frac{1}{n}\, \sigma^2_{E_i}$.

d/ Composition of a line term and a slice term.

In the preceding case c/, it very often happens that the real grades of the drives are not known perfectly, but are only estimated from a regular grid of samples with a spacing of a (a < h). Hence the errors which arise in estimating the lines with the samples, and S itself from the lines (supposed to be known) are independent of each other. The estimation variance is then :

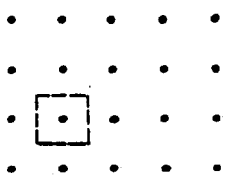
(2-26) $$\sigma^2_E = \frac{1}{N}\, \sigma^2(a) + \frac{\Sigma_i\, b_i^2\, \sigma^2_{E_i}}{(\Sigma\, b_i)^2}$$

$\sigma^2(a)$ is the extension variance (2-23) of a point sample within its segment of influence, and N is the total number of samples. $\frac{1}{N}\, \sigma^2(a)$ is the line term, the variance of the error made, in estimating the lines by the samples.

The second term, given in (2-25) is the slice term, the variance of the error made in extending the grade of the lines to their slices of influence.

This composition principle (2-26) is valid provided that a is less than h. It is particularly applicable to the case of a reconnaissance on a rectangular grid pattern a, h.

e/ Case of a square grid. The extension variance of a borehole to its square of influence, at the centre of which it is located, is :



(2-27) $$\sigma_E^2 = 2\,Q\left(\frac{a}{2}, \frac{a}{2}\right) - F(a,a)$$

For a square grid, it can be shown that the errors made in estimating each square by the sample set at its centre are independent. Hence the estimation variance of N samples is

$$\sigma_N^2 = \frac{1}{N}\,\sigma_E^2$$

with σ_E^2 given by (2-27).

f/ Graphs. For each isotropic scheme, i.e. for each $\gamma(r)$, results are given in the form of graphs :

1/ The function F(a,b) defined in a/ which enables the variance of a sample within the rectangle a, b to be calculated.

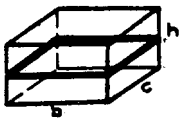
2/ The extension variance (2-24) of a median of length b to the rectangle b,h - and in particular the case where b = 0, in which case the elementary extension variance is obtained once again

3/ The extension variance (2-27) to the square a,a. These three graphs allow any estimation variance to be calculated quickly.

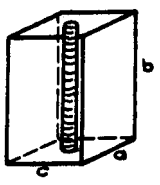
Case of the three dimensions. Exactly as above but now with three kinds of terms :

- line term : extension of punctual samples to lines.
- (slice or) section term : extension of lines to sections.
- block terms : extension of sections to blocks.

The following graphs (in addition to those above) have to be drawn up :



- extension variance of a median plane b,c to its right-angled parallelepiped of influence.



- extension of a borehole of length b to its right prism of influence a, b, c. This function F(a,b,c,) will be used to calculate the variance of a sample within this parallelepiped.

2-7 THE NUGGET EFFECT.

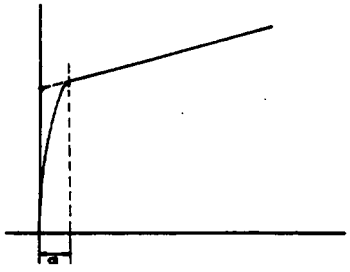
2-7-1 Genesis of the nugget effect.

A semi-variogram of finite range characterizes what is called a transition phenomenon ; beyond the distance a, independence is reached, and tne range gives the scale of the elementary structures of the corresponding regionalized phenomenon. There is often in addition the superimposition of several structures of quite different scales, intermeshed with each other. The experimental variogram then exhibits a succession of thresholds and levels, the analysis of which allows the reconstitution of the hierarchy of these "gigogne structures". The notion of scale plays a fundamental part here. At the scale of a dozen or a hundred meters, a transition phenomenon which has, for instance, a range of the order of centimeters is no longer apparent on the experimental $\gamma(h)$, except as a discontinuity at the origin, or "nugget effect". In a general way, all nugget effects are reflections of a transition structure, the dimensions fo which are considerably exceeded by the working scale : the details and the characteristics of this prior structure have long since ceased to be perceptible, and the larger scale has barely preserved a single parameter - the nugget constant - which gives a kind of overall undifferentiated measure of the "intensity" of this hidden structure.

To analyze the genesis of a nugget effect, start at the point level and imagine that on a primary structure of dimension a, a macro-regionalization - that is a secondary structure of much larger dimensions - is superimposed. If the primary structure only existed, the corresponding Re. V. could be described as a realization of a R.F. having a covariance C(h) of range a, or a semi-variogram :

$$\gamma_1(h) = C - C(h)$$

of range a, and such that $\gamma_1(\infty) = C - C(0)$. To take the macro-regionalization into account, we must add a second component $\gamma_2(h)$ which represents the secondary structure and which varies only very slowly at the scale of the primary structure :



$$\gamma(h) = C - C(h) + \gamma_2(h)$$

This $\gamma(h)$ exhibits near the origin a zone of very rapid increase, the dimension of which is of the same order as a. At the scale of the macro-structure, this $\gamma(h)$ hence exhibits a nugget effect of

amplitude C.

2-7-2 Macroscopic influence of the nugget effect.

Let us examine the consequences of this nugget effect. At first, at the macroscopic level, the samples taken will no longer be points, but their volume will be quite large in regard to a. We shall determine the semi-variogram $\gamma_v(h)$ of these samples v (the only one experimentally accessible). γ_v is the sum of the very continuous component γ_2 (which has not been sensibly altered by this regularization) and of what we obtain by applying formula (2-12) to the component $C - C(h)$. Therefore this nugget component is :

$$\gamma_p(h) = \frac{1}{v^2} \int_v dx \int_v C(x-y)\, dy - \frac{1}{v^2} \int_v dx \int_v C(h+x-y)\, dy$$

We shall now consider γ_p . We know that $\gamma_p(o) = 0$, but when h exceeds the dimensions of the volume v then $|h+x-y| \geq a$ if $x \in v$ and $y \in v$, and $C(h+x-y) = 0$. Thus only the first term remains as long as h is not very small, and a discontinuity is observed experimentally at the origin, the value σ_p^2 of which (the nugget variance) is

$$\sigma_p^2 = \frac{1}{v^2} \int_v dx \int_v C(x-y)\, dy$$

On the other hand, by hypothesis the dimensions of v are large in respect to a, and

$$\int_v C(x-y)\, dy = C(h)\, dh$$

(integral extended through the whole space) except when x is at a distance less than a from the boundary of v. But these points occupy a negligible volume, and their influence, which is of the order of a, can be disregarded. Thus there remains

$$\sigma_p^2 = \frac{1}{v^2} \int_v dx \quad C(h)\, dh = \frac{1}{v} \int C(h)\, dh$$

The nugget constant σ_p^2 observed experimentally is <u>in inverse ratio to the volume</u> of samples:

$$\sigma_p^2 = \frac{A}{v}$$

and the coefficient $A = \int C(h)\, dh$, which is the integral of the covariance C(h) of the micro-structures, is the only indication of them remaining at the scale of the volumes v.

In the same way the nugget effect increases the variance of v within V, and the extension and estimation variances. For $\sigma^2(v/V)$ for instance, the contribution of the nugget effect will be

$$\frac{1}{v^2} \int_v dx \int_v C(x-y)\, dy - \frac{1}{V^2} \int_V dx \int_V C(x-y)\, dy$$

i.e., from the previous calculation :

$$A \left(\frac{1}{v} - \frac{1}{V}\right)$$

In the estimation variance of V by N samples of size v, a similar nugget component is found, equal to :

$$A \left(\frac{1}{Nv} - \frac{1}{V}\right)$$

In every case, the variance due to the nugget effect is in inverse ratio to the volume of the samples. It is as if the Re. V. itself has two independent components, one very regular and corresponding to the semi-variogram γ_2, the other completely random and discontinuous and taking into account the nugget effect.

Representation of a nugget effect by a Dirac measure.

As it has been seen, the macroscopically observable effects of a nugget effect depend only on the constant :

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

and not on the exact form of the micro-covariance C(h). At the macroscopic level, the nugget effect can thus be symbolized by a Dirac measure $A\delta$ (mass A set at 0) replacing the function C(h). The exercises 16 to 20 on the transitive methods have made us familiar with this point of view. In the probabilistic version, the punctual variogram γ is replaced by $\gamma_1(h) - A\delta$ (with the sign - , as it concerns a variogram). The usual formulae immediately produce again the nugget term A/v observed at the macroscopic scale.

Strictly speaking, to justify the introduction of this Dirac measure into the expression of the variogram or of the punctual covariance, it would be necessary to refer to the random measures theory (and no longer to a R.F.), which we do not do here (cf. [4]). In fact, the Dirac measure is the covariance of a random measure corresponding, for example, to a Poissonian setting of points in an n-dimensional space. Some details about this will be found in the exercises (9 and 10).

2-8 THE DE WIJSIAN SCHEME.

The De Wijsian scheme is defined by the intrinsic function (isotropic)

$$\gamma(r) = 3\,\alpha \log r$$

The coefficient α is the absolute dispersion : it characterizes the dispersion of the pure state phenomenon, i.e. independently of all the influence from the geometry of the field or of the support. It will be noticed that log r tends to $-\infty$ when $r \to 0$, so that the function log r cannot, strictly speaking, be the variogram of an I.R.F. In fact, this logarithmic variogram (log r is a measure, and no longer a function) characterizes a random measure, and no longer a R.F. But the regularizations of this random measure are again R.F., and it can be shown that their variograms are deduced from the logarithmic function according to the rules valid for R.F. In practice, this means that the De Wijsian scheme can be used as if it were an I.R.F., provided that we never reason at the punctual level, but always on regularizations.

2-8-1 Linear equivalents.

Put :

$$F(v) = \frac{1}{v^2} \int_v dx \int_v \log |x-y|\, dy$$

(mean value of log |h| within the volume v : it is a function which depends not only on the measure of the volume v, but also on its form). We get from relation (2-18) :

$$\sigma^2(v/V) = 3\,\alpha[F(V) - F(v)] \tag{2-28}$$

In the same De Wijsian orebody, if two samples v and v' obey F(v) = F(v'), they will have the same variance in any panel. They will be said to be equivalent. If we take as a set.v a straight segment of length ℓ , we get

$$F(\ell) = \log \ell - \frac{3}{2}$$

(this can be demonstrated by means of the auxiliary functions of 2-5-2, a/). We will call the linear equivalent of a sample v the length of the segment ℓ equivalent to v, that is to say the length defined by

$$\log \ell - \frac{3}{2} = F(v)$$

In the same way, let L be the linear equivalent of the field V

$$\log L - \frac{3}{2} = F(v)$$

By difference we have

$$F(V) - F(v) = \log (L/\ell)$$

and (2-28) then gives :

(2-29) $$\sigma^2(v/V) = 3 \alpha \log (L/\ell)$$

This formula gives a quick way of calculating the variance. It is often used in the opposite way, to evaluate the absolute dispersion from the experimental variance of the samples within their field : if v and V are geometrically similar, we have $V/v = (L/\ell)^3$, and (2-28) is reduced to the De Wijs formula : $\sigma^2(v/V) = \alpha \log (V/v)$.

Approximate calculation of linear equivalents. - With excellent precision, the linear equivalent of the rectangle of sides a and b is a+b.

In the case of a right-angled parallelepiped (a,b,c), c referring to the smallest side, we get this time with a rather rough approximation, the linear equivalent as a+b + 0,7 c. For greater precision, refer to graph 1.

Application : behaviour of the semi-variogram of the samples v when h is large with respect to the dimensions of v (the De Wijsian scheme is never employed at the point level, for log r is infinite at r = 0, but only through regularizations[1]). Let γ_v be the semi-variogram of samples v distant h from each other. $2\,\gamma_v(h)$ is the extension variance of these two samples, and formula (2-14) leads to

$$\gamma_v(h) = \frac{3\alpha}{v^2} \int_v \int_v \log |h+x-y| \, dx \, dy - 3\,\alpha\, F(v)$$

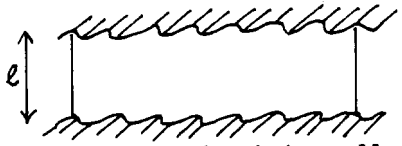
When h is greater than the dimensions of v, the first term is not very different from $3\,\alpha \log r$. The second term is $3\,\alpha(\log \ell - 3/2)$, ℓ being the linear equivalent of v. Then for h great enough in regard to v, we get :

(2-30) $$\gamma_v(r) = 3\,\alpha \log \frac{r}{\ell} + 9\,\frac{\alpha}{2}$$

This formula is very important in practical applications.

2-8-2 The two dimensional De Wijsian scheme.

In the preceding application, we can take v as a segment of length ℓ moving in a direction parallel to itself. After (2-30), we get :



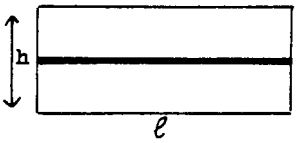
$$\gamma_\ell(h) = 3\,\alpha \log \frac{r}{\ell} + \frac{9\alpha}{2} \qquad \text{when } h \gg \ell$$

On the contrary, when h is small, the grading rule $\log r \to \pi\, r/\ell$ gives

(2-31) $$\gamma_\ell(h) = 3\,\alpha\,\pi\,\frac{h}{\ell}$$

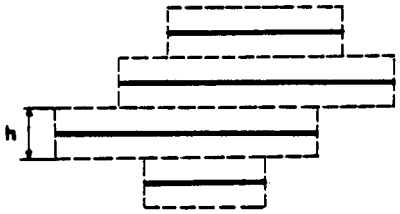
From the exact expression of $\gamma_\ell(h)$ (deduced from log r by grading) the extension variance ℓ to its rectangle of influence ℓ, h is computed : graph 2. When h is small compared with ℓ (in practice, it is enough that $h \leq \ell$), we can use (2-31) to calculate this extension variance directly, which leads to :

1 - At the point level, the De Wijsian scheme is no longer a function but a random distribution. But the regularizations of this scheme are again R.F's.



(2-32) $\qquad \sigma_E^2 = \frac{\pi}{2} \frac{h}{\ell}$

Application : Stratiform orebody developed by parallel drives.



a/ Let us first of all assume that the real grades of the drives are known and calculate the slice-term. If h is the equidistance of the levels, the extension variance $\sigma_{E_o}^2$ of the level of length ℓ_i to its influence slice can be read on graph 2. The slice term is then :

$$\sigma_E^2 = \frac{\Sigma\, \ell_i^2\, \sigma_{E_i}^2}{(\Sigma\, \ell_i)^2}$$

If each one of the levels has a length ℓ_i greater than h, $\sigma_{E_i}^2$ is given by (2-32) and, putting L = $= \Sigma\, \ell_i$ (total developed length), we get

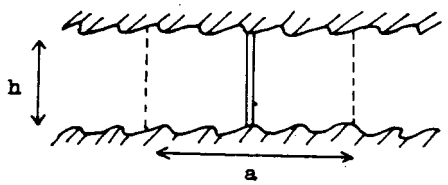
$$\sigma_E^2 = \alpha \frac{\pi}{2} \frac{\Sigma\, \ell_i^2 \frac{h}{i}}{L^2} = \alpha \frac{\pi}{2} \frac{h}{L}$$

or still with S = L h (total mineralized area)

$$\sigma_E^2 = \alpha \frac{\pi}{2} \frac{S}{L^2}$$

Consequence : an orebody four times bigger than another one, other things being equal, needs less than half the valuation cost per ton of ore for the same precision.

b/ If the mean values of the levels are estimated by means of channel samples regularly spaced at an interval a, a line term has to be added. When evaluating it, it is no longer possible to neglect the thickness p of the formation (p is small, by hypothesis, with regard to h or to the level lengths ℓ_i , but not with respect to a). Again, graph 2 gives the extension variance $\sigma^2(a)$ of a channel of length p to its rectangle of influence a,p. The line term is then :



$$\frac{1}{n}\, \sigma^2(a) = \frac{a}{L}\, \sigma^2(a)$$

Note : In most of statiform or lode orebodies, the third dimension, that of thickness, is not equivalent to the two others. There is an anisotropy. The length p which is to be assigned to the channel can thus be different from the real one. This equivalent thickness p is determined by noticing, from (2-30), that the experimental semi-variogram of the channels has the value

$$\gamma(h) = 3\,\alpha\,[\log(\tfrac{h}{p}) + \tfrac{3}{2}] \qquad \text{when } h > p$$

It is the value so determined which is to be used in the calculation of $\sigma^2(a)$.

A thin seam proved by boreholes on a square grid.- Let p be the equivalent thickness of the formation, defined as in the preceding note. We have to calculate the extension variance of a bore to its square (a,a) of influence. By assimilating this square to a circle having the same surface, we get (cf. Exercise 11) :

$$\sigma_E^2 = 3\,\alpha\,[\log\tfrac{2a}{p} - \tfrac{1}{2}]$$

We deduce the estimation variance by dividing by the number of boreholes n :

$$\sigma_n^2 = \tfrac{3\alpha}{n}\,[\log\tfrac{2a}{p} - \tfrac{1}{2}]$$

2-8-3 The De Wijsian scheme in the three-dimensional space.

Graph 1 gives the linear equivalent of a right-angled parallelepiped.

Estimation of mass. - The line term and the section term are calculated as above (from graph 2). For the block term, as the case may be, graph 3, 4 or 5 will be used (extension variances of large sections ab, medium sections ac and small sections bc, $a \geq b \geq c$).

Massive deposits proved by boreholes on a square grid. - The extension variance $\sigma_{E_i}^2$ of a bore of thickness h_i to its right prism h_i,a,a of influence is read from graph 6. We get then

$$\sigma_E^2 = \frac{\Sigma\, h_i^2\, \sigma_{E_i}^2}{(\Sigma\, h_i)^2}$$

(if the h_i are equal or nearly so, then : $\frac{1}{n}\sigma^2_{E_i}$).

2-9 THE SPHERICAL SCHEME

To represent a transition phenomenon, schemes of the following form can be used :

$$\gamma(h) = A[K(o) - K(h)]$$

where K is the geometric covariogram of the volume v (cf. Exercise 9, ch. 1). If we take as v the sphere of diameter a, we get (cf. ex. 9 of the transitive methods) :

$$K(h) = \begin{cases} \frac{\pi}{6} a^3 \left(1 - \frac{3}{2}\frac{h}{a} + \frac{1}{2}\frac{h^3}{a^3}\right) & (|h| < a) \\ 0 & (|h| \geq a) \end{cases}$$

The spherical scheme is thus defined by the semi-variogram :

$$\gamma(r) = \begin{cases} C\left(\frac{3}{2}\frac{r}{a} - \frac{1}{2}\frac{r^3}{a^3}\right) & \text{when } r < a \\ C & \text{when } r \geq a \end{cases}$$

The <u>range</u> is a, the <u>sill</u> is $C = \gamma(\infty)$, the <u>slope at the origin is</u> $\frac{3}{2}\frac{C}{a}$.

In practical applications, a and C are determined by successive approximations : we start from values a_1 and C_1 obtained by simple interpolation on the experimental semi-variogram ; then <u>graph 7</u> is used to calculate the theoretical variance of a sample in its field ; if this variance is, for example, greater than the experimantal variance of the samples in their field, then we can either lower C_1, or increase a_1. The solution is quickly reached.

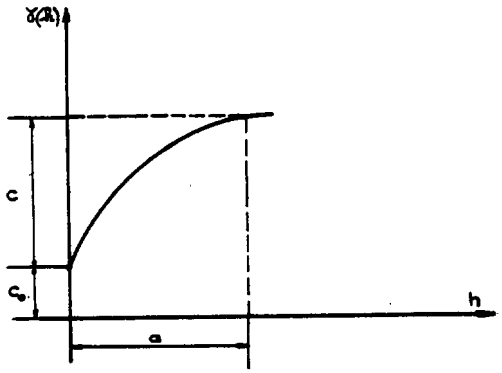
We have at our disposal only the graphs relative to the one or two-dimensional case :

<u>Graph 7</u> : variance of a point within a rectangle.

Graph 8 : élementary extension variance of a point to its segment of influence (calculation of the line term) ; extension variance of a point to its square of influence (square grid patterns).

Graph 9 : extension variance of a median segment to its rectangle of influence (calculation of the slice term).

Remark - In practice, transition phenomena are very often accompanied by a nugget effect : in this case they can be represented by a spherical scheme with nugget effect C_o :

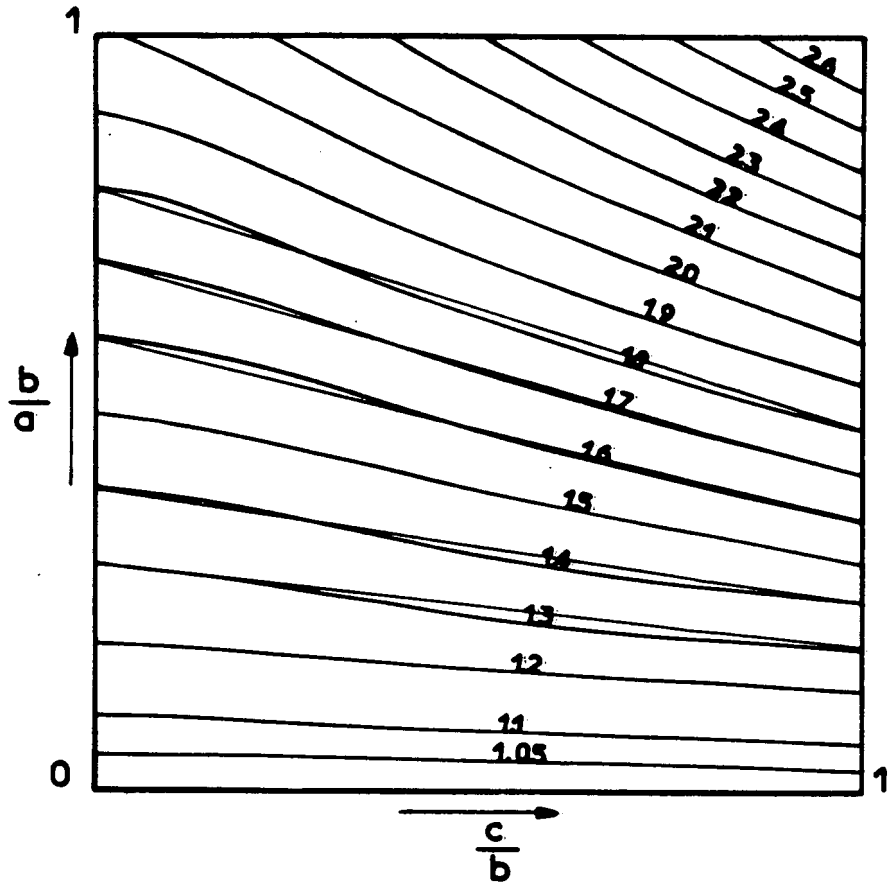


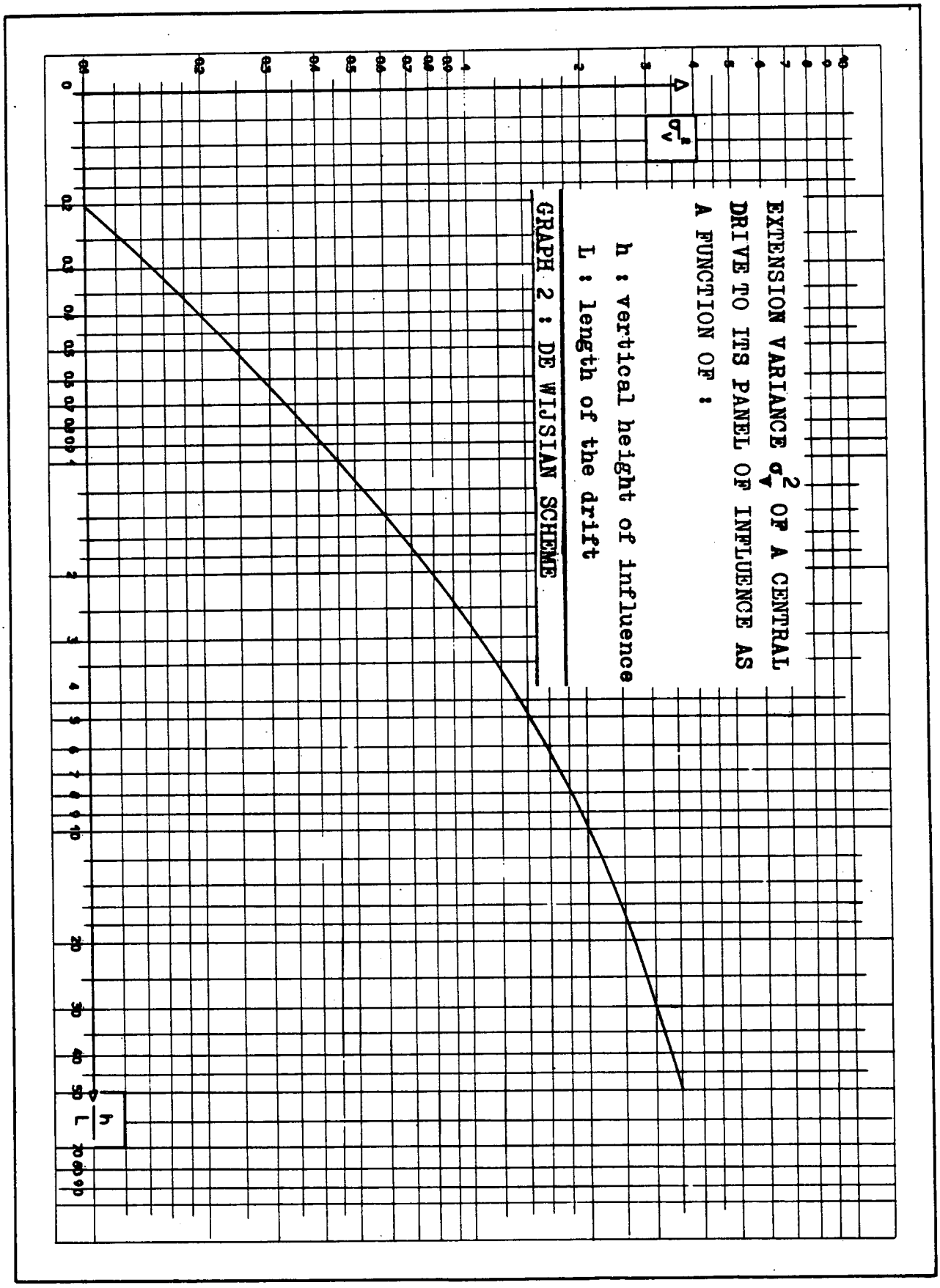
G R A P H I

LINEAR EQUIVALENCE IN A DE WIJSIAN SCHEME

The parallelepiped $a \times b \times c$ $(a \geq b \geq c)$ is equivalent to the linear sample λa

The curves are graduated in terms of λ.

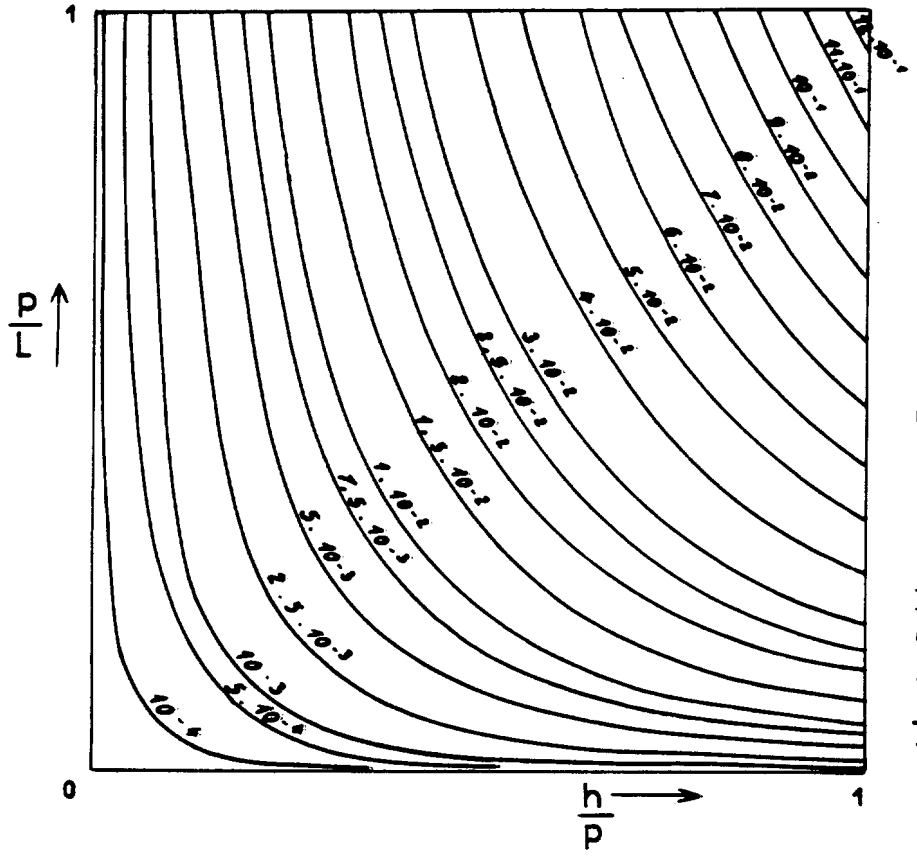




GRAPH 2 : DE WIJSIAN SCHEME

G R A P H 3

DE WIJSIAN SCHEME



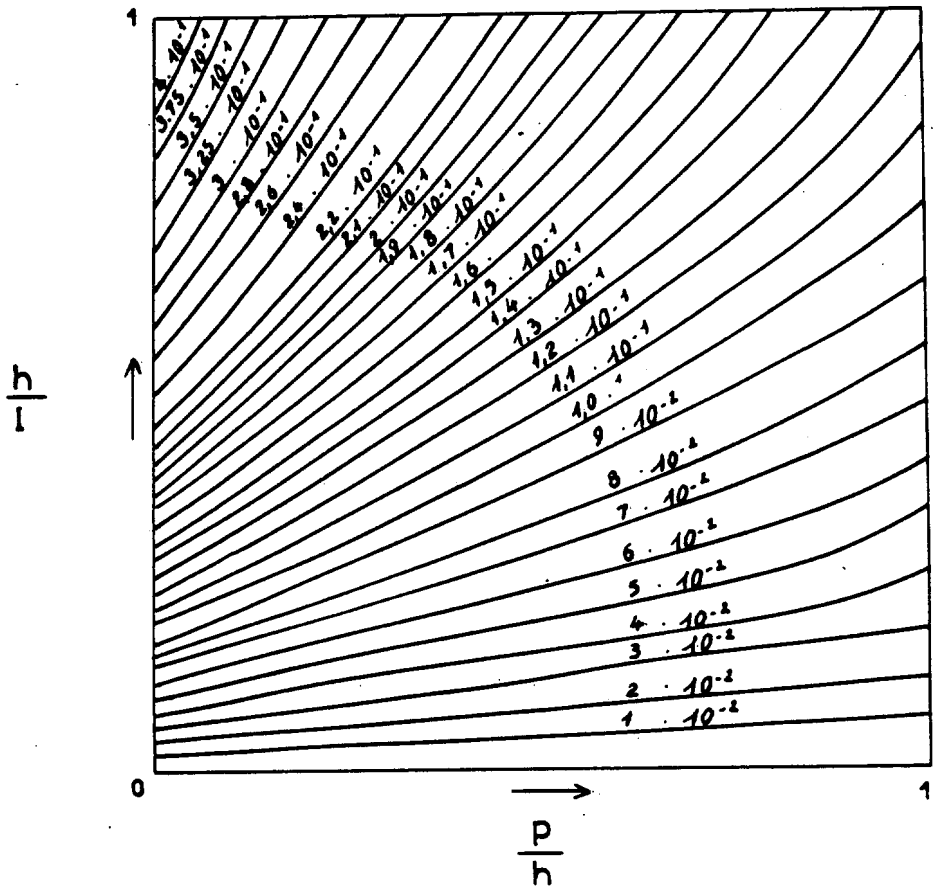
GRAPH 3 -

Extension variance of a large section L × p within the influence zone rise :

h : L ≥ p ≥ h.

The curves are graduated in terms of

$\frac{1}{3^{\alpha}} \sigma_E^2$



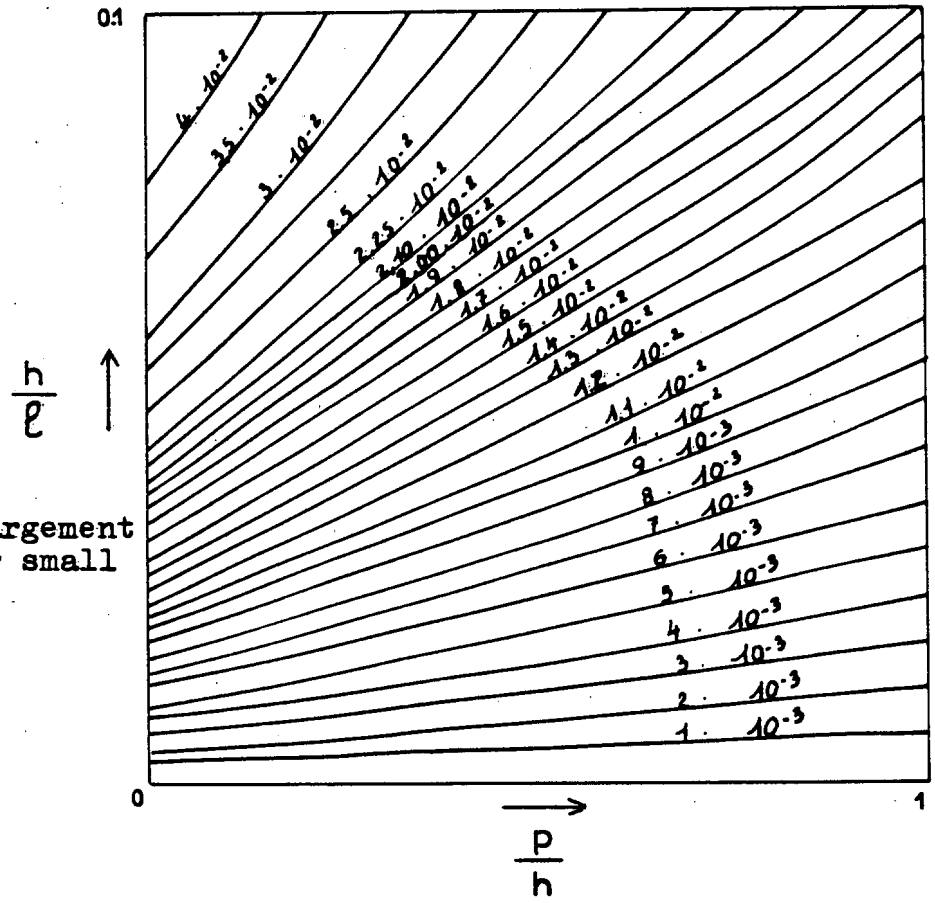
G R A P H 4

DE WIJSIAN SCHEME

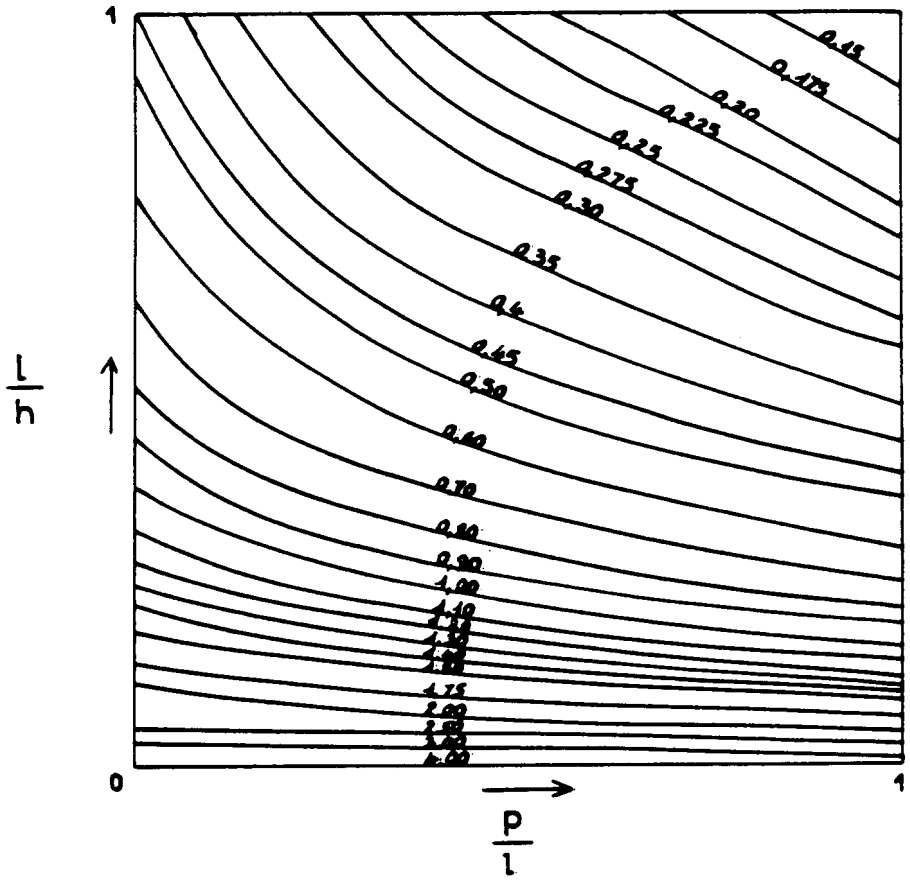
Graph 4 - Extension variance of an intermediate section l × p within its zone of influence of ris h = l ≥ h ≥ p

The curves are graduated in terms of

$$\frac{1}{3^{\alpha}}\,\sigma_E^2$$



Graph 4 - Enlargement of graph 3 for small values of $\frac{h}{l}$



G R A P H 5

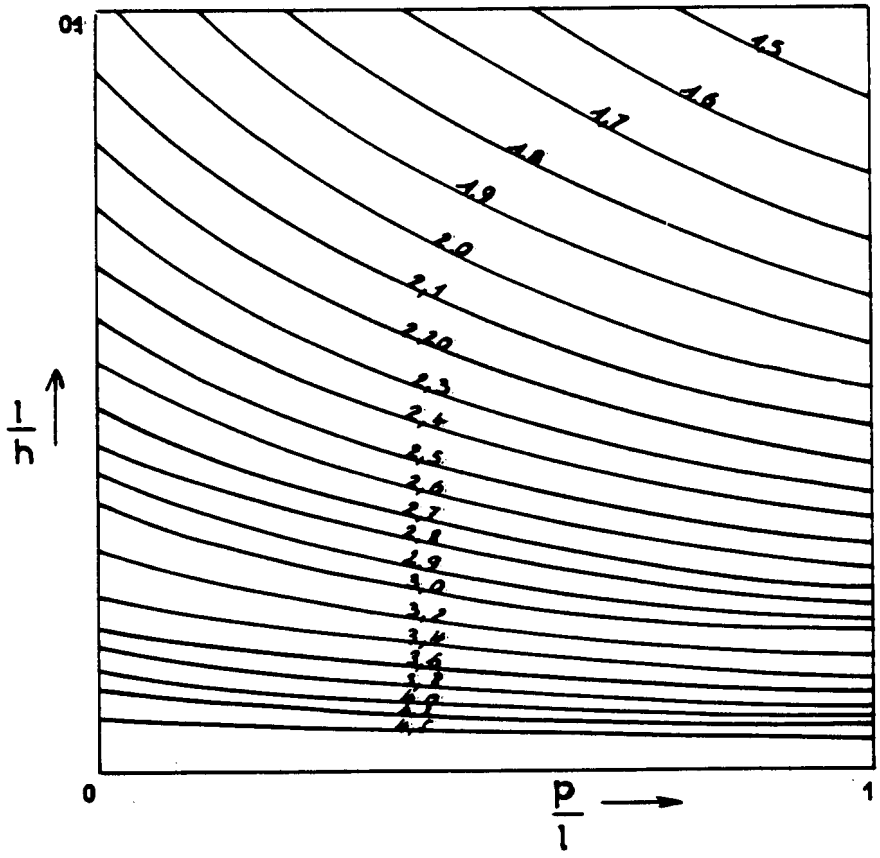
DE WIJSIAN SCHEME

Graph 5 - Extension variance of a small section l × p to its zone of influence of rise

$h = h \geq l \geq p$

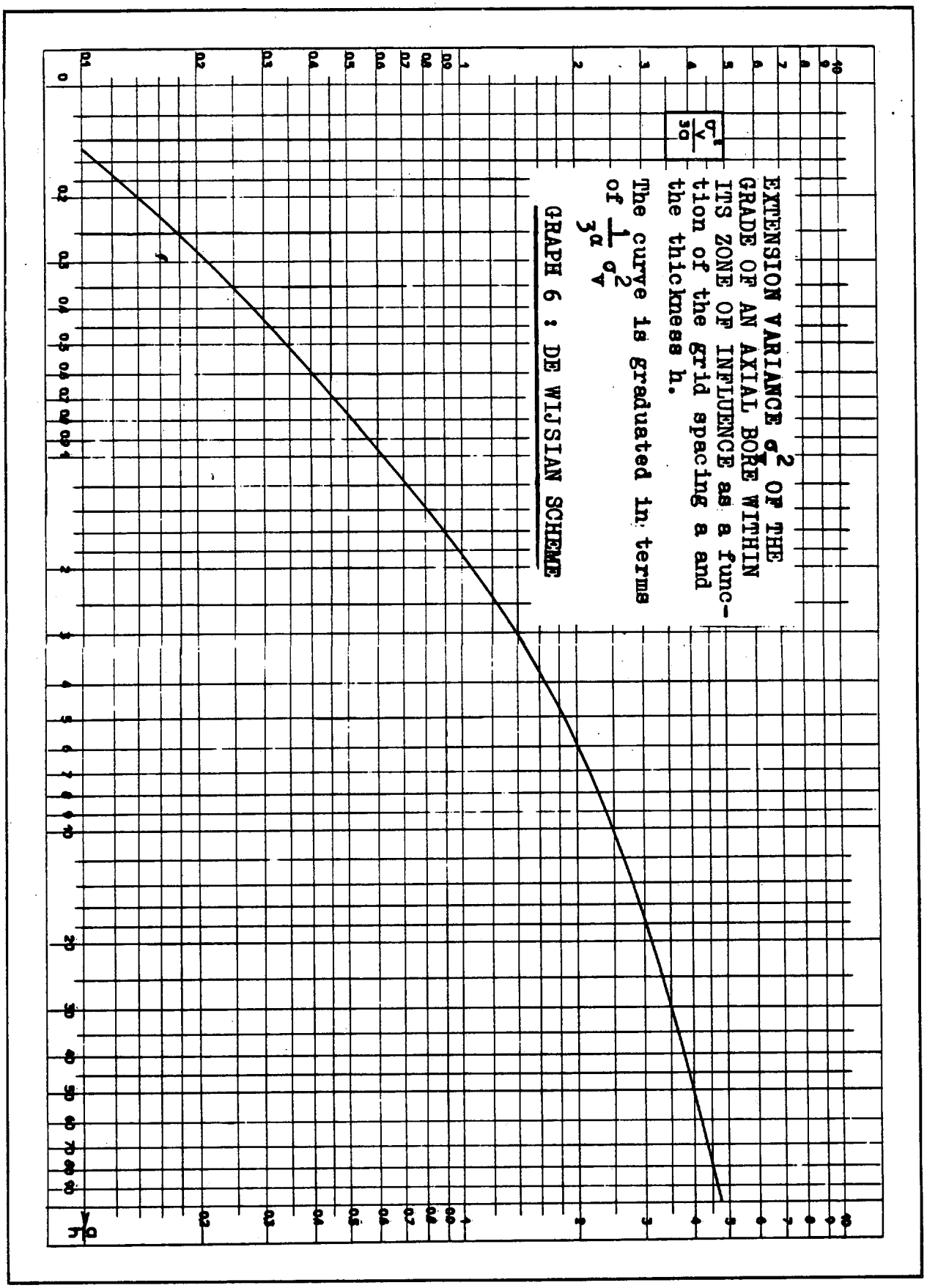
The curves are graduated in terms of

$$\frac{1}{3^{\alpha}} \sigma_E^2$$



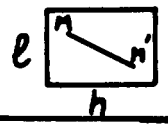
Graph 5 -

Enlargement of graph 5 for small values of $\frac{l}{h}$

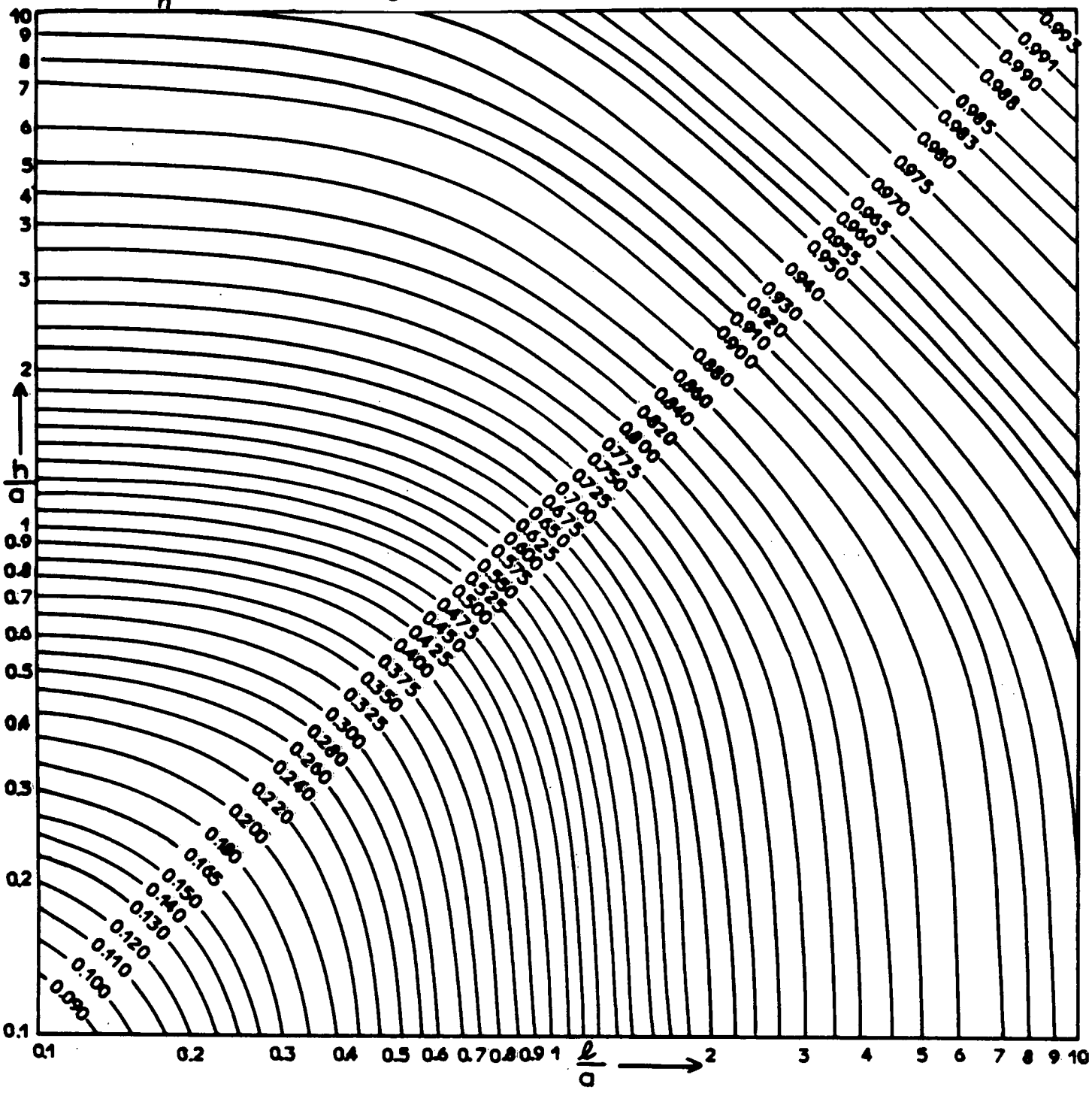


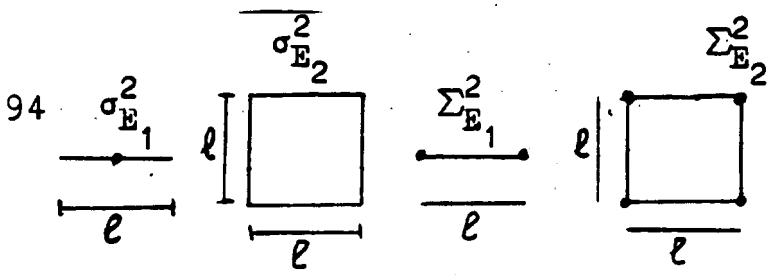
GRAPH 6 : DE WIJSIAN SCHEME

G R A P H 7 - SPHERICAL SCHEME



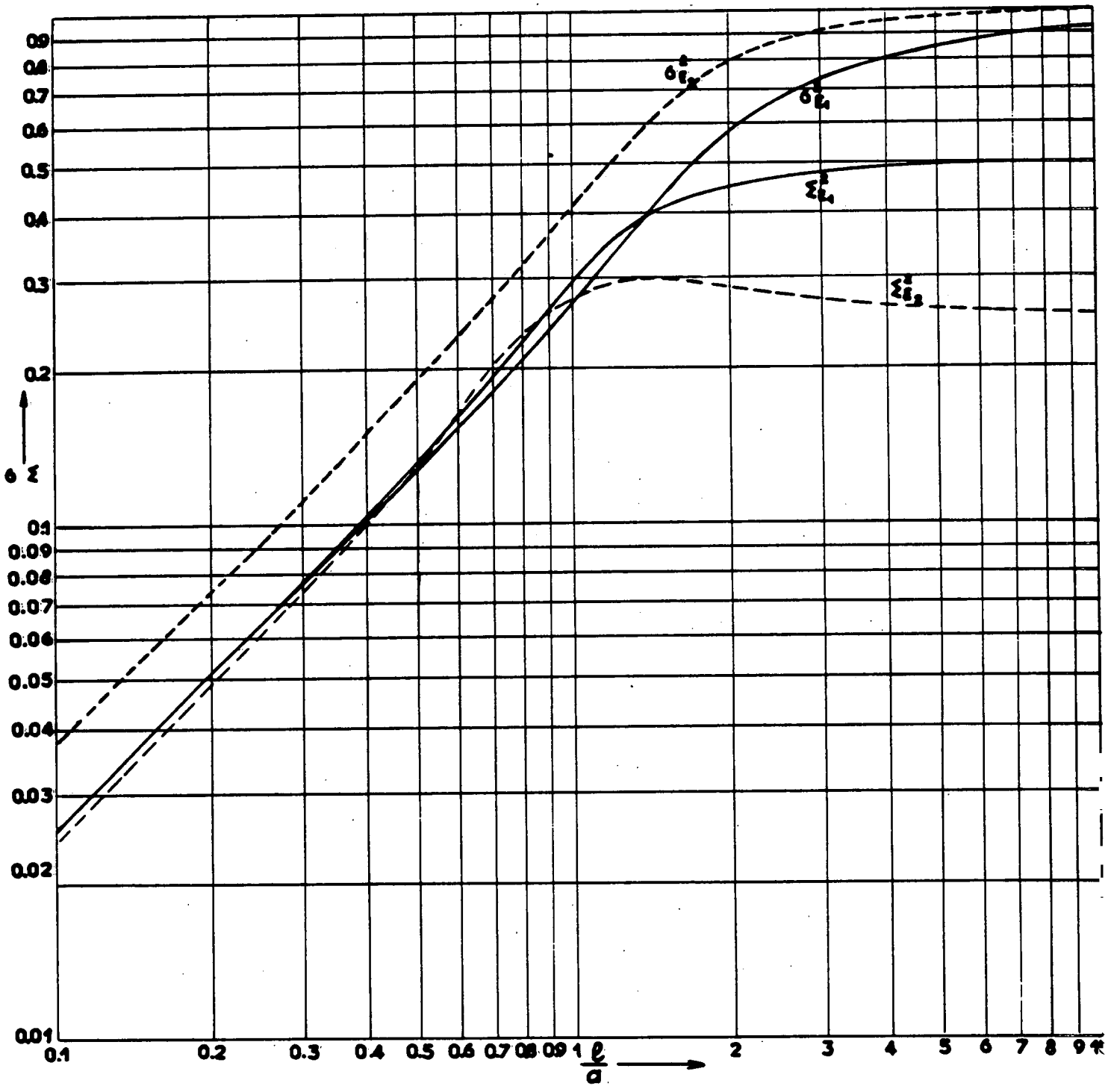
Function $\frac{1}{C}$ F $\frac{\ell}{a} \cdot \frac{h}{a}$ = variance of a point within the rectangle $\ell \times h$. (i.e. for C = 1)





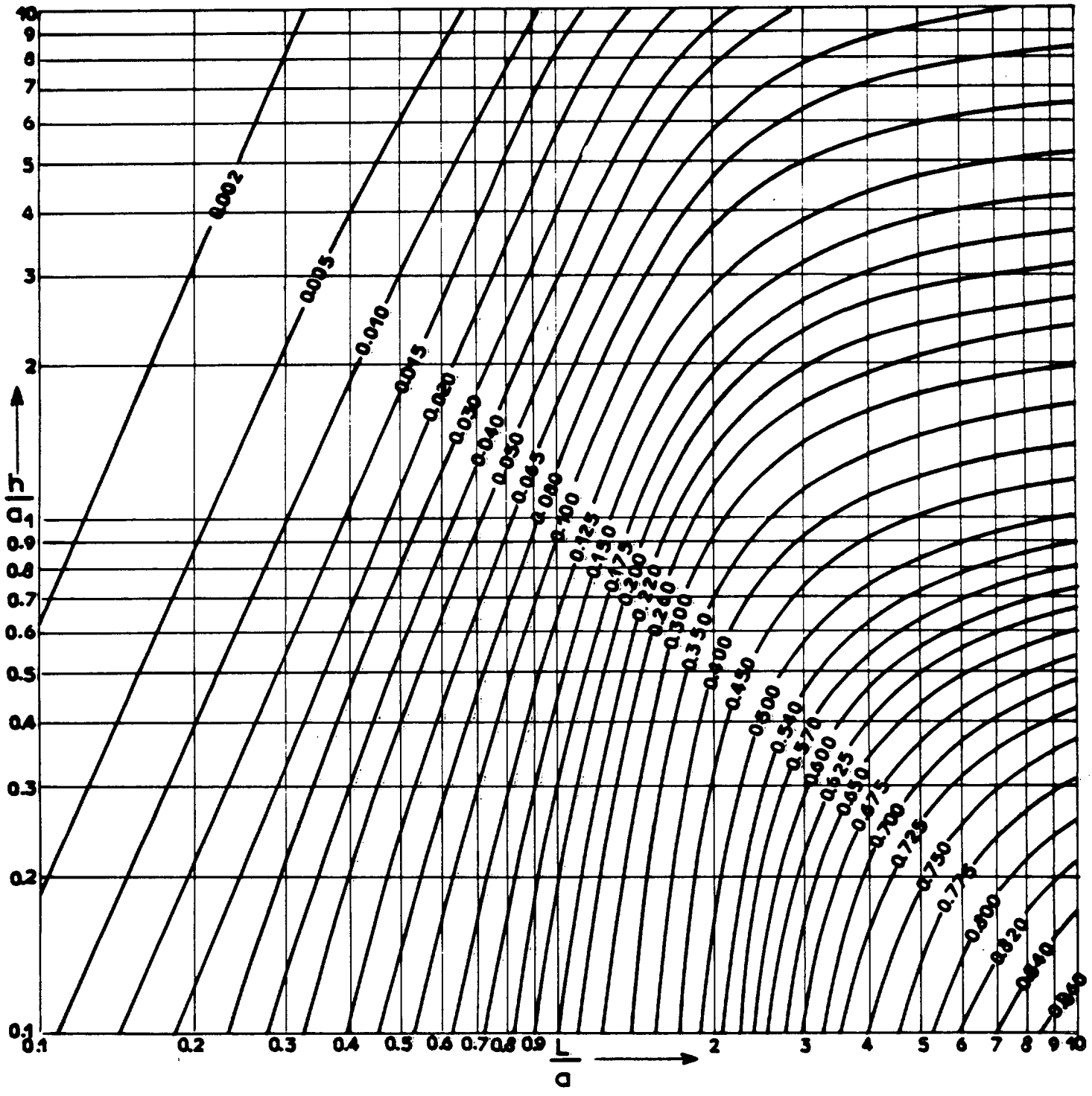
G R A P H 8 · SPHERICAL SCHEME

Various Extension Variances



. G R A P H 9 : SPHERICAL SCHEME

Extension variance of the median drift h within its rectangle L × h



2-10 STATISTICAL INFERENCE AND QUASI-STATIONARITY.

2-10-1 Quasi-stationary R.F.

Let us now return to the methodological problem raised in paragraph 1-4-6. We have to show - thanks to a rather weak hypothesis like the one of quasi-stationarity - that the problem of the global estimation of a given field V is solvable. This arises from the two following points :

~ The estimation variance (at least for small grid spacings) only depends upon the behaviour near the origin of a "mean" covariance or covariogram.

~ This behaviour is experimentally obtainable, in other words we can arrive at what we want to know by statistical inference from a single realization, provided that the R.F. is not differentiable q.m.

It should be possible to take as a basic minimal hypothesis the quasi-intrinsic character, i.e. the existance of a locally stationary variogram γ which deforms only slowly in space. Here we shall limit ourselves to the slightly stronger hypothesis of quasi-stationarity now being specified : a random function $Z(x)$ will be quasi-stationary if it has an expectation $m(x)$ and a centred covariance $C(x,y)$ such that :

a/ $m(x)$ is a very regular function slowly varying in space (at the scale of the grid) : more precisely, $m(x)$ can be considered as constant in a domain of the size of the grid.

b/ There is a function of three arguments $K(h;x,y)$ such that $C(x,y) = K(x-y;x,y)$ and such that, for a given h, $K(h;x,y)$ is a very regular and slowly varying (in the same sense as in a/) function of the two arguments x and y. In other words, when x and y belong to the same domain whose dimensions are those of the grid, $K(h;x,y)$ only depends on h, and it is as if the covariance C was locally stationary.

In fact, we shall even make a somewhat more restrictive hypothesis. We shall assume that in the domain V we want to estimate, the function K has an expansion as an integral series of the form :

$$K(h;x,y) = \sum_n \varpi_n(x)\, \varpi_n(y)\, C_n(h)$$

where the ϖ_n designate regular functions slowing varying at the scale of the grid, and the $C_n(h)$ are

covariance functions.

It is then possible to find stationary R.F.'s of the second order $Y_n(x)$, with zero expectation and mutually independent, such that :

$$Z(x) = m(x) + \sum_n \omega_n(x)\ Y_n(x)$$

This expansion is sufficiently general for what we have in mind. But problems like the computation of an estimation variance, estimation of a variogram and variance of this estimation; etc... will be seen to be linearly dependent on the independent components $Y_n(x)$. It will be sufficient if we handle the components $Y_n(x)$ separately, and sum up the results obtained. We thereby get an important simplification, for we are reduced to the case when $Z(x)$ is of the form

(2-33) $$Z(x) = m(x) + \omega(x)\ Y(x)$$

and thus has the expectation $m(x)$ and a centred covariance of the form

(2-34) $$C(x,y) = \omega(x)\ \omega(y)\ C_o(x-y)$$

Let us now examine how our two problems appear when relationships (2-33) and (2-34) are verified.

2-10-2 Computation of the estimation variance.

For simplification, the argument will be given for the one-dimensional case when the interval (0,L) is estimated from n samples on a regular grid of spacing a = L/n with a centred sampling arrangement. Let $x_i = (i - \frac{1}{2})a$ be the sampling points and

$$Z_i = \frac{1}{a} \int_{(i-1)a}^{ia} Z(x)\ dx$$

be the true assay of the influence zone number i. The estimation error associated with the estimator $\frac{1}{n} \sum_n Z(x_i)$ is

$$\frac{1}{n} \sum_i [Z(x_i) - Z_i]$$

In the term $Z(x_i) - Z_i$, the contribution of $m(x)$ is negligible, for $m(x)$ is considered as being constant over the influence zone : $m(x)$ plays no part in this problem. There remains

$$Z(x_i) - Z_i = \omega(x_i)\ Y(x_i) - \frac{1}{a} \int_{(i-1)a}^{ia} \omega(x)\ Y(x)\ dx$$

and, $\omega(x)$ being considered as constant over an influence zone :

$$Z(x_i) - Z_i = \omega(x_i) \left[Y(x_i) - \frac{1}{a} \int_{(i-1)a}^{ia} Y(x)\ dx \right] = \omega(x_i)\ [Y(x_i) - Y_i]$$

The variance of this individual error is thus :

$$D^2[Z(x_i) - Z_i] = [\omega(x_i)]^2\ \sigma^2_{E_o}$$

where $\sigma^2_{E_o}$ is the extension variance of x_i to its influence zone computed with the covariance C_o of $Y(x)$. Now, the approximation methods introduced in paragraph 2-5 show that these individual errors can be considered as independent. It follows that the estimation variance is :

$$\sigma^2_{Est} = \sigma^2_{E_o}\ \frac{1}{n^2}\ \sum_i\ \omega^2(x_i) \neq \frac{1}{n^2}\ \sigma^2_{E_o}\ \frac{1}{a} \int_o^L \omega^2(x)\ dx$$

i.e.

$$\sigma^2_{Est} = \frac{1}{n}\ \sigma^2_{E_o}\ \frac{1}{L} \int_o^L \omega^2(x)\ dx$$

Hence, this estimation variance would be the same if we had the stationary covariance :

$$C(h) = C_o(h)\ \frac{1}{L} \int_o^L \omega^2(x)\ dx$$

Let then $\overline{C}(h)$ be the (pseudo-stationary) mean covariance :

$$\overline{C}(h) = \frac{1}{L-h} \int_o^{L-h} C(x, x+h)\ dx = \frac{C_o(h)}{L-h} \int_o^{L-h} \omega(x)\ \omega(x+h)\ dx$$

When $h \leq a$, $\omega(x+h)$ is not different from $\omega(x)$, and we have :

$$\frac{1}{L-h}\int_0^{L-h} \omega(x)\,\omega(x+h)\,dx \neq \frac{1}{L}\int_0^{L} \omega^2(x)\,dx$$

Thus, when $h \leq a$, $C(h) = \overline{C}(h)$ and the estimation variance is the same as that for a stationary R.F. with a covariance which would be precisely this mean covariance $\overline{C}(h)$.

When a is small, we still have to show that it is possible to estimate $\overline{C}(o) - \overline{C}(h)$.

2-10-3 Possibility of statistical inference.

Put $\overline{\gamma}(h) = \overline{C}(o) - \overline{C}(h)$, $\gamma_o(h) = C_o(o) - C_o(h)$ etc... We are trying to estimate the behaviour of $\overline{C}(h)$ near the origin, i.e. of $\overline{\gamma}(h)$ when a is small. Therefore, we take as an estimator :

$$\gamma^*(h) = \frac{1}{2(L-h)}\int_0^{L-h} [\omega(x+h)\,Y_{x+h} - \omega(x)\,Y_x]^2\,dx \neq \frac{1}{2(L-h)}\int_0^{L-h} \omega^2(x)\,[Y_{x+h} - Y_x]^2\,dx$$

With the approximations already made, this estimator is without bias :

$$E[\gamma^*(h)] = \overline{\gamma}(h)$$

Its variance remains to be evaluated. Here, it is necessary to,introduce an assumption about the distribution of the R.F., for moments of order 4 will necessarily intervene. The simplest way is to limit ourselves to the Gaussian case as in Exercise 16. The computations made in paragraphs a/ and b/ of this exercise remain valid here and give :

$$(2\text{-}35)\quad D^2[\gamma^*(h)] = \frac{1}{2(L-h)^2}\int_0^{L-h} \omega^2(x)dx \int_0^{L-h} \omega^2(y)dy\,[\gamma_o(x-y+h) + \gamma_o(x-y-h) - 2\,\gamma(x-y)]^2\,dy$$

Let us evaluate this integral when a si small. Obviously it is equivalent to

$$\frac{1}{2L^2}\int_0^{L}\int_0^{L} \omega^2(x)\,\omega^2(y)\,[\gamma_o(x-y+h) + \gamma_o(x-y-h) - 2\,\gamma(x-y)]^2\,dx\,dy$$

Put

$$F(u) = [\gamma_o(u+h) + \gamma_o(u-h) - 2\,\gamma_o(u)]^2$$

and let R be the transitive covariogram of the function equal to $\varpi^2(x)$ on (0,L) and 0 elsewhere. From Cauchy's algorithm, it follows that :

$$D^2[\gamma^*(h)] = \frac{1}{L^2} \int_0^L R(u)\ F(u)\ du$$

When $u > h$, F(u) is a regular function of u, equivalent to $h^4[\gamma''(u)]^2$. Hence the variance is of the form

$$D^2[\gamma^*(h)] = A\ h^4 + \frac{R(o)}{L^2} \int_0^h F(u)\ du$$

Everything depends now on the behaviour of F(u) near 0. Let $b\ h^\lambda$ (eventually $b\ h^2 \log h$) the principal part of γ_o at h = 0. That of F(u) is :

$$b^2\ h^{2\lambda}\ [(1 + \frac{u}{h})^\lambda + (1 - \frac{u}{h})^\lambda - 2\ u^\lambda]^2$$

Hence, the variance will be of the form :

(2-36) $$D^2[\gamma^*(h)] = A\ h^4 + \frac{B}{L^2}\ h^{1+2\lambda}$$

Thus, two cases are to be distinguished :

a/ $\lambda < 2$ (R.F. non differentiable q.m.) The relative variance, for h small, is :

$$\frac{D^2[\gamma^*(h)]}{[\gamma(h)]^2} = A'\ h^{4-2\lambda} + \frac{B'}{L^2}\ h$$

It is thus infinitely small compared with h and statistical inference is possible. In the limiting case when $\gamma(h)$ has a behaviour term in $h^2 \log h$, a relative variance in $A'/\log h$ would be found : Theoretically, statistical inference would still be possible, but would be very difficult to do in practice because of the very slow rate of decrease of $1/\log h$.

b/ $\lambda = 2$ (R.F. differentiable q.m.) The relative variance $D^2(\gamma^*)/\gamma^2$ tends to a constant A' different from 0 when h tends to 0. This means (unless L is very great with respect to the range of $C_o(h)$) that statistical inference is impossible.

In short, statistical inference is always possible when the R.F. is not differentiable q.m.,

and only in this case is the problem solvable in all generality. This quite singular circumstance which penalizes R.F!s which are too regular and prohibits the inference of their covariance when the realization is not very extensive will be appreciated in full when we study universal kriging and in particular the notion of random drift.

2-11 EXERCISES ON INTRINSIC RANDOM FUNCTIONS.

2-11-1 Construction of I.R.F's.

Exercise 1 - An origin x_0 having been chosen at random on (0,a), the straight line is divided into segments of length a with subdivision points $x_0 + ka$ (k a positive or negative integer). Let Y(x) be a R.F. taking, for each of these segments of length a, a random constant value Y. The values are drawn by chance independently from one segment to another according to the same law of probability, with mean value m and variance σ^2 . Calculate the probability that x and x+h belong to the same segment a.
(Solution : $1 - |h|/a$ when $|h| \leq a$, 0 otherwise).
Derive from this the semi-variogram of Y(x). ($\frac{|h|}{a} \sigma^2$ when $|h| \leq a$ and σ^2 otherwise).

Exercise 2 - Same question as in 1, but the lengths of the segments are now randomly distributed, independently from each other, and follow the same exponential law $e^{-\lambda h}$ (in other words, the points of discontinuity constitute a Poisson process).
(Solution : Prob {x and x+h $\in$ same segment} = $e^{-\lambda h}$. Then $\gamma(h) = (1 - e^{-\lambda h})\, \sigma^2$.

Exercise 3 - Consider a Poisson process on the line (random points separated by independent segments obeying the same law of density a e^{-ax}). A R.F. Y(x) is defined (except for a constant) by taking $Y(x) = C^{ste}$ outside the poissonian points x_i and, at each x_i , $Y_+(x_i) - Y_-(x_i) = X_i$, with independent r v 's x_i obeying the same distribution of expectation m and variance σ^2 .
Show that the R.F. Y(x) is intrinsic (but not stationary of order 2), has a drift m a h and semi-variogram $\frac{1}{2}\, a(m^2 + \sigma^2)|h|$ (linear).
(Solution : This process is a compound Poisson process with independent and stationary increments. To compute the drift, calculate the expectations $E[Y(x+h) - Y(x)]$ and

$E[Y(x+h) - Y(x)]^2$ conditionally when the number n of poissonian points falling within x and x+h is given, and then decondition with respect to n).

Exercise 4 - a/ Let Y(x) be a stationary R.F. of order 2, m its expectation, C(h) its covariance. Show that the R.F. $Z(x) = \int_0^L Y(y)\,dy$ is intrinsic (but not stationary of order 2) and that it has a drift mh and a semi variogram $\gamma(h) = \int_0^h (h-x)\,C(x)dx$.
(Solution : start from $\gamma'' = C$).

b/ Application to the process with random slopes : consider poissonian discontinuity points as in Ex. 3, with the slope Π remaining constant in the intervals between those points. These slopes Π are random, independent, have a zero expectation and a variance ϖ_2 in each of these intervals. Show that the process Y(x) so defined (except for a constant) is an I.R.F. (non stationary od order 2) without drift, and with semi-variogram $\varpi_2 \frac{e^{-ah} - 1+ah}{a^2}$. Notice the parabolic behaviour near h = 0 of the variogram.
(Solution : this process is of the type of Ex. 2, and it is sufficient to apply a/ to Y'(x)).

2-11-2 Exercises on estimation variances.

Exercise 5 - Consider the semi-variogram h^λ $(0 < \lambda < 2)$ in a one-dimensional space. Calculate the auxiliary functions X and F, the elementary extension variance of a central point to its segment of influence, and the extension variance to this segment of its two extremities (closed arrangement). Compare and discuss.
Solution :

$$X(h) = \frac{h^\lambda}{\lambda+1} \quad ; \quad F(h) = \frac{2\,h^\lambda}{(\lambda+1)(\lambda+2)} \quad ; \quad \frac{2a^\lambda}{\lambda+1}\left(\frac{1}{2^\lambda} - \frac{1}{\lambda+2}\right) \quad ; \quad \left(\frac{2}{\lambda+2} - \frac{1}{2}\right)a^\lambda$$

We have $\sigma_E^2 > \sigma_E^2$, when $\lambda > 1$, whereas the expressions are equal when $\lambda = 1$: when $\lambda > 1$, there is high continuity, and a single well-placed sample is better than two ill-placed ones ; when $\lambda < 1$, we are not far from the purely random situation and two samples, even badly placed, are always more informative than a signle one. When $\lambda = 1$, the two conditions are equivalent.

Exercise 6 - In a one-dimensional space with $\gamma(h) = h^\lambda$, compute the estimation variance of the segment L = na by n point samples, in the three following cases :

- regular grid (closed arrangement) $\left(\frac{1}{n}\frac{2}{\lambda+1}\left[\frac{1}{2^\lambda} - \frac{1}{\lambda+2}\right] a^\lambda\right)$
- stratified random grid $\left(\frac{1}{n}\frac{2\,a^\lambda}{(\lambda+1)(\lambda+2)}\right)$
- purely random grid $\left(\frac{1}{n}\frac{2}{(\lambda+1)(\lambda+2)}\ L^\lambda\right)$

Exercise 7 - Consider, in a two-dimensional space, the semi-variogram r^λ and an orebody developed by drives, whose lengths are greater than the level interval h.

a/ Calculate the extension variance of a level of length ℓ to its rectangle of influence ℓ,h.

(grading under constant thickness ℓ transforms r^λ into $A\, r^{\lambda+1}/\ell$ (or log r into $\pi r/\ell$); thus :

$$\sigma_E^2 = A\frac{2}{\lambda+2}\left|\frac{1}{2^{1+\lambda}} - \frac{1}{\lambda+3}\right|\frac{h^{1+\lambda}}{\ell} = B\frac{h^{1+\lambda}}{\ell}$$

b/ Deduce from this the estimation variance (weight by the squares of the lengths). Put $L = \Sigma\, \ell_i$ and S = L h, and write this variance in the form

$$B\frac{S^{1+\lambda}}{L^{2+\lambda}} \quad - \quad \text{(De Wijsian case : } A = \pi,\ B = \frac{\pi}{6}\text{ , }\frac{\pi}{6}\ \frac{S}{L^2}\text{)}$$

Exercise 8 - Same problem as in 7, but assume in addition that the levels are estimated from channel samples with a spacing a. Compute the line term $\left(\frac{C}{L}\, a^{1+\lambda}\right.$ with $C = \frac{2}{\lambda+1}\left[\frac{1}{2^\lambda} - \frac{1}{\lambda+2}\right]\right)$.
Economic optimization : let ϖ be the cost per meter of drive, p_o the cost of each sample, and $\ell_o = p_o/\varpi$. Optimize the precision for a given net cost or, equivalently, the cost for a given precision.
(It amounts to minimizing $C\, h\, a^{1+\lambda} + B\, h^{2+\lambda}$ under the condition $\frac{1}{n\ell_o} + \frac{1}{ha} = C$. We get the following relationship :

$$B(2+\lambda)\ h^{1+\lambda} = \lambda\ C\ a^{1+\lambda} + C(1+\lambda)\ \frac{a^{2+\lambda}}{\ell_o}$$

which gives the spacing h of the levels as a function of the spacing a of the channel samples.

2-11-3 Exercises on the nugget effect, De Wijsian and spherical schemes.

Exercise 9 - Nugget effect in the pure state.- We are given nuggets set at points distributed in space according to a Poisson scheme (definition : the number N(v) of nuggets contained in v is a Poisson r.v. of expectation λv ; if v and v' are disjointed, N(v) and N(v') are independent).

a/ Take out volumes v, and put $Y(x) = N(v_x)$, v_x designating the translate of v located at point x. Show that the covariance of Y(x) and Y(x+h) is the variance of $N(v_x \cap v_{x+h})$ or $\lambda K(h)$, K(h) designating the geometric covariogram of the volume v.

b/ The nuggets are now supposed to have random independent weights (mean p_o, variance σ_p^2), and Y(x) is the sum $P_1 + \ldots P_N$ of the weights of the $N = N(v_x)$ nuggets contained in v_x. Calculate the mean and the variance of Y(x) ($E(Y) = \lambda v\, p_o$, $D^2(Y) = \lambda_v(p_o^2 + \sigma_p^2)$. What is the covariance of Y(x) and Y(x+h) ? (replace v by K(h)).

Exercise 10 - Dilution scheme.- We are given Poisson points as in Ex. 9. Let f(x) be a function and $Y(x) = \sum_i f(x-x_i)$, x_i designating the location of the different nuggets. Thus Y(x) is a dilution of these nuclei. Find the covariance K(h) of Y(x).
(Solution : $\lambda\, g(h)$, with $g = f * \check{f}$).

(Only the case where f has a compact support can be examined : consider two points x_o and $x_o + h$ and a bounded domain V containing the translates by x_o and $x_o + h$ of the support of f. First reason conditionally when the number n of poissonian points falling within V is fixed, then decondition with respect to n. Notice also how interpretation by means of the covariance measure $\lambda\delta$ of the poissonian points makes this result intuitive).

Exercise 11 - Consider the De Wijs scheme $\gamma(r) = \log r$ in a two-dimensional space. It will be recalled that if g is a harmonic function inside a circle, its mean value within the circle (or on its circumference) is equal to its value at the centre : log r is a harmonic function except for r = 0.

a/ C(R), X(R), F(R), $C_o(R)$ designate the mean value of log r, respectively, on the circumference of a circle of radius R, between this circle and its circumference, within this circle, and between this circle and its centre. Show that C(R) = X(R) = log r ; F(R) = log R - 1/4 ; $C_o(R)$ = log R - 1/2 (the first two relations result from the theory of harmonic functions. To calculate F(R), vary $R^4F(R)$ and show that $\frac{d}{dR} R^4F(R)$ =

$= 4\ R^3\ X(R)$. For $C_o(R)$, integrate $2\pi r\ X(r)$).

b/ Compute the estimation variance of a circular panel from its circumference ($\sigma_E^2 = 2\ X(R) - F(R) - C(R) = \frac{1}{4}$).

c/ Extension variance to the circle of a small central element whose linear equivalent is $\ell(\log R/\ell + 3/4)$. If L is the linear equivalent of the circle, this estimation variance is also $\log (L/\ell) - 1/2$. Whence, deduce that the extension variance of a sample equivalent to in the square (a,a) at whose centre it is set, is approximately $\log \frac{2a}{\ell} - 1/2$.

Exercise 12 - Spherical scheme in the one-dimensional space.- a/ Compute the auxiliary functions of the spherical scheme :

$$\gamma(\ell) = \frac{3}{2}\ \frac{\ell}{a} - \frac{1}{2}\ \frac{\ell^3}{a^3} \qquad \text{when } \ell \leq a, \quad 1 \text{ when } \ell \geq a$$

$$X(\ell) = \frac{3}{4}\ \frac{\ell}{a} - \frac{1}{8}\ \frac{\ell^3}{a^3} \qquad " \qquad 1 - \frac{3}{8}\ \frac{a}{\ell}$$

$$F(\ell) = \frac{1}{2}\ \frac{\ell}{a} - \frac{1}{20}\ \frac{\ell^3}{a^3} \qquad " \qquad 1 - \frac{3}{4}\ \frac{a}{\ell} + \frac{1}{5}\ \frac{a^2}{\ell^2}$$

b/ Compute the elementary extension variance of a sample within its segment of influence b when $b \leq a$ and $b \geq 2a$, and interpret the results.

$$\frac{1}{4}\ \frac{b}{a} + \frac{3}{160}\ \frac{b^3}{a^3} \quad \text{and} \quad 1 - \frac{3}{4}\ \frac{a}{b} - \frac{1}{5}\ \frac{a^2}{b^2}$$

2-11-4 Exercises on large grids.

The three following exercises constitute a test. If the reader is able to understand the first two exercises and do the corresponding computations (which are easy in themselves) he has certainly assimilated the mental processes of a geostatistician. Moreover, if he can accept the critical conclusions of the last exercise without becoming confused (for large grids the principle of composition of line and slice terms is not valid, and the principle of composition of elementary extension variances applies only very approximately) this shows his comprehension goes deeply and allows him to weight the significance and limits of each of the approximate assumptions which

render geostatistics operative.

From an epistemological point of view, it will be noted that for large grids (i.e. greater than the range) the exact form of the covariance or of the $\gamma(h)$ loses all its importance : K(h) only intervenes by virtue of its value C at the origin and by its first moments (A_0 and A_1 in R^1, A_0, A_1, A_2 and A_3 in R^2 ; in R^3, we must go up to A_5). Thus, in R^2 for example, all the schemes are reduced to a unique type depending only on the four essential parameters A_0, A_1, A_2 and A_3 and on the parameter C (which only intervenes as a multiplying factor). For large grids, we are relatively near the conditions of validity of classical statistics : to the estimation variance C/n which classical statistics would give, geostatistical computation adds corrective terms whose coefficients are the A_1 themselves.

Exercise 13 (1 dimension) - Let $\gamma(h) = \gamma(r)$ be an (isotropic) variogram of the transitive type, i.e. of the form :

$$\gamma(r) = C - K(r) \qquad \text{with } K(r) = 0 \text{ when } r > \text{range}.$$

1/ In a one-dimensional space, when $\ell >$ range, we have :

$$C - F(\ell) = \frac{2}{\ell^2} \int_0^{\ell} (\ell - x)\, K(x)\, dx = \frac{2}{\ell^2} \int_0^{\infty} (\ell - x\, K(x)\, dx$$

Deduce from this :

$$F(\ell) = C - \frac{A_0}{\ell} + \frac{A_1}{\ell^2}$$

$$\chi(\ell) = C - \frac{A_0}{2\ell}$$

with :

$$A_0 = 2\int_0^{\infty} K(x)\, dx \quad , \quad A_1 = 2 \int_0^{\infty} x\, K(x)\, dx$$

2/ Extension variance for a grid spacing a > range :

$$\sigma_E^2 = 2\ \chi\ \frac{a}{2} - F(a) = C - \frac{A_0}{a} - \frac{A_1}{a^2}$$

For a drive of length ℓ = na, the estimation variance is

$$\frac{1}{n}\,\sigma_E^2 = \frac{a}{\ell}\,C - \frac{A_0}{\ell} - \frac{A_1}{a\ell}$$

Exercise 14 (2 dimensions) - 1/ Function F(a,b). The mean value of $\gamma(h) = \gamma(r)$ in the rectangle (a,b) est :

$$F(a,b) = \frac{4}{a^2\,b^2}\int_0^a\int_0^b (a-x)\,(b-y)\,\gamma\,(\sqrt{x^2+y^2})\,dx\,dy$$

2/ Let $\gamma(h) = C - K(h)$ be a transition scheme (i.e. $K(h) = 0$ when $r = |h| > \varepsilon$, ε being the range). For large grids (a,b > range) 1/ gives :

$$C - F(a,b) = \frac{4}{a^2\,b^2}\int_0^\infty\int_0^\infty (a-x)\,(b-y)\,K(\sqrt{x^2+y^2})\,dx\,dy$$

Deduce from this :

$$F(a,b) = C - \frac{\pi\,A_1}{ab} + \frac{2\,A_2}{ab}\;\frac{1}{a} + \frac{1}{b}\; - A_3\,\frac{1}{a^2\,b^2}$$

with :

$$\begin{cases} A_1 = \frac{1}{\pi}\int_{R^2} K(h)\,dh = 2\int_0^\infty r\,K(r)\,dr \\ A_2 = 2\int_0^\infty r^2\,K(r)\,dr \\ A_3 = 2\int_0^\infty r^3\,K(r)\,dr \end{cases}$$

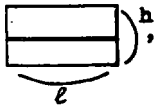
3/ Deduce from 2/ X(a;b), $\gamma(a;b)$ and Q(a;b) (still for a,b > range).($X(a;b) = X_b(a)$). Apply :

$$X(a;b) = \frac{1}{a}\,\frac{\partial}{\partial a}\;\frac{a^2\,F}{2} \qquad \gamma(a;b) = \frac{\partial^2}{\partial a^2}\left(\frac{a^2\,F}{2}\right)$$

$$Q(a;b) = \frac{1}{ab}\;\frac{\partial^2}{\partial a\,\partial b}\left(\frac{a^2\,b^2\,F}{4}\right)$$

this must give : $X = C - \frac{\pi\,A_1}{2\,ab} + \frac{A_2}{a\,b^2}$, $\gamma = C$

$$Q = C - \frac{\pi\,A_1}{4\,ab}$$

4/ Extension variance of a drive ℓ in $\ell \times h$,  , $\ell, \frac{h}{2}$ > range.

$$\sigma_E^2 = 2\,\bar{\gamma}\left(\frac{h}{2}, \ell\right) - F(h,\ell) - F(\ell)$$

$$\sigma_E^2 = \frac{A_o}{\ell} - A_1\left(\frac{\pi}{\ell h} + \frac{1}{\ell^2}\right) + 2\,A_2\left(\frac{1}{h\ell^2} - \frac{1}{h^2\ell}\right) + A_3\,\frac{1}{h^2\ell^2}$$

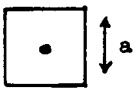
5/ Estimation variance of S by drives of length ℓ_i > range (equidistance of the drives h > range). Put $\Sigma\,\ell_i = L$, N number of drives.

$$\frac{\Sigma\,\ell_i\,\sigma_{E_i}^2}{(\Sigma\,\ell_i)^2} = \frac{A_o}{L} - A_1\left(\frac{\pi}{Lh} + \frac{N}{L^2}\right) + 2\,A_2\left(\frac{N}{hL^2} - \frac{1}{h^2L}\right) + A_3\,\frac{N}{h^2L^2}$$

$$= A_o\,\frac{h}{S} - A_1\left(\frac{\pi}{S} + \frac{h^2N}{S^2}\right) + 2\,A_2\left(\frac{hN}{S^2} - \frac{1}{hS}\right) + A_3\,\frac{N}{S^2}$$

6/ Estimation of S by drives (equidistance h > range) sampled at a spacing a > range :

$$\sigma_{Est}^2 = \frac{ah}{S}\,C - A_1\left(\frac{h}{aS} + \frac{\pi}{S} + \frac{h^2N}{S^2}\right) + 2\,A_2\left(\frac{hN}{S^2} - \frac{1}{hS}\right) + \frac{A_3N}{S^2}$$

7/ Extension variance of a borehole 

$$\sigma_E^2 = 2\,\bar{\gamma}\left(\frac{a}{2}, \frac{a}{2}\right) - F(a,a) = C - \frac{\pi\,A_1}{a^2} - 4\,\frac{A_2}{a^3} + \frac{A_3}{a^4}$$

hence the estimation variance of S :

$$\frac{1}{n}\,\sigma_E^2 = C\,\frac{a^2}{S} - \pi\,\frac{A_1}{S} - 4\,\frac{A_2}{aS} + \frac{A_3}{a^2S}$$

Exercise 15 (Critical study) - 1/ Let us take a critical look : for h = a, formula 6/ is not consistent with the result found in 7/ (except for the first term $\frac{C}{n}$ which is the approximation of classical statistics). Between the two principles of approximation :

~ composition of slice term and line term (used in 6/)

~ composition of elementary extension variances (used in 7/)

At least one is unusable in the case of large grids.

On second thought, it appears that the estimation variance associated with a rectangular grid (a,b) must depend only on the surface ab of the rectangle and not on the ratio a/b (a > b) when a and b are greater than the range. Formula 6/ makes this ratio a/h intervene explicitly . It can be concluded that the principle of composition of line terms and slice terms cannot be applied to large grids, and that formula 7/ must be preferred.

2/ To verify this rigourously, start from the general formula :

$$\sigma^2_{Est} = \frac{1}{S^2} \int_S \int_S K(x-y)\, dx\, dy - \frac{2}{NS} \sum_i \int_S K(x_i-y)\, dy + \frac{1}{N^2} \sum_{i,j} K(x_i-x_j)$$

When the distances $|x_i-x_j|$ between the boreholes and the distances from the boreholes to the boundary of S are greater than the range, this can be reduced to the (rigourous) formula :

$$\sigma^2_{Est} = C - F(S) + \frac{C}{N} - \frac{2\pi A_1}{S}$$

which is valid whatever the shape of the grid (and which can be applied, in particular, even to irregular grids).

C - F(S) depends only on the surface S (and not on the grid) and its principal part (for S large) is $\pi A_1/S$; hence the approximation formula

$$\sigma^2_{Est} = \frac{C}{N} - \pi \frac{A_1}{S}$$

3/ Now look whether the principle of composition of elementary extension variances gives better results. We get :

$$\frac{1}{N} \left[2\, Q\left(\frac{a}{2}, \frac{b}{2}\right) - F(a,b)\right] = \frac{C}{N} - \pi \frac{A_1}{S} - \frac{2A_2}{S} \left(\frac{1}{a} + \frac{1}{b}\right) + \frac{A_3}{Sab}$$

The first two terms are good. The terms of higher order (in A_2 and A_3) are not acceptable, for they still depend on a and b whereas the exact formula only makes the geometry of S intervene. Nevertheless this formula is better than 6/ (in which the first term only is good).

4/ As in paragraph 2/ above, show that the estimation variance of a segment L by a grid of spacing a > range, is (rigourously) $C - F(L) + \frac{C}{N} - 2\frac{A_0}{L}$.

Deduce from this that the estimation variance for one dimension (Ex. 13 § 2/) is in fact

$$\sigma^2_{Est} = \frac{C}{N} - \frac{A_o}{L} - \frac{A_1}{L^2}$$

Thus , for one dimension, the principle of composition of elementary extension variances gives the exact value of the first two terms, but not that of the third one.

5/ Conclude from the above that only the first two terms of the formula of paragraph 5/ Ex. 14 are valid (estimation variance of S by perfectly sampled drives).
To restore the exact formula, S will be supposed to be the rectangle $\ell \times h$ and the argument will be given directly on the variogram γ' deduced by grading from γ (under constant thickness). For h greater than the range, the function F'(h) of this one-dimensional transitive scheme is of the form

$$F'(h) = C' - \frac{A'_o}{h} + \frac{A'_1}{h^2}$$

(cf. Ex. 13). Show directly that the new constant C' is

$$C' = C - F(\ell,o) = \frac{A_o}{\ell} - \frac{A_1}{\ell^2}$$

To determine the new coefficients A'_o and A'_1 , proceed by identification with $F(\ell,h)$ (Ex. 14, par. 2/), which gives :

$$A'_o = \pi \frac{A_1}{\ell} - 2\frac{A_2}{\ell^2}$$

$$A'_1 = 2\frac{A_2}{\ell} - \frac{A_3}{\ell^2}$$

Show by applying the result of paragraph 4/ above that the estimation variance for drives of length ℓ and at equal level interval of $h = \frac{H}{N}$ is :

$$\sigma^2_{Est} = \frac{C'}{N} - \frac{A'_o}{H} - \frac{A'_1}{H^2}$$

$$\sigma^2_{Est} = \frac{A_o}{N\ell} - A_1\left(\frac{1}{N\ell^2} + \frac{\pi}{S}\right) + 2\frac{A_2}{S}\left(\frac{1}{\ell} - \frac{1}{H}\right) + \frac{A_3}{S^2}$$

Compare with Ex. 14, Paragraph 5/ .

2-11-5 Statistical inference for R.F.'s.

Exercise 16 (Estimation of a variogram) : Let $Y(x)$ be an I.R.F. on the straight line $\mathbb{R}^1$, $\gamma(h)$ its semi-variogram. A realization of $Y(x)$ is supposed to be known in the interval $(0,L)$ and to estimate $\gamma(h)$, the following estimator is formed :

$$\gamma^*(h) = \frac{1}{2(L-h)} \int_0^{L-h} [Y(x+h) - Y(x)]^2 \, dx$$

a/ Show that γ^* is an unbiased estimator : $E[\gamma^*(h)] = \gamma(h)$.

b/ (Lemma) Let X_1, X_2, X_3 and X_4 be four Gaussian r.v.'s with zero expectation, and let σ_{ij} be their covariance matrix. Show that

$$E(X_1 \, X_2 \, X_3 \, X_4) = \sigma_{12} \, \sigma_{34} + \sigma_{13} \, \sigma_{24} + \sigma_{14} \, \sigma_{23}$$

(proceed by identification of the term in $u_1 \, u_2 \, u_3 \, u_4$ in the characteristic function with four variables. This lemma can as well be admitted without demonstration).

c/ The increments of the I.R.F. $Y(x)$ are supposed to have Gaussian distributions (N.B. : the object of this hypothesis is simply to get the computation of moments of order 4). By using the lemma b/, establish the relationship :

$$E[(Y_{x+h}-Y_x)^2 \, (Y_{y+h}-Y_y)^2] = 4 \, \gamma^2(h) + 2 \, [\gamma(x-y+h) + \gamma(x-y-h) - 2\gamma(x-y)]^2$$

and deduce from this the variance $D^2[\gamma^*(h)]$ of the estimator $\gamma^*(h)$:

$$D^2[\gamma^*(h)] = \frac{\omega^2}{(L-h)^2} \int_0^{L-h} dx \int_0^x [\gamma(x-y+h) + \gamma(x-y-h) - 2\gamma(x-y)]^2 \, dy$$

(use the symmetry in x and y of the function to be integrated).

d/ Do the explicit computations for $\gamma(h) = \omega|h|$ (the process $Y(y)$ is then the brownian movement). Therefore, show first (with $h \geq 0$) that

$$D^2[\gamma^*(h)] = \frac{4 \, \omega^2}{(L-h)^2} \int_0^{L-h} dx \int_{\mathrm{Sup}(0,(x-h))}^x [h-x+y]^2 \, dy$$

and distinguish two cases :

For $h \leq L/2$ $\quad D^2[\gamma^*(h)] = \left(\frac{4}{3}\frac{h^3}{L-h} - \frac{1}{3}\frac{h^4}{(L-h)^2}\right)\omega^2$

For $h \geq L/2$ $\quad D^2[\gamma^*(h)] = \left(2h^2 + \frac{1}{3}(L-h)^2 - \frac{4}{3}h(L-h)\right)\omega^2$

In particular, when h is very small, the relative variance is :

$$\frac{D^2[\gamma^*(h)]}{[\gamma(h)]^2} \neq \frac{4}{3}\frac{h}{L}$$

It tends to 0 with h, and consequently statistical inference is possible in normal circumstances in so far as it concerns the behaviour of $\gamma(h)$ near the origin, and this is true even if we have only a single realization at our disposal. On the contrary, as soon as h is not very small compared with L, the relative variance becomes very large and statistical inference is no longer possible : the experimental variogram γ^* will "usually" differ considerably from its expectation $\gamma(h)$. For example, for $h = L/2$, the relative variance is equal to 1 (i.e. very large).

Exercise 17 (Estimation of the variance $\sigma^2(O/L)$). - In the same conditions as above, the variance of the punctual samples in L is estimated by

$$S^2 = \frac{1}{L}\int_0^L [Y(x) - \overline{Y}]^2\,dx \qquad \text{with} \qquad \overline{Y} = \frac{1}{L}\int_0^L Y(x)\,dx$$

a/ Put S^2 in the form $\frac{1}{L^3}\int_0^L\int_0^L\int_0^L (Y_x - Y_y)(Y_x - Y_{y'})\,dx\,dy\,dy'$ and show that $E(S^2) = \sigma^2(O/L)$.

b/ Compute the variance $D^2(S^2)$ when the increments of Y(x) are supposed to be Gaussian. To do this, deduce the expression for S^4 (which is a sextuple integral), use lemma b/ of Ex. 16 to obtain the argument of this integral, and show that the result can be expressed with the help of the auxiliary functions X and F in the form

$$D^2(S^2) = 2[F(L)]^2 + \frac{4}{L^2}\int_0^L (L-x)\,\gamma^2(x)\,dx - \frac{8}{L^3}\int_0^L x^2\,X^2(x)\,dx$$

$$- \frac{8}{L^3}\int_0^L x(L-x)\,X(x)\,X(L-x)\,dx$$

c/ Apply b/ to the case $\gamma(h) = \omega|h|$, and show that the relative variance on S^2 is equal to 4/5 (that is : very large) : the fluctuations of the experimental variance around its expectation always have a very large amplitude.

Exercise 18 (pseudo-covariance).- It is very often assumed without a thorough examination, that the Re. V;'s can be considered as realizations of stationary R.F.'s of order 2 (whereas very often a more accurate study shows that only the increments of this R.F. have moments of order 2 - in other words, very often there is no covariance C(h), but only a variogram $\gamma(h)$). Now it happens that the procedures of statistical inference which are used, as a consequence of the biases they produce, systematically make this hypothesis plausible, even when it is completely false.

We will only demonstrate this phenomenon by means of an example : that of the brownian movement with semi-variogram $\gamma(h) = |h|$, a realization of which is known in the interval (0,L). Thus we suppose that :

$$\begin{cases} E[Y(x+h) - Y(x)] = 0 \\ \\ \frac{1}{2} D^2[Y(x+h) - Y(x)] = |h| \end{cases}$$

Usually, to estimate the hypothetical covariance C(h), which in fact does not exist here, the (experimental) values of the following expressions are computed :

$$\overline{Y} = \frac{1}{L}\int_0^L Y(x)\,dx$$

$$C^*(x,y) = [Y(x) - \overline{Y}][Y(y) - \overline{Y}]$$

and the "experimental covariance" is deduced from them :

$$C^*(h) = \frac{1}{L-h}\int_0^{L-h} C^*(x+h,x)\,dx$$

Now with the semi-variogram $\gamma(h) = |h|$, the expectation of $C^*(x,y)$ is :

$$E[C^*(x,y)] = \frac{2}{3}L + \frac{x^2+y^2}{L} + 2\,\mathrm{Sup}(x,y)$$

For $h \geq 0$, we have thus

$$E[C^*(x+h,x)] = \frac{2}{3}L + \frac{x^2+(x+h)^2}{L} - 2x - 2h$$

and, by integrating with respect to x, we see that "the experimental covariance" has for an expectation :

$$E[C^*(h)] = \frac{1}{3} L - \frac{4}{3} h + \frac{2}{3} \frac{h^2}{L}$$

It is a <u>parabola</u>, with slope $-\frac{4}{3}$ at the origin. Hence, an apparent variance $C(0) = \frac{1}{3} L$ will be found, linked only to the length L of the segment considered, and which constitutes a pure artefact (since the true variance is infinite). Although there is no stationary covariance, but only a linear variogram, the biases introduced by this procedure of estimation result in an apparent confirmation of the existence of this covariance. It will be noted that the structure of the phenomenon is extremely distorted : not only is the straight line replaced by a parabola, but even <u>the slope at the origin is changed</u> ($\frac{4}{3}$ instead of 1). The $C^*(h)$ is a pure artefact, and retains almost nothing of the true structure of the phenomenon.;

It will be noted that the <u>experimental variogram</u> :

$$2\,\gamma^*(h) = \frac{1}{L-h} \int_0^{L-h} [Y(x+h) - Y(x)]^2 \, dx$$

has expectation

$$E[\gamma^*(h)] = \gamma(h) = |h|$$

and so is not changed by the preceding bias, and <u>consequently provides a safer method than the covariance</u>.

CHAPTER 3

KRIGING

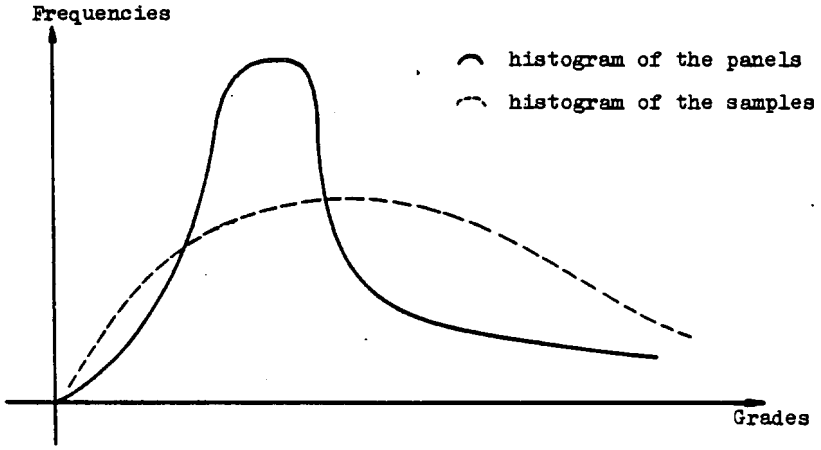
3-1 THE OBJECT OF KRIGING.

In mining terms, the problem of kriging is to find the best linear estimator possible of the grade of a panel, taking into account all the available information, i.e. the assay values of the different samples that have been taken, either inside or outside the panel we want to estimate. Kriging amounts to weighting, i.e. assigning a weight to the assay of each sample, these weights being computed in order to minimize the resultant estimation variance according to the geometrical characteristics of the problem (shapes, dimensions, and relative setting of the panel and the samples). Roughly, and naturally enough, kriging will assign a low weight to distant samples, and vice-versa. But this intuitive rule can only be partly true when more complex phenomena such as "screen effect" or "transfer of influence" appear. Of course, in order to solve a problem of kriging, i.e. to compute effectively the proper optimal weight which is to be assigned to each sample, it is necessary to make certain assumptions about the geostatistical characteristics of the orebody under study, i.e., essentially to give oneself the covariance function or the variogram of the R.F. of which the punctual assays are supposed to constitute a realization. In principle, it is not necessary to introduce a stationary or intrinsic hypothesis, and the kriging equations have a general scope.(In practice, of course, we have to begin with the estimation of the covariance or of the variogram, starting from the experimental data, and here the hypothesis in question returns with all its importance). In the following chapter, the general case of the non-stationary R.F. will be dealt with, and we shall limit ourselves for the moment to intrinsic or stationary of order 2 R.F's.

The first interesting thing about kriging comes from the definition itself. By minimizing the estimation variance, we are sure to make the best use of the available information or, in other words, to obtain the most precise possible estimation of the panel concerned. This is often a great advantage, but it would not always justify the additional complications necessarily introduced by weighting. By far, the most important practical point of kriging is not that it provides the best

estimation possible, but that in addition it avoids any systematic-errors. In most ore-bearing deposits, we select for exploitation a certain number of panels which are juged payable, and abandon others deemed unpayable. D.G. Krige has shown that a selection based only on inside samples necessarily resulted - on average- in an over-estimation of the selected panels.

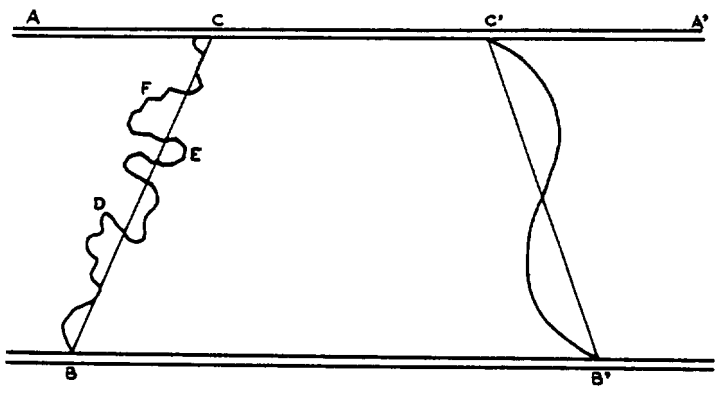
The reason for this general phenomenon is that the variance of the panel's true grades is always lower than the variance of the inside samples. In other words, the histogram of the actual grades of the panels always contains fewer extreme grades (rich or poor) and more intermediate values than the histogram deduced from the inside samples. If we compute what would be the effect of a selection of panels on this last histogram, the rejected panels would be in fact less poor than foreseen, and the chosen ones less rich.



The notion of kriging allows this phenomenon to be understood and its consequences corrected. For the very reason that we have selected a rich panel, the "aureole" of outside samples will have, in general, a lower grade than the inside ones ; yet its influence on the panel to be estimated is not negligible since kriging has assigned a non-zero weight to it. Failing to take this external aureole into account inevitably introduces a cause of systematic error - one of over-evaluation. To explain how this occurs, let us imagine a vein-type deposit developed by two drives AA' and BB'.

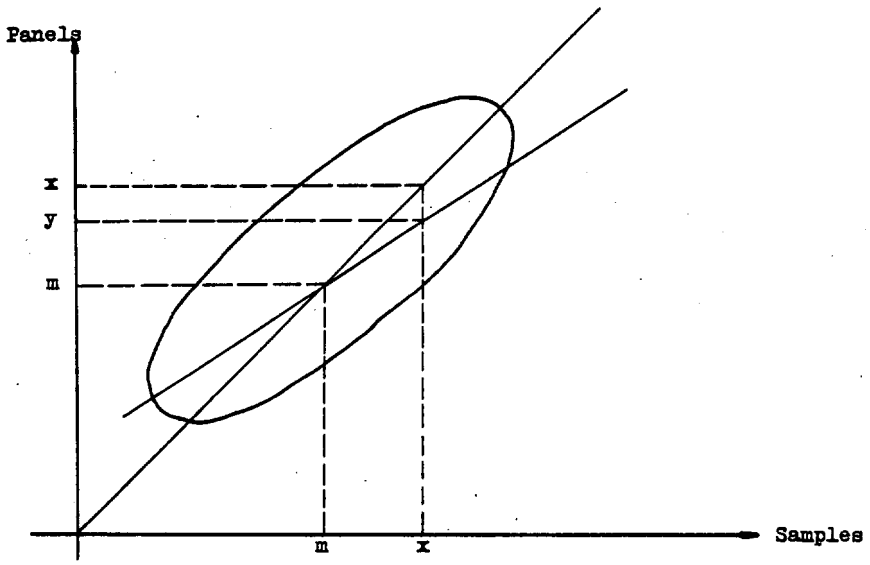
The grades of BB' are deemed payable, while those of AA' are payable only over a segment CC'. If we merely take into account the mean grades of BB' and CC', we are sure to make an error of over-estimation : the two segments AC and C'A' - although poor - have a not negligible effect on the grade of the trapezium BB', CC' selected for mining. If it were possible to draw the precise boundary between payable and unpayable ore, the real boundary would not lie along the straight lines BC or B'C' but would be, in general, a very irregular curve like B D E F ... Moreover, poor enclaves would re-

main inside the rich panel and vice-versa, so that while working we would have to abandon certain rich parts and mine some poor ones : thus, expressed in a concrete manner, is the influence of the poor segments AC and C'A' on the rich panel. (As this is a homogeneous orebody, and the existence of a precise boundary as defined by the grades seems doubtful, we shall speak of kriging. The word "smudging" is used when rich and poor parts, geologically heterogeneous, are separated by a continuous boundary which can be represented by a fairly regular curve such as B'D'E' and which, as before, need not follow B'C'). In this case, kriging consists in weighting the grades of BB', CC' and the two segments AC, C'A' by convenient coefficients, which would be, for example, 60% for BB', 27% for CC', and 13% for AC and C'A' .



The essential effect of such a weighting procedure is to eliminate - on average- any systematic error of over-evaluation, which is particularly dreadful. Compared with this prime objective, the improvement of the precision itself appears as a relatively minor result.

Let us now give some indications on the manner in which D.G. Krige has formulated the problem at the beginning of the fifties. South African miners had appreciated this phenomenon of the over-evaluation of rich panels and applied empirical corrective coefficients. To re-determine these coefficients, Krige started from the hypothesis that the sampling process was correct, in other words that the expectation of the sample assays taken inside the panel was equal to the true mean grade of this panel. Hence, if the variables are Gaussian (in the South African case, they were in fact lognormal, but this does not change anything essential), the regression line giving the con itional expectation of the sample as a function of the panel grade is identical to the first bisector.



The other regression line, giving the expectation of the panel as a function of the sample (and this is the interesting one), has thus a slope less than 1. If the sample assay x is greater than the general average m, the panel expectation y is thus inferior to x, and conversely. Hence Krige corrected this systematic error by using the equation

(3-1) $$y = m + \beta(x-m)$$

of this second regression line, with a regression coefficient $\beta < 1$. More precisely, if ρ is the correlation coefficient, σ_x and σ_y the standard deviations of the samples and the panels, the two regression coefficients (respectively 1 and β) are

$$1 = \rho \frac{\sigma_x}{\sigma_y} \quad , \quad \beta = \rho \frac{\sigma_y}{\sigma_x}$$

hence :

$$\beta = \frac{\sigma_y^2}{\sigma_x^2} < 1$$

The coefficient β is equal to the ratio of the variances (within the whole orebody) of the samples and the panels.

A more thorough examination of (3-1) suggests other possibilities. In general, the mean grade m is unknown, and is estimated by taking the arithmetic mean $m = \frac{1}{n} \Sigma X_i$ of the available samples

within the orebody. But if we replace m with m^*, relationship (3-1) takes the form :

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

of a linear combination of the grades X_i of the available samples, and with the condition :

$$\Sigma a_i = 1$$

to which we will return at length. But then, instead of attributing to the outside samples - near or distant - the same indifferentiable weight $a_i = \frac{1-a}{n}$, it seems natural to try to attribute to each one an appropriate weight a_i, taking into account its location with respect to the panel to be estimated. Thus we are led to the search for optimal weights a_i, i.e. to the notion of kriging as introduced at first.

3-2 NOTATION .

Throughout this chapter and the following one, the same system of notation will be used :

~ Z(x) will designate a (stationary or not) R.F. with a non-zero expectation in general, and Y(x) a (not necessarily stationary nor intrinsic) F.R. with zero expectation : for example, very often Y(x) will be

$$Y(x) = Z(x) - E[Z(x)]$$

~ S will designate a set of $\mathbb{R}^n$ on which a realization of the R.F. is known experimentally. In the case where S is a finite set, the experimental points will be designated by Greek indices, i.e. :

$$S = \{x_\alpha \ , \ \alpha = 1,2,\ldots N\}$$

~ As the case may be, the estimation of different values may be wanted : either the value $Z(x_o)$ of the R.F. at a point $x_o \notin S$, or the mean value $\frac{1}{V} \int_V Z(x)\, dx$ within a domain V different from S, or more generally (to condense all possible cases into a single script) a mean value

$$Z_o = \int p(dx)\ Z(x) \qquad (\int p(dx) = 1)$$

of Z(x) weighted by a measure p with sum 1 and with a support separate from S.

~ So as to do this, linear estimators will be formed starting from the available data, which will be the Z(x), $x \in S$. In the discrete case, we sill always denote

$$Z_\alpha = Z(x_\alpha) \qquad (x_\alpha \in S)$$

and in the same way, for functions f(x), C(x,y), etc... :

$$f_\alpha = f(x_\alpha) \quad , \quad C(x_\alpha, y_\beta) = C_{\alpha\beta} \quad \text{etc ...}$$

In the discrete case, the summation convention will always be applied : the summation must be performed on all indices appearing twice (in general, once in low or covariant position, and once again in the upper or contravariant position). Hence, $\lambda^\alpha Z_\alpha$ will be written instead of $\sum_\alpha \lambda^\alpha Z_\alpha$. The estimators which we will use will thus always be written as

$$Z^* = \lambda^\alpha Z_\alpha$$

When S is infinite, measures <u>with support in S</u> will be used to form the estimators :

$$Z^* = \int_S \lambda(dx)\ Z(x)$$

In general, this can even be written as $\int Z(x)\ \lambda(dx)$, for the measure λ has a support in S by definition.

Let us now set out the kriging equations, distinguishing three cases (stationary R.F. with zero or known expectation ; stationary R.F. with unknown expectation ; intrinsic R.F. without drift). <u>In all cases, the covariance or the variogram will be supposed to be known</u>.

3-3 STATIONARY R.F. WITH ZERO OR A PRIORI KNOWN EXPECTATION.

Let $Y(x)$ be a R.F. with zero expectation, $\sigma(x,y)$ its covariance, $S = \{x_\alpha\}$ the set of experimental points (supposed at the start to be finite), $Y_o = \int p(dx)\ Z(x)$ the variable to be estimated. As an estimator, we shall use the linear combination

$$Y_K = \lambda_K^\alpha\ Z_\alpha$$

and determine the coefficients λ_K^α from the conditions which minimize the expectation of $(Y_o - Y_K)^2$, i.e. (as the expectation of the R.F. is equal to 0) the variance $D^2[Y_o - Y_K]$. But this variance is :

$$E[Y_o - Y_K]^2 = D^2(Y_o) - 2\ \lambda_K^\alpha\ \sigma_{\alpha Y_o} + \lambda_K^\alpha\ \lambda_K^\beta\ \sigma_{\alpha\beta}$$

with :

$$\begin{cases} \sigma_{\alpha Y_o} = \int p(dx)\ \sigma(x_\alpha, x) \\ \\ D^2(Y_o) = \int\int p(dx)\ \sigma(x,y)\ p(dy) \end{cases}$$

Equating to zero the partial derivatives with respect to λ_K^α of this quadratic form, we get the system :

$$\lambda_K^\beta\ \sigma_{\alpha\beta} = \sigma_{\alpha Y_o} \qquad (3\text{-}2)$$

This system of N equations with N unknowns is regular, and has a unique solution if and only if the covariance matrix $\sigma_{\alpha\beta}$ is positive definite (thus with determinant > 0), which we will always assume. The variance σ_K^2 (or kriging variance) of this optimal estimate is equal to the value of the quadratic form $E(Y_o - Y_K)^2$ when we take the solution of (3-2) as the coefficients λ_K^α. But (3-2) involves :

$$\lambda_K^\alpha\ \lambda_K^\beta\ \sigma_{\alpha\beta} = \lambda_K^\alpha\ \sigma_{\alpha Y_o}$$

At the optimum, the cross and quadratic terms are thus equal, and we get :

$$\sigma_K^2 = D^2(Y_o) - \lambda_K^\alpha\ \sigma_{\alpha Y_o} \qquad (3\text{-}3)$$

In the case of punctual kriging (estimation of $Y(x_o)$ at $x_o \notin S$) the system is reduced to :

$$(3\text{-}4) \qquad \begin{cases} \lambda_K^\beta \, \sigma_{\alpha\beta} = \sigma_{\alpha, x_o} \\ \\ \sigma_K^2 = \sigma_{x_o, x_o} - \lambda_K^\alpha \, \sigma_{\alpha, x_o} \end{cases}$$

The solution depends of course on x_o, and the corresponding estimator :

$$Y_K(x_o) = \lambda_K^\alpha(x_o) \; Y_\alpha$$

is an <u>exact interpolator</u>, i.e. $Y_K(x_{\alpha_o}) = Y_{\alpha_o}$ if x_o coïncides with an experimental point $x_{\alpha_o} \in S$: this can be verified directly with the help of system (3-4), but it is a priori obvious that the estimator Y_{α_o} is optimal, for its variance is zero.

When the set S is <u>infinite</u>, we try to estimate Y_o with the help of an estimator of the form :

$$(3\text{-}5) \qquad Y_K = \int_S \lambda_K(dx) \; Y(x)$$

and the system (3-2) becomes

$$(3\text{-}6) \qquad \begin{cases} \forall \; y \in S \qquad \int_S \lambda_K(dx) \; \sigma(x,y) = \sigma_{y, Y_o} \\ \\ \sigma_K^2 = D^2(Y_o) - \int_S \lambda_K(dx) \; \sigma_{x, Y_o} \end{cases}$$

But here, some reservations are necessary. If <u>there exists</u> a measure λ_K with support on S which verifies (3-6), this solution is unique, and (3-5) is the optimal estimator. But such a measure λ_K does not necessarily exist (at least when the covariance is very regular, cf. Exercise 8). However, it can be shown (cf. [6]) that a unique optimal estimator Y_K always exists, which belongs to the <u>Hilbertian space</u> H(S) generated by the Y(x), $x \in S$ (i.e. : limit in the mean square of finite linear combinations of the Y(x), $x \in S$) . but Y_K does not necessarily have a representation of the form (3-5) with a measure λ_K .

This result can be explained by noticing that the systems (3-4) or (3-6) can be written as :

$$(3\text{-}7) \qquad \begin{cases} \text{Cov} \; [Y_K - Y_o, \; Y(y)] = 0 \qquad (\forall \; y \in S) \\ \\ \sigma_K^2 = D^2(Y_o) - \text{Cov} \; (Y_K \; , \; Y_o) \end{cases}$$

Hence Y_K is the only element of the Hilbertian space H(S) such that $Y_K - Y_o$ is orthogonal to every element $Y \in H(S)$ (i.e. with a covariance equal to 0 with every $Y \in H$). This property characterizes Y as being the orthogonal projection of Y_o in H(S), hence the existence and uniqueness, but obviously this does not imply (except in the finite case when H is a Euclidean space) the existence of an integral representation of the form (3-5).

Let Z(x) now be a R.F. with centred covariance $\sigma(x,y)$ and a constant expectation $m = E[Z(x)]$, non zero, but known. We are led immediately to the preceding case by using the R.F. $Y(x) = Z(x) - m$. Thus the optimal estimator is :

$$Z_K = m + \lambda_K^\alpha (Z_\alpha - m) \tag{3-8}$$

or

$$Z_K = m + \int_S \lambda_K(dx) [Z(x) - m] \tag{3-8'}$$

with coefficients λ_K^α or a measure λ_K verifying the same systems (3-2) or (3-6) , and the variance σ_K^2 still having the same expression.

3-4 STATIONARY RANDOM FUNCTION WITH UNKNOWN EXPECTATION.

3-4-1 Kriging equations.

Let Z(x) now be a R.F. with constant but unknown expectation m, and $\sigma(x,y)$ its centred covariance. $Z_o = \int p(dx)\ Z(x)$ is to be estimated from the Z_α, the values of the realization on a finite set $S = \{x_\alpha, \alpha = 1,2,\ldots N\}$, with the help of a linear combination of the form :

$$Z^* = \lambda^\alpha Z_\alpha$$

As the expectation m is unknown, it is necessary to impose on the coefficients λ^α the said condition of universality ::

$$\sum_\alpha \lambda^\alpha = 1 \tag{3-9}$$

Indeed, the best possible linear combination is the one that minimizes the expectation of $(Z_o - Z^*)^2$. We have

$$E(Z_o - Z^*)^2 = E(Z_o^2) - 2\,E(Z_o\,Z^*) + E(Z^*)^2 =$$

$$= m^2(1 - \sum_\alpha \lambda^\alpha)^2 + D^2(Z_o) - 2\,\lambda^\alpha\,\sigma_{\alpha Z_o} + \lambda^\alpha\,\lambda^\beta\,\sigma_{\alpha\beta}$$

As m is unknown, this expression can only be minimized if it does not depend on m, in other words if (3-9) is true.

This condition (3-9) can also be justified by requiring the estimator Z^* to be unbiased whatever the (unknown) value of m (universal estimator). As

$$E(Z_o - Z^*) = m - m \sum_\alpha \lambda^\alpha$$

again we find that this expectation is zero whatever m may be, if and only if (3-9) is true.

When the condition of universality (3-9) is verified, $E(Z_o - Z^*)$ is equal to 0, and consequently

$$E(Z_o - Z^*)^2 = D^2(Z_o - Z^*) = D^2(Z_o) - 2\,\lambda^\alpha\,\sigma_{\alpha, Z_o} + \lambda^\alpha\,\lambda^\beta\,\sigma_{\alpha\beta}$$

In expressing that this quadratic form is minimum, regard being had to condition (3-9), we get the following system containing a Lagrange parameter μ :

$$(3\text{-}10) \qquad \begin{cases} \lambda^\beta\,\sigma_{\alpha\beta} = \sigma_{\alpha, Z_o} + \mu \\ \\ \sum_\alpha \lambda^\alpha = 1 \end{cases}$$

Multiplying the first equation (3-9) and taking into account the second one, we get :

$$\lambda^\alpha\,\lambda^\beta\,\sigma_{\alpha\beta} = \lambda^\alpha\,\sigma_{\alpha, Z_o} + \mu \sum_\alpha \lambda^\alpha = \lambda^\alpha\,\sigma_{\alpha, Z_o} + \mu$$

Hence the expression of the kriging variance, where the Lagrange factor intervenes :

$$(3\text{-}10') \qquad D^2(Z^* - Z_o) = D^2(Z_o) + \mu - \lambda^\alpha\,\sigma_{\alpha, Z_o}$$

In the caze of punctual kriging at a point x_o, this system becomes :

$$(3\text{-}11) \qquad \begin{cases} \lambda^\beta \sigma_{\alpha\beta} = \sigma_{\alpha,x_o} + \mu \\ \sum_\alpha \lambda^\alpha = 1 \\ D^2[Z(x_o) - Z^*] = \sigma_{x_o x_o} + \mu - \lambda^\alpha \sigma_{\alpha,Z_o} \end{cases}$$

Here too it can be verified that punctual kriging constitutes an exact interpolator.

It can be shown that the system (3-10) is always regular. In the continuous case, there always existsa unique element $Z^* \in H(S)$, the limit in the mean square of finite linear combinations verifying condition (3-9), and such that :

$$\text{Cov}\ (Z^* - Z_o\ ,\ Z(y)) = 0 \qquad \forall\ y \in S$$

This unique element is the projection of Z_o in the closed linear variety defined in H(S) by the universality condition, hence the existence and uniqueness of the optimal estimator Z^*. Here also, Z^* has not necessarily a representation of the form $\int_S \lambda(dx)\ Z(x)$ with a measure λ with support in S verifying

$$(3\text{-}12) \qquad \begin{cases} \int_S \lambda(dx)\ \sigma(x,y) = \sigma_{y,Z_o} + \mu \qquad \forall\ y \in S \\ \int_S \lambda(dx) = 1 \end{cases}$$

But, if we can find a measure λ obeying (3-12), then $Z^* = \lambda(dx)\ Z(x)$ is the unique solution of the problem.

3-4-2 Optimal estimation of m.

Instead of estimating a spatial mean of the type $\int p(dx)\ Z(x)$, we can also try to estimate the expectation $m = E[Z(x)]$ itself. We are going to determine the estimator of m and, in the following paragraph, to examine the relation existing between m^* and Z^*. We shall limit ourselves to the case where S is finite ; the extension to the infinite case can then be deduced very

easily, with the usual reservations about the existence of a representation of our estimator, using measures having support in S.

To estimate m, we take the linear combination

$$m^* = \lambda_o^\alpha \; Z_\alpha$$

We next impose on the coefficients λ_o^α the universality condition

$$\Sigma \; \lambda_o^\alpha = 1$$

which expresses that m^* is without bias, whatever the unknown value of m, and the coefficients λ_o^α are chosen so as to minimize $E(m-m^*)^2 = D^2(m^*)$ by taking this condition into account. From

$$D^2(m-m^*) = \lambda_o^\alpha \lambda_o^\beta \sigma_{\alpha\beta}$$

we deduce that the λ_o^α constitute the unique solution of the following system, in which there appears a Lagrange factor μ_o :

$$\text{(3-13)} \qquad \begin{cases} \lambda_o^\beta \sigma_{\alpha\beta} = \mu_o \\ \\ \sum\limits_\alpha \lambda_o^\alpha = 1 \end{cases}$$

At the optimum, we get $\lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} = \mu_o \sum\limits_\alpha \lambda^\alpha = \mu_o$ so that the Lagrange factor μ_o is equal to the variance of the optimal estimator

$$\text{(3-13')} \qquad D^2(m^*) = \mu_o$$

3-4-3 The additivity theorem.

At the end of paragraph 3-3, we have indicated how to form the optimal estimator when the expectation is not zero but is known. The optimal coefficients λ_K^α are computed for the

case m = 0 , and a very simple correction, which is to replace Y(x) by Z(x) - m, is made :

$$Z_K = m + \lambda_K^\alpha \, (Z_\alpha - m) \tag{3-8}$$

Let us show that the optimal estimator Z^* for the case when m is unknown (paragraph 3-4-1) and the optimal estimator m^* of m itself are related by the following relationship obtained by taking (3-8):

$$Z^* = m^* + \lambda_K^\alpha \, (Z_\alpha - m^*) \tag{3-14}$$

with the same coefficients λ_K^α , the solution of the system (3-2). This result has the following significance : we can krige as if m was known, on the condition that we replace the real unknown value m by its optimal estimate m^*.

We shall consider the right hand side of (3-14) :

$$m^* + \lambda_K^\alpha \, (Z_\alpha - m^*) = [\lambda_K^\alpha + \lambda_o^\alpha \, (1 - \sum_\beta \lambda_K^\beta] \, Z_\alpha$$

The quantities

$$\lambda'^\alpha = \lambda_K^\alpha + \lambda_o^\alpha \, (1 - \sum_\beta \lambda_K^\beta)$$

obey the universality condition, for :

$$\sum_\alpha \lambda'^\alpha = \sum_\alpha \lambda_K^\alpha + \sum_\alpha \lambda_o^\alpha \, (1 - \sum_\beta \lambda_K^\beta) = 1$$

taking into account $\sum_\alpha \lambda_o^\alpha = 1$ (system (3-13)).

Let us then form the expression :

$$\lambda'^\beta \, \sigma_{\alpha\beta} = \lambda_K^\beta \, \sigma_{\alpha\beta} + (1 - \sum_\gamma \lambda_K^\gamma) \, \lambda_o^\beta \, \sigma_{\alpha\beta}$$

From (3-2) and (3-13), we get :

$$\lambda'^\beta \, \sigma_{\alpha\beta} = \sigma_{\alpha, Z_o} + (1 - \sum_\gamma \lambda_K^\gamma) \, \mu_o$$

Hence, the λ'^β do follow the first relationship (3-11), with the Lagrange factor

(3-15) $$\mu = (1 - \sum_{\gamma} \lambda_K^{\gamma})\ \mu_o$$

From the uniqueness of the solution, we do indeed have $\lambda'^{\alpha} = \lambda^{\alpha}$, and relationship (3-14) follows from it.

Concerning the variances, we have also an additivity relationship. From (3-14), we have :

$$Z^* - Z_o = (\lambda_K^{\alpha} Z_{\alpha} - Z_o) + (1 - \sum_{\beta} \lambda_K^{\beta})\ m^*$$

But system (3-2) expresses precisely that $\lambda_K^{\alpha} Z_{\alpha} - Z_o$ has a zero covariance with every Z_{β}, thus also with all their linear combinations and in particular, with m^*. Hence we have :

$$D^2(Z^* - Z_o) = D^2(\lambda_K^{\alpha} Z_{\alpha} - Z_o) + (1 - \sum_{\beta} \lambda_K^{\beta})^2\ D^2(m^*)$$

i.e.

(3-16) $$D^2(Z^* - Z_o) = \sigma_K^2 + (1 - \sum_{\beta} \lambda_K^{\beta})^2\ D^2(m^*)$$

The first term is the kriging variance σ_K^2 when m is known. The second one gives an exact measure of the loss of accuracy we suffer relative to Z_o by not knowing the true value of m.

3-5 CASE OF AN I.R.F. WITHOUT COVARIANCE.

Let Z(x) be an I.R.F. without drift, having a variogram γ but no covariance. To estimate $Z_o = \int p(dx)\ Z(x)$ with $\int p(dx) = 1$, we look for a linear combination

$$Z^* = \lambda^{\alpha} Z_{\alpha}$$

such that :

a/ the "error" $Z^* - Z_o$ is an authorized linear combination (paragraph 2-2-1), i.e. one having a finite variance.

b/ this estimation variance is minimum.

Condition a/ gives $\sum_\alpha \lambda^\alpha - \int p(dx) = 0$, i.e. the same universality condition as in the preceding paragraph :

$$\sum_\alpha \lambda^\alpha = 1$$

When this condition is verified, we know from the mechanism of paragraph (2-2-1) that the variance of $Z^* - Z_o$ can be computed as if there existed a covariance equal to $-\gamma$. Thus the system (3-11) can be transposed directly : the coefficients λ^α of the optimal estimator constitute the unique solution of the system

$$\text{(3-17)} \qquad \begin{cases} \lambda^\beta \gamma_{\alpha\beta} = \int p(dx)\ \gamma(x, x_\alpha) - \mu \\ \sum_\alpha \lambda^\alpha = 1 \end{cases}$$

and the corresponding variance σ_K^2 is :

$$\text{(3-17')} \qquad \sigma_K^2 = -\iint p(dx)\ \gamma(x,y)\ p(dy) + \mu + \lambda^\alpha \int p(dx)\ \gamma(x, x_\alpha)$$

In the continuous case (with the usual restrictions about the existence of a representation of Z^* by means of a measure), we have likewise :

$$\text{(3-18)} \qquad \begin{cases} \int_S \lambda(dx)\ \gamma(x,y) = \int p(dx)\ \gamma(x,y) - \mu \qquad (\forall\ y \in S) \\ \int_S \lambda(dx) = 1 \\ \sigma_K^2 = -\iint p\ \gamma\ p + \mu + \iint p\ \gamma\ \lambda \end{cases}$$

Finally, in the punctual case, we still obtain an exact interpolator.

Note - Let $\lambda^\alpha(x_o)$ be the solution of the kriging of the point x_o :

$$\text{(3-19)} \qquad \begin{cases} \lambda^\beta(x_o)\ \sigma_{\alpha\beta} = \gamma(x_o, x_\alpha) - \mu(x_o) \\ \sum_\alpha \lambda^\alpha(x_o) = 1 \\ \sigma_K^2(x_o) = \mu(x_o) + \lambda^\alpha(x_o)\ e(x_o, x_\alpha) \end{cases}$$

From the linear character of the right hand side of (3-17), the solution of the kriging of $\int p(dx)\ Z(x)$ is :

$$\lambda^{\alpha} = \int p(dx)\ \lambda^{\alpha}(x)$$

There is superimposition, or linear combination of the punctual krigings : this relationship applies also to the Lagrange factor :

$$\mu = \int p(dx)\ \mu(x)$$

On the other hand, the variance (3-17') cannot be obtained by linear combination of the variances $\sigma_K^2(x_0)$ of the punctual krigings.

3-6 EXERCISES ON KRIGING.

Exercise 1 (Kriging of a segment of length ℓ) - a/ Consider, on the real line $\mathbb{R}$, an I.R.F. without covariance with semi-variogram $\gamma(h)$ and four sampling points x_1, $x_2 = x_1 + \ell$, $x_3 = x_2 + \ell$, $x_4 = x_3 + \ell$. We want to krige the segment (x_2, x_3) of length ℓ from the values Z_1, Z_2, Z_3, Z_4 of the realization at the points x_1, x_2, x_3, x_4. Write the system (3-17) with the help of the auxiliary functions X and F of paragraph 2-5-2.
[Show that $\lambda_1 = \lambda_4 = \frac{\lambda}{2}$, $\lambda_2 = \lambda_3 = \frac{1-\lambda}{2}$, which is obvious by symmetry, with λ the solution of :

$$\frac{\lambda}{2}\gamma(\ell) + \frac{1-\lambda}{2}\gamma(\ell) + \frac{\lambda}{2}\gamma(2\ell) = X(\ell) + \mu$$

$$\frac{1-\lambda}{2}\gamma(\ell) + \frac{1-\lambda}{2}\gamma(2\ell) + \frac{\lambda}{2}\gamma(3\ell) = 2\,X(2\ell) - X(\ell) + \mu$$

Eliminate the Lagrange parameter μ, hence :

$$\lambda\left[\frac{1}{2}\gamma(3\ell) - \gamma(2\ell) - \frac{1}{2}\gamma(\ell)\right] = 2\,X(2\ell) - 2\,X(\ell) - \frac{1}{2}\gamma(2\ell)$$

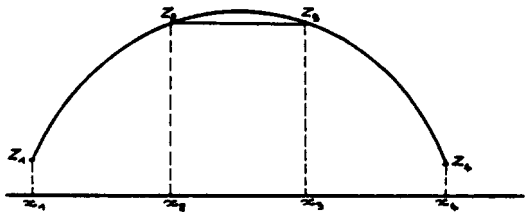
b/ Application to the case $\gamma(h) = |h|^{\alpha}$ $(0 \leq \alpha < 2)$ and interpretation.

(Solution : use Exercise 5 of Chapter 2. We get

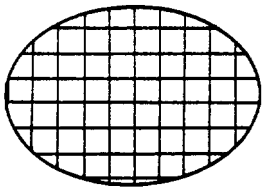
$$\lambda = \frac{2^{\alpha} - \frac{4}{\alpha+1}(2^{\alpha}-1)}{1 + 2^{\alpha+1} - 3^{\alpha}}$$

For $\alpha = 0$, $\lambda = \frac{1}{2}$ (pure nugget effect, the optimal estimator is the arithmetic mean of the four samples). When α increases, λ decreases, is equal to 0 when $\alpha = 1$ (Markovian property of the linear variogram), and becomes negative when $1 < \alpha < 2$: when $\alpha > 1$, the realization shows good continuity and the figure below explains why ; for $\frac{Z_1+Z_4}{2} < \frac{Z_2+Z_3}{2}$ (for example), the estimator must be greater than $\frac{Z_2+Z_3}{2}$, thus $\lambda < 0$:

(Compare with Exercise 12, Chapter 4)



Exercise 2 (Orebody developed by drifts and raises).- Consider an orebody developed by two regular grids of drifts and raises, parallel respectively to two perpendicular directions.



Let Z_M and Z_T be the mean grades of the drifts and of the raises, Z be the unknown grade of the orebody and $\sigma_M^2 = D^2(Z-Z_M)$ and $\sigma_T^2 = D^2(Z-Z_T)$ be the estimation variances of the orebody by the drifts and by the raises. We assume (this is true with good approximation) that $E[(Z-Z_M)(Z-Z_T)] = 0$.

a/ Show that the optimal estimator (kriging) of Z is $Z^* = \lambda Z_M + (1-\lambda) Z_T$ with $\lambda = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_M^2}$ (weighting by the reciprocal of the respective estimation variances) and that the corresponding variance is $\frac{\sigma_M^2 \sigma_T^2}{\sigma_T^2 + \sigma_M^2}$.

(the coefficients are of the form λ and $\lambda-1$ because of the universality condition . Start with

$$D^2(Z-Z^*) = D^2[\lambda(Z-Z_M) + (1-\lambda)(Z-Z_T)] = \lambda^2 \sigma_M^2 + (1-\lambda)^2 \sigma_T^2$$

and minimize with respect to λ).

b/ Application to the De Wijsian case when the length of the drifts and raises are greater than their equidistance.

[Call L_M and L_T the (total) length of the drifts and raises, and use the relationship $\sigma_E^2 = \alpha \frac{\pi}{2} \frac{S}{L^2}$ of 2-8-2. Then $\lambda = \frac{L_M^2}{L_M^2 + L_T^2}$ (weighting by the squares of the lengths, and $\sigma_K^2 = \alpha \frac{\pi}{2} \frac{S}{L_M^2 + L_T^2}$).

Exercise 3 (Markovian property of the exponential covariance and of the linear variogram).- This preparatory exercise may help to understand the following exercises.

a/ A R.F. Z(x) on the real line ($\mathbb{R}$) is Markovian if, for any x_o , the Z(x), $x > x_o$ and the Z(x'), $x' < x_o$ are conditionally independent when $Z(x_o)$ is given. Show that a stationary R.F. of order 2 with Gaussian distribution is Markovian if and only if its centred covariance is of the form $C\, e^{-a|h|}$.

[Two Gaussian r.v.'s are independent when their correlation coefficient is equal to 0. If X_1, X_2, X_3 are Gaussian, the correlation coefficient of X_1 and X_3 for X_2 fixed is : $\frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{(1-\rho_{12}^2)(1-\rho_{23}^2)}}$ Deduce therefrom that the R.F. is Markovian if and only if $C(h+h') = C(h)\, C(h')$ (h, h' positive)]

b/ Let Z(x) be an I.R.F. on the real line. Z(x) is said to have independent increments if the $Z(x_i) - Z(y_i)$ are independent when the intervals (y_i, x_i) are disjoint or have a single common point. Show that an I.R.F. with Gaussian increments has independent increments if and only if its variogram is linear (the I.R.F. is then a Wiener-Levy process or brownian motion).

[a simple calculation shows that the increments are without correlation of and only if the variogram is linear]

Exercise 4 (exponential covariance, continuous case). - a/ On a line, let Y(x) be a R.F. with zero expectation and $C(h) = e^{-a|h|}$ be its covariance. The realization of Y(x) is known in the interval (-R, +R). Krige the point $x_o = R + h$ ($h > 0$).

[If Y(x) is Gaussian, we can deduce from the Markovian property of Exercise 3 that the optimal estimator is $e^{-a|h|}Y(r)$ with variance $1 - e^{-2a|h|}$. This property, related to the only covariance, remains if Y(x) is not Gaussian. It can be shown directly that

the measure $\lambda_K = e^{-a|h|}\ \delta_R$ (Dirac placed at R) is indeed the solution of (3-6) for any $y \leq R$, and σ_K^2 can be computed directly as well.

b/ Let Z(x) be a R.F. with unknown expectation m and $e^{-a|h|}$ its centred covariance. Show that the optimal estimator m^* of m (when the realization is known on (-R, +R) verifies :

$$\begin{cases} m^* = \frac{aR}{1+aR}\ \overline{Z} + \frac{1}{1+aR}\ \frac{Z_R + Z_{-R}}{2} & \left(\overline{Z} = \frac{1}{2R}\int_{-R}^{R} Z(x)\ dx\right) \\ D^2(m^*) = \frac{1}{1+aR} \end{cases}$$

[establish the relationships $\frac{1}{2}\ (\delta_R+\delta_{-R})\ e^{-a|x-y|} = e^{-aR}\ \mathrm{chay}$ and $\int_{-R}^{R} e^{-a|x-y|}\ dx = \frac{2}{a} - \frac{2}{a}\ e^{-aR}\ \mathrm{chay}$ for $-R \leq y \leq R$. Deduce from this that the measure $\nu_o = \frac{a}{2}\ dx + \frac{1}{2}(\delta_R+\delta_{-R})$ verifies $\int_{-R}^{R} \nu_o(dx)\ e^{-a|x-y|} = 1$ and that the measure $\lambda_o = \frac{1}{1+aR}\ \nu_o$ is the solution required with the Lagrange parameter $\mu_o = D^2(m^*) = \frac{1}{1+aR}$]

c/ With the same conditions as in b/ , krige the point $x_o = R + h$ $(h > 0)$ by applying the additivity relationship.

$$[Z^* = e^{-a|h|}\ Z_R + (1- e^{-a|h|})m^*\ ;\ D^2(Z_{x_o} - Z^*) = 2 - e^{-2ah} + \frac{(1 - e^{-ah})^2}{1+aR}\ ]$$

Exercise 5 (exponential covariance, discrete case).- a/ Let Y(x) be a R.F. distributed along a line with zero expectation and a stationary covariance $e^{-a|h|}$. The realization is known at the n+1 points of abscissae 0, 1,... n. Compute the punctual kriging of the point $x_o = i + \varepsilon$ $(0 < \varepsilon < 1)$ $(i < n)$.

[From the Markovian property, we must look for a solution of the form $Y_K = \lambda\ Y_i + \lambda'\ Y_{i+1}$. Show that the system (3-4) is true for $\lambda_j = 0$ $(j \neq i,\ j \neq i+1)$, $\lambda_i = \lambda = \frac{\mathrm{Sh}(1-e)a}{\mathrm{Sha}}$, $\lambda_{i+1} = \lambda' = \frac{\mathrm{Sh}\ \varepsilon a}{\mathrm{Sha}}$]

b/ With the same conditions, let Z(x) be a R.F. of covariance $e^{-a|h|}$ and unknown expectation m. Show that the optimal estimator is of the form

$$m^* = (1-b)\ \overline{Z} + b\ \frac{Z_o+Z_n}{2} \qquad \left(\overline{Z} = \frac{1}{n+1}\ \sum_{i=0}^{n}\ Z_i\right)$$

with $b = \dfrac{2}{(n+1)\,e^{a} - (n-1)}$, $\mu_o = D^2(m^*) = \dfrac{1 + a}{(n+1)\,e^{a} - (n-1)}$

[First establish the relationship $\sum_{j=0}^{n} e^{-a|i-j|} = \dfrac{e^{a}+1}{e^{a}-1} - \dfrac{e^{-ai} + e^{-(n-1)a}}{e^{a}-1}$; look for a solution of the form $\lambda^{\beta} = A + B(\delta_o^{\beta} + \delta_n^{\beta})$ for the system (3-13) and proceed by identification].

c/ With the same conditions, krige a point x_o between i and i+1 when m is unknown.

[Apply the additivity relationship].

Exercise 6 (linear variogram).- a/ Consider, along a line, an I.R.F. with no drift and $\gamma(h) = |h|$ its variogram. The realization is known at the point x = R, and at a finite or infinite number of points < R. Show that the kriging of $x_o = R+h$ (h > 0) is Z(R) itself with $\sigma_K^2 = 2|h|$.

[this solution is suggested by the Markovian property of |h|. Verify that the measure $\lambda(dx) = \delta_R$ is the true solution of system (3-18) for any y < R] .

b/ The realization is now supposed to be known at two points x_1 and x_2 and at any number of points outside the interval (x_1,x_2). Show that the kriging of x_o between x_1 and x_2 is :

$$Y^*(x_o) = \frac{x_2-x_o}{x_2-x_1}\ Y(x_1) + \frac{x_o-x_1}{x_2-x_1}\ Y(x_2) \qquad (x_1 \le x_o \le x_2)$$

$$\sigma_K^2 = 2\,\frac{(x_2-x_o)(x_o-x_1)}{x_2-x_1}$$

[The Markovian property suggests a solution of the type $\lambda\, Y_1 + (1-\lambda)\, Y_2$. Show that with $\lambda = \dfrac{x_2-x_o}{x_2-x_1}$ (linear interpolation between the two nearest known points), the measure $\nu = \delta_{x_1} + (1-\lambda)\,\delta_{x_2}$ verifies $\int \nu(dx)|x-y| = |x_o-y|$ when $y \le x_1$ or $y \ge x_2$. The Lagrange factor is equal to 0, and the computation of σ_K^2 follows easily].

Exercise 7 (linear variogram or nugget effect). - On the segment (-R,+R) a realization of an I.R.F. is known, with a variogram which comprises a linear term and a nugget term. As we are in the continuous case, we will take $\gamma = |h| - C\delta$. (nugget effect represented by a Dirac measure). Krige the point $x_o = R+h$ (h > 0). To do this :

a/ Establish the relationships $\int_{-R}^{R} |x-y|\ \mathrm{chax}\ dx = \dfrac{2R}{a}\ \mathrm{sha}\ R + \dfrac{2}{a^2}\ \mathrm{chay}$, $\int_{-R}^{R} |x-y|\ \mathrm{shax}\ dx$

$= - 2 \frac{\text{cha } R}{a} y + \frac{2}{a^2}$ shay. Hence deduce that with $a = \sqrt{\frac{2}{C}}$:

$$\int_{-R}^{R} (|x-y|dx - C\delta(dx)) \text{ chax} = \frac{2R}{a} \text{ sha } R, \qquad \int_{-R}^{R} (|x-y|dx - C\delta(dx)) \text{ shax} = - 2 \frac{\text{cha } R}{a} y$$

b/ Show that the function $\lambda(x) = \frac{a \text{ shax}}{2 \text{ chaR}} + \frac{a \text{ chax}}{2 \text{ shaR}}$ verifies the system :

$\int_{-R}^{R} \lambda(x)dx = 1$, $\int_{-R}^{R} (|x-y|dx = C\delta(dx)) \lambda(x) = - y + R$. Hence deduce that the optimal estimator is $Z^*(x_o) = \int_{-R}^{R} \lambda(x) Z(x)dx$, with the Lagrange parameter $\mu = x_o - R = h$ and $\sigma_K^2 = 2h + \frac{1}{a} \text{ cha } R$.

[Note that the existence of a nugget effect removed the Markovian property. The density $\lambda(x)$ gives positive weights to all the samples. The nugget effect is said to raise the screens].

Exercise 8 (Gaussian exponential). - When the covariance is very regular, it may be the case that the kriging solution cannot be represented by a measure. For example, consider on a line, a stationary R.F. of covariance $C(h) = e^{-\frac{h^2}{2}}$ and zero expectation. We want to krige the point $x_o = R + h$, $h > 0$, knowing the realization on $(-R,+R)$. If the optimal estimator is of the type $Y_K = \int_{-R}^{+R} \lambda(dx) Y(x)$, the measure λ verifies

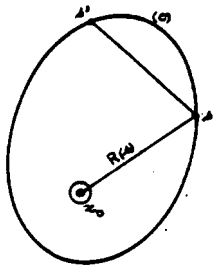
$$\int_{-R}^{R} \lambda(dx) \quad e^{-\frac{(x-y)^2}{2}} = e^{-\frac{(x_o-y)^2}{2}} \qquad \forall y \in (-R, +R)$$

Show that a measure λ obeying this relationship cannot exist.

[The measure $\nu(dx) = e^{-\frac{x^2}{2}} \lambda(dx)$ should verify $\int_{-R}^{R} e^{xy} \nu(dx) = e^{-\frac{x_o^2}{2} + x_o y}$. According to the properties of the Laplace transformation, the only measure obeying this relationship would be $e^{-x_o^2/2} \delta_{x_o}$, but its support is not in $(-R,+R)$].

Exercise 9 (De Wijsian kriging in $\mathbb{R}^2$). - In $\mathbb{R}^2$, the De Wijsian scheme has a property of Markovian character : when the realization is known on a closed contour C, there is no correlation between the inside and the outside of C : when kriging an element inside C, the knowledge of information outside C does not change the solution. There is a total screen effect, which an electrical analogy allows us to understand (log r is the harmonic potential of the plane, and plays the same rôle in $\mathbb{R}^2$ as the Newtonian potential

$1/r$ in $\mathbb{R}^3$).



a/ Let $Z(x)$ be a De Wijsian I.R.F. in $\mathbb{R}^2$, C a closed contour on which the realization is known; s the curvilinear abscissa of a point of C, $r(s,s')$ the distance between two points of abscissae s and s' on C. We want to krige a small circle centred at a point x_0 inside C. Let R(s) be the distance between x_0 and a moving point on C. Show that the optimal estimator is $Z^* = \int_C \lambda(s)\, Z(s)$ with a function $\lambda(s)$ such that :

$$\begin{cases} \int_C \lambda(s) \log r(s,s')\, dr = \log R(s') + \mu \\ \int \lambda(s)\, dx = 1 \end{cases}$$

Hence deduce that $\lambda(s)$ is the <u>density</u> induced by a mass -1 placed at x_0 (when the potential is log r) in other words that (C) is an equipotential curve when a mass -1 is placed at x_0 and the density $\lambda(s)$ on C.

We know that the potential is constant not only on C but also outside C. Hence, deduce that if y is any point outside C, we have still

$$\int_C \lambda(s) \log r(s,y) ds = \log r(x_0,y) + \mu$$

and that the density $\lambda(s)$ still gives the kriging solution when the realization is known on C and on any set S' inside of C (total screen effect).

b/ The Green function $G(x)$ relative to the point x_0 is characterized by the two following properties : 1/ $G(x)$ is constant on C

2/ $G(x) - \log r(x_0,x)$ is a regular and harmonic function at any point x inside C, including x_0. We know that the induced density $\lambda(x)$ is given by $\frac{1}{2\pi} \frac{dG}{dn}$ (derivative with respect to the positive normal).

Hence deduce the kriging of a point x_0, knowing the realization on a straight line. [The Green function is $\log \frac{r}{r'}$, cf. figure a, and $\lambda(x) = \frac{1}{\pi} \frac{h}{x^2+h^2}$].

c/ Same question for an inside point x_0, knowing the realization on a circle of radius R.

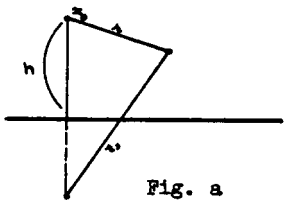


Fig. a

[here $G = \log \frac{r}{r'}$, with x_o' the inverse of x_o in an inversion of modulus R^2]. Use polar coordinates :

$$G = \frac{1}{2} \log \frac{\rho^2+a^2 - 2\,a\,\rho\cos\theta}{\rho^2 + \frac{R^4}{a^2} - 2\frac{R^2}{a}\rho\cos\theta}$$

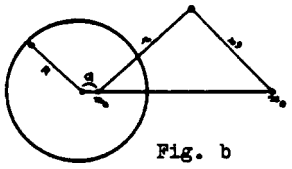


Fig. b

We get the density

$$\lambda(\theta) = \frac{R^2-a^2}{2\pi R}\frac{1}{r^2} = \frac{R^2-a^2}{2\pi R}\frac{1}{\rho^2 + a^2 - 2\,a\,\rho\cos\theta}\ ]$$

Exercise 10 (Kriging for large grids) (cf. Ex. 13 to 15, Chapter 2).- A surface S is to be kriged from n+k samples, with k samples taken inside S and n outside. The distance between two distinct samples, and the distance between each sample and the boundary of S are supposed to be greater than the range. See notations of Exercises 13 to 15, Chapter 2.

a/ Write the kriging equations. Show that they reduce to :

$$C\,\lambda_i = \sigma_{si} + \mu$$
$$\Sigma\,\lambda_i = 1$$

with $\sigma_{si} = 0$ for the outside samples i and $\sigma_{si} = \pi\frac{A_1}{S}$ for the inside samples.

Deduce that the k inside samples have the same weight λ and the n outside samples the same weight λ' with :

$$\lambda = \frac{1}{n+k} + \frac{n}{n+k}\pi\frac{A_1}{CS}$$

$$\lambda' = \frac{1}{n+k} - \frac{k}{n+k}\pi\frac{A_1}{CS}$$

b/ Compute the kriging variance :

$$(\sigma_K^2 = C - F(S) - \frac{2k}{n+k}\pi\frac{A_1}{S} - \frac{nk}{n+k}(\pi\frac{A_1}{S})^2 \times \frac{1}{C} + \frac{C}{n+k})$$

In particular, when S is great, we have the approximation formula :

$$\sigma_K^2 \sim \frac{C}{n+k} + \frac{n-k}{n+k}\pi\frac{A_1}{S}$$

c/ Compare these results with formula (3-1) of paragraph 3-1, and show that in the case of large grids, the optimal estimator is the same as that obtained initially by Krige with his linear regression method.

[The regression coefficient is $\beta = \frac{\sigma_y^2}{\sigma_x^2} = \frac{C-F(S)}{C/k} = \frac{k}{C} \frac{\pi A_1}{S}$. When m is known, the optimal estimator is identical to that given by the linear regression :

$$Z_K = m + \beta \left(\frac{1}{k} \sum_i X_i - m \right)$$

X_i grade of the inside samples. From (3-4-3), we have to take

$$Z^* = m^* + \beta \left(\frac{1}{k} \sum_i X_i - m^* \right)$$

with $m^* = \frac{1}{n+k} \left(\sum_i X_i + \sum_j X_j \right)$ (Y_j outside samples). Thus, we get λ and λ' as before].

Compute the variance from the additivity relationship.

[With m known, the variance is $\sigma_y^2 (1-\rho^2)$. Show that $\rho^2 = \beta = \frac{k}{C} \pi \frac{A_1}{S}$. Hence deduce that :

$$\sigma_K^2 = \sigma_y^2 (1-\beta) + (1-\beta)^2 D^2(m^*) = (1-\beta) \pi \frac{A_1}{S} + (1-\beta)^2 \frac{C}{n+k}$$

and verify that this result is the same as in paragraph b/].

CHAPTER 4

UNIVERSAL KRIGING

4-1 INTRODUCTION

4-1-1 Review of least squares methods.

In this section, our object will be to set out in terms of non-stationary random functions the problem of estimating drifts[1] (or trends), and to show that this problem has an optimal solution which is different from those produced by the methods of "trend surface analysis". The latter, which consist in general in adjusting a polynomial by a least squares method, create a feeling of unease ; indeed, they produce the impression that nature has been violated by having a polynomial imposed by force, and there is no a priori reason why a polynomial should represent the actual structure of the phenomenon to be represented by it. The three main criticisms of the least square methods are :

Firstly, the methods often show a confusion between the concept and the mode of operation. Few authors bother to define what they mean by the "trend" which they are going to estimate by a least squares method. The impression is often created that this famous trend is nothing more than a numerical result brought about by the mode of operation - i.e perhaps a pure and simple artefact. Upon reading the literature refering to trend, it seems that it can refer to at least three distinct ideas : if $Z(x)$ is the regionalized variable concerned, and $P(x)$ is the polynomial adjusted by the method of least squares to fit the known experimental values at points x_1, x_2,..., the value of x in $P(x)$ can be given one or the other of the following incompatible meanings :

a/ $P(x)$ is (an estimation of) the a priori expectation $E[Z(x)]$: it is always in this sense that the term "drift" will be used here.

b/ $P(x)$ is an estimate of the true value (unknown) $Z(x)$ taken by the regionalized variable at point x : it is in this sense that the term "punctual kriging" will be used here. More precisely, punctual kriging will give the best linear estimator for $Z(x)$ from the given data $Z(x_1)$, $Z(x_2)$....

1 - In order to avoid any anthropomorphical connotation, we shall always use the term "drift" instead of "trend".

c/ Lastly, $P(x)$ is sometimes given the meaning "moving average". $P(x_o)$ would thus be (an estimate of) the mean value of $Z(x)$ in an area of variable size (to be specified) surrounding the given point x_o. Here, kriging will designate the best linear estimator of this moving average.

These distinctions are very important theoretically and practically. It is essential to define precisely whether a/, b/, or c/ above is what is meant at any particular time. In submarine contouring problems , the true values $Z(x)$ are to be estimated and mapped (case b/). In mining, it is the mean ore grade within a panel of given size which is of interest (case c/). In certain studies of a more fundamental nature, where the object is to reconstitute the mechanisms which have produced the phenomenon being studied, the drift itself would give a reflection of the structure of these mechanisms (case a/). In geophysics for example, the notion of regional anomaly corresponds to the concept of drift.

This brings us to the second objection against the methods of least squares. It is impossible for the same polynomial $P(x)$ to solve the three problems set in a/, b/ and c/. It can be seen that $P(x)$ solves the problem a/. But in general, it is not the best solution possible ; "master" methods, which always use the same polynomials whatever the structural characteristics of the phenomenon under study, have in general no chance to produce the optimum solution.

Thirdly, least squares methods have no way of determining the magnitude of the error made by estimating the drift from the polynomial $P(x)$. Contrary to what is sometimes believed, the variance of the residuals is not an estimation variance ; the variance of the differences $Z(x_i) - P(x_i)$ at points such as x_i where the data are actually known is, by construction, systematically less (and very much less) than that of the differences $Z(x) - P(x)$ at points x whose values are not actually known. Thus, the variance of the residuals is not the estimation variance of $Z(x)$ (problem b/ above). Moreover, neither is it the estimation variance of the drift (problem a/), because in this case there is no conceptual link between the difference $Z(x_i) - P(x_i)$ and the quality of $P(x_i)$ considered as an estimator of the a priori expectation $E[Z(x_i)]$.

However, even if the methods of least squares are unsatisfactory, the purpose they had in view, but which they attain so badly or not at all, is concerned with a real and important problem. There are actual phenomena which absolutely cannot be represented by (realizations of) stationary random functions. In submarine contouring for example, it is quite certain that depth increases with increasing distance from the shore. We shall try to set out and solve the important problems raised by such drifts in terms of the theory of non-stationary random functions.

4-1-2 Statement of the problem and general hypotheses.

The regionalized variable $z(x)$ to be studied will be interpreted as the realization of a random function $Z(x)$ which in general is non-stationary. In the first set of hypotheses, it will be assumed that $Z(x)$ has for its first and second order moments :

$$(I) \qquad \begin{cases} E[Z(x)] = m(x) \\ \\ E[Z(x)\ Z(y)] = m(x)\ m(y) + C(x,y) \end{cases}$$

We shall say that the function $m(x)$ is the drift of the R.F. $Z(x)$. In fact, it seems that the only definition of drift which is clear in concept is this one : the first moment of a non-sationary R.F. The function $C(x,y)$ depends on the two points x and y and is the usual (non-stationary) covariance.

In many cases (as shown in practice by Geostatistics), the hypotheses set out in (I) are still too restrictive, and are better replaced by the hypotheses in (I'), which show that the increments of the R.F. $Z(x)$ (and not $Z(x)$ itself) have the following first and second order moments :

$$(I') \qquad \begin{cases} E[Z(x) - Z(y)] = m(x) - m(y) \\ \\ \frac{1}{2} D^2[Z(x) - Z(y)] = \gamma(x,y) \end{cases}$$

In this case, the drift $m(x)$ is defined except for a constant only, for, in general, the expectation $E[Z(x)]$ no longer exists : hence the point of view of hypotheses (I') amounts to studying the random function $Z(x)$ except for an additive constant. The function $\gamma(x,y)$ is the usual semi-variogram. (In one dimension, Brownian motion or Poisson processes are examples of an R.F. having no expectation, which verifies the hypotheses of (I'), but not of (I)).

The two following problems, related but distinct, have to be solved : knowing the numerical values in a set S (i.e. the set of points where experimental data are available) taken by a realization $z(x)$ of the R.F. $Z(x)$, we have to :

1/ estimate the function $m(x)$ (over S and outside S) : this is the problem a/ above.

2/ estimate $z(x)$ at a point $x \notin S$ (problem b/), or, more generally, estimate a "moving average" $\int \mu(dx)\ z(x)$ where μ is a measure whose support is distinct from S (problem c/).

Moreover, it must be possible to represent by means of estimation variances the errors made in these estimations, and to choose the estimators in order to minimize (as far as can be done) this estimation variance.

Concerning this last point, it is emphasized that we shall try to find the best possible linear estimator of $m(x)$ or $\int \mu(dx)\, z(x)$ that can be constructed from the numerical values of $z(y)$, $y \in S$: the non-linear estimators are far too complicated to be used in these problems, and on the other hand their properties do not depend only on the first and second order moments which appear in expressions (I) and (I') but make the whole distribution of the R.F. $Z(x)$ intervene. (Moreover, in the case where this distribution is Gaussian, it is known that the best possible estimator coïncides with the best linear estimator).

Thus, stated in completely general terms (functions $m(x)$ and $C(y,x)$ being totally unknown), the problem obviously cannot be solved, and is also probably meaningless. But in actual problems, the idea of drift can only have real meaning if the corresponding function $m(x)$ varies continuously and regularly in relation to the scale at which we are working (and to the available experimental data): if the function $m(x)$ is irregular and apparently chaotic at this scale, it must be considered as the realization of a new R.F. From the point of view of physical interpretation (i.e. non mathematical) the idea of drift is obviously related to that of scale. Taking topography as an example, at a scale of tens of metres, the idea of mountain is expressed by a drift ; at a scale of tens of kilometers, it corresponds to a random function, and at this scale, the drift would rather express the notion of a chain of mountains (see also the concept of gigogne structure [7]).

But the condition of regularity, imposed a priori on the function $m(x)$ so that the idea of drift has an actual physical content, means as well that a local estimate of $m(x)$ is always possible to some extent. In fact, this condition states that in a certain vicinity V of a given point x_0, the function $m(x)$ can be approximated with excellent precision by a function of the form :

$$m(x) \simeq f(x) = \sum_{\ell=0}^{k} a_\ell f^\ell(x) = a_\ell f^\ell(x) \qquad \text{(II)}$$

where the $f^\ell(x)$ are known functions, chosen once and for all (e.g. polynomials), and the a_ℓ are unknown coefficients : the neighbourhood V of x_0 where the approximation (II) is acceptable without being too large must - if the problem makes sense - contain enough experimental points so that the $k+1$ unknown coefficients $a_0, a_1, \ldots a_k$ can be estimated.

There remains the problem of the function $C(x,y)$, or $\gamma(x,y)$, about which nothing is known a priori. But here again and for the same reasons, it can be assumed that C or γ can be assimilated to functions of known types, at least locally, and change only slowly with distance on the working scale.

The most favourable case will be that of a function γ (or C) of the form $\gamma(x,y) = \varpi\ \gamma_0(x-y)$ where γ_0 is a known function, and ϖ is a factor which changes slowly, and which can be regarded as being constant in the neighbourhood V of x_0 : as the equations which determine the optimum estimators are linear and homogeneous, these estimators will only depend on γ_0 and not on the factor ϖ; the latter only affects the estimation variance, and not the estimators themselves. In certain cases, the following simple expression can be used :

$$\gamma(x,y) = \varpi\ r \qquad (r = |x-y|)$$

(linear variogram). In this case, it is sufficient that the true variogram or true covariance has a linear behaviour near $x = y$ and up to distances comparable with the dimensions of the neighbourhood V described above ; this happens more often than would be thought at first sight.

In general, however, it will be convenient to introduce one or more supplementary parameters besides the factor ϖ. e.g. :

$$\gamma(x,y) = \varpi\ r^{\alpha-1} \qquad (0 < \alpha < 2)$$

or else

$$C(x,y) = \varpi\ e^{-ar}$$

In the first case, the parameter α is related to the degree of continuity of the Re. V. In the second case,(exponential covariance), the parameter a, or rather its inverse, gives a measure of the range of the phenomenon (i.e. the distance beyond which correlations vanish).

Of course, the experimental control of this kind of hypothesis, and the estimation of the corresponding parameters (ϖ and above all α or a) will raise some delicate problems of mathematical statistics : in spite of their prime importance for practical applications, it is not possible to discuss these questions completely here, and we will evoke them only briefly.

Summing up the problem at hand, it may be set out as follows : we have a R.F. Z(x) verifying either hypothesis (I) or (I'). In the neighborhood V of a point x_o, the drift is of the form (II), the coefficients a_ℓ being unknown. The covariance C(x,y) or the variogram $\gamma(x,y)$ are known a priori (eventually except for a factor). Lastly, the numerical values of (i.e. the realization of) Z(x) at points x belonging to the set $S \subset V$ are known. The problem is to find the best linear estimators of :

1°/ the unknown coefficients a_ℓ of the drift.

2°/ the numerical value of (i.e. the realization of) Z(x) at $x_o \notin S$ (but $x_o \in V$) or of $\int \mu(dx)$ Z(x) where the support of the measure μ is separate from S.

3°/ and to be able to control the validity of the hypothesis made when choosing a particular mathematical expression for the covariance or the variogram, and to estimate the parameters on which this expression depends.

To aid understanding, the theory will not be expressed in terms of Hilbertian spaces. The reader will be referred to [6] for certain demonstrations (especially for the establishment of the theorems of existence and uniqueness). The notations used are those of paragraph 3-2. In fact, real difficulties will arise only when the experimental data set S is infinite: the existence (in the sense of Hilbertian spaces) of our optimal estimators is always ensured, but they do not necessarily have a representation of the type $\int \lambda(dx)$ Z(x) with measures μ with support in S. In general, in what follows, the integral equation relative to the continuous case will be written without explicitly repeating this essential reservation, which the reader should always bear in mind.

4-2 OPTIMAL ESTIMATION OF THE DRIFT.

4-2-1 Estimation of the drift at a point.

Let us first examine the case where the hypotheses (I) and (II) of paragraph 4-1-2 apply : existence of a drift having the representation :

$$m(x) = a_\ell f^\ell(x) \qquad (\ell = 0, 1, \ldots k)$$

in a certain neighbourhood V of the experimental data domain S, and existence of a centred covariance $\sigma(x,y)$. The numerical values of the k+1 coefficients a_ℓ are unknown, and we want to estimate the value $m(x_o) = a_\ell f^\ell(x_o)$ of the drift at a point x_o (which can, indifferently, belong to S or not, but must belong to V).

For that purpose, a linear estimator of the form

$$m^*(x_o) = \lambda^\alpha Z_\alpha \quad \text{or} \quad \int_S \lambda(dx)\, Z(x)$$

will be constructed, and two conditions will be imposed on the coefficients λ^α or on the measure λ :

a/ The estimator must be unbiased whatever the (unknown) values of the coefficients a_ℓ (universality condition).

b/ Taking the universality condition into account, the variance of this estimation will be minimal (optimality condition).

In general, the argument will be given for the case where the set S is finite, and the results so obtained will only be transposed to the continuous case (with the reservations expressed in the preceding paragraph, concerning the representation of the estimators by means of measures).

Let us first examine the universality condition. It is written $E[m^*(x_o)] = m(x_o)$, i.e. :

$$a_\ell \lambda^\alpha f_\alpha^\ell = a_\ell f^\ell(x_o)$$

whatever the a_ℓ , thus

$$\lambda^\alpha f_\alpha^\ell = f^\ell(x_o) \tag{4-2-1}$$

or, in the continuous case :

$$\int_S \lambda(dx)\, f^\ell(x) = f^\ell(x_o)$$

When this condition is verified, $E[m-m^*]^2$ is equal to the variance $D^2(m-m^*)$ and hence no longer depends on the unknown coefficients a_ℓ . This variance is :

$$D^2(m^*) = \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta}$$

The coefficients λ^α remain to be determined by minimizing this quadratic form, taking the universality condition (4-2-1) into account. There appear k+1 Lagrange factors μ_ℓ , and the following system is obtained :

$$(4\text{-}2\text{-}2) \qquad \begin{cases} \lambda^\beta \sigma_{\alpha\beta} = \mu_\ell f_\alpha^\ell \\ \\ \lambda^\alpha f_\alpha^\ell = f^\ell(x_o) \end{cases} \qquad (\ell = 0,1,\ldots k)$$

The covariance matrix is always supposed to be strictly positive definite. It can then be shown (cf. [6]) that the system (4-2-2) is <u>regular</u> and has thus one and only one solution, if and only if the k+1 functions f_ℓ are <u>linearly independent on S</u> (in the algebraic sense), i.e. if :

$$c_\ell \ f_\alpha^\ell = 0 \ \Rightarrow \ c_\ell = 0$$

In what follows, this condition will always be supposed to be fulfilled : in the applications however, it will always be necessary to make sure that it is so.

In the <u>continuous</u> case, with the usual reservations, an estimator will be sought, which is of the form $m^*(x_o) = \int_S \lambda(dx)\ Z(x)$ with a measure λ with support in S verifying the following system :

$$(4\text{-}2\text{-}3) \qquad \begin{cases} \int \lambda(dx)\ \sigma(x,y) = \mu_\ell f^\ell(y) \qquad , \qquad \forall\ y \in S \\ \\ \int \lambda(dx)\ f^\ell(x) = f^\ell(x_o) \end{cases}$$

Let us now look for the <u>variance</u> of the estimator. At the optimum, the system (4-2-2) is verified. By multiplying the first relationship of this system by λ^α , and using the second relationship, we get :

$$\lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} = \mu_\ell \ \lambda^\alpha f_\alpha^\ell = \mu_\ell \ f^\ell(x_o)$$

Hence, in the discrete case as in the continuous case, the variance of the estimator is related to the Lagrange parameters by the formula :

(4-2-4) $$D^2[m^*(x_o)] = \mu_\ell \, f^\ell(x_o)$$

Case when a variogram exists, but no covariance.

In all that follows, the function f^o corresponding to $\ell = 0$ will always be the function identically equal to 1, $f^o(x) = 1$, and the coefficient a_o will play a particular role. Indeed the drift, as the R.F. itself, can in fact be defined except for a constant only, and a_o is really indeterminate when no covariance exists. In principle, it is not possible to estimate $m(x_o)$ itself. This can also be seen when noticing that for $\ell = 0$, the universality condition (4-2-1) is reduced to $\sum_\alpha \lambda^\alpha = 1$: thus this is not a permitted linear combination (paragraph 2-2-1). On the contrary, it is possible to estimate $m(x_o) - m(y_o)$ for any couple of points x_o and y_o. Indeed, let us fix x_o and consider the R.F. :

$$Z'(x) = Z(x) - Z(y_o)$$

Its drift is $m(x) - m(y_o) = m'(x_o)$ and its covariance :

$$C(x,y) = -\gamma(x,y) + \gamma(x,y_o) + \gamma(y,y_o)$$

(See paragraph 2-2-1). This time we have :

$$m'(x) = a_\ell \, f^\ell(x) - f^\ell(y_o) \qquad (\ell = 1,2,\ldots k)$$

the constant a_o being eliminated. The argument can then be repeated and $m'(x_o)$ can be estimated with the help of a permitted linear estimator :

$$\begin{cases} m'^*(x_o) = \lambda^\alpha Z'_\alpha = \lambda^\alpha Z_\alpha \\ \\ \sum_\alpha \lambda^\alpha = 0 \end{cases}$$

The universality conditions are then :

$$\lambda^\alpha \varphi^\ell_\alpha = \varphi^\ell(x_o) \qquad \ell = 1,2,\ldots k$$

(the equation in $\ell = 0$ disappears, since $\varphi^o = 1 - 1 = 0$). The variance $D^2(m'^*)$ is :

$$D^2(m'^*) = \lambda^\alpha \lambda^\beta (-\gamma_{\alpha\beta} + \gamma_{\alpha x_o} + \gamma_{\beta y_o}) = -\lambda^\alpha \lambda^\beta \gamma_{\alpha\beta}$$

because of the condition $\sum \lambda^\alpha = 0$, and we get the system :

$$\begin{cases} -\lambda^\beta \gamma_{\alpha\beta} = \mu_\ell \varphi_\alpha^\ell + \mu_o' \\ \lambda^\alpha \varphi_\alpha^\ell = \varphi^\ell(x_o) \\ \sum \lambda^\alpha = 0 \end{cases}$$

Besides, $\mu_\ell \varphi_\alpha^\ell = \mu_\ell f_\alpha^\ell - \mu_\ell f^\ell(y_o)$, and the constant $\mu_\ell f^\ell(y_o)$ may be included into the Lagrange parameter μ_o'. Likewise, $\lambda^\alpha \varphi_\alpha^\ell = \lambda^\alpha f_\alpha^\ell$ because $\sum \lambda^\alpha = 0$. Finally, we obtain the following system, deduced from (4-2-2) by replacing $\sigma_{\alpha\beta}$ by $-\gamma_{\alpha\beta}$, restricting the index ℓ to the values $1,2,\ldots k$, and introducing in counterpart the supplementary universality condition $\sum \lambda^\alpha = 0$ and the corresponding parameter μ_o :

(4-2-5)
$$\begin{cases} \lambda^\beta \gamma_{\alpha\beta} = -\mu_\ell f_\alpha^\ell - \mu_o \\ \lambda^\alpha f_\alpha^\ell = f^\ell(x_o) - f^\ell(y_o) \qquad (\ell = 1,2,\ldots k) \\ \sum_\alpha \lambda^\alpha = 0 \end{cases}$$

The corresponding estimator $\lambda^\alpha Z_\alpha$ (which estimates the increment $m(x_o) - m(y_o)$) has for a variance:

(4-2-6)
$$D^2(\lambda^\alpha Z_\alpha) = \mu_\ell [f^\ell(x_o) - f^\ell(y_o)]$$

4-2-2 Estimation of the coefficients of the drift.

First, consider the case where a covariance exists. The solution of (4-2-2) $\lambda^\alpha = \lambda^\alpha(x_o)$ obviously depends on x_o, and examination of the right hand sides shows that it depends linearly on $f^\ell(x_o)$. When x describes V, or even simply S, the $f^\ell(x)$ are linearly independent. Hence there exist a unique matrix λ_ℓ^α and a matrix $\mu_{\ell s}$ such that :

$$\lambda^\alpha(x_o) = \lambda_\ell^\alpha f^\ell(x_o) \ , \ \mu_\ell(x_o) = \mu_{\ell s} f^s(x_o)$$

and this unique matrix is the solution of the system deduced from (4-2-2) by equating to 0 the coefficient of $f^{\ell}(x_o)$, i.e. :

$$(4\text{-}2\text{-}7) \qquad \begin{cases} \lambda_{\ell}^{\beta} \ \sigma_{\alpha\beta} = \mu_{\ell s} \ f_{\alpha}^{s} \\ \\ \lambda^{\alpha} \ f_{\alpha}^{s} = \delta_{\ell}^{s} \end{cases} \qquad (\ell = 0,1,\ldots k)$$

The estimator $m^*(x_o) = \lambda_{\ell}^{\alpha} \ z_{\alpha} \ f^{\ell}(x_o)$ is thus of the form $A_{\ell} \ f^{\ell}(x_o)$ with

$$(4\text{-}2\text{-}8) \qquad \begin{cases} A_{\ell} \ = \lambda_{\ell}^{\alpha} \ z_{\alpha} \\ \\ E(A_{\ell}) = a_{\ell} \end{cases}$$

(whatever the unknown values of the a_{ℓ} are). Hence, A_{ℓ} is a universal estimator of a_{ℓ}, characterized by the fact that for any vector b^{ℓ}, the variance of $A_{\ell} \ b^{\ell}$ is minimal (taking the universality consition into account). Thus A_{ℓ} is the optimal universal estimator of the coefficient a_{ℓ}.

From (4-2-4), for any vector b^{ℓ} :

$$D^2 \ (A_{\ell} \ b^{\ell}) = \mu_{\ell s} \ b^{\ell} b^{s}$$

Hence the matrix $\mu_{\ell s}$ of the Lagrange factors is identical to the covariance matrix of the optimal estimators A_{ℓ} (in particular, it is always symmetrical) :

$$(4\text{-}2\text{-}9) \qquad \mu_{\ell s} = \text{Cov} \ (A_{\ell} \ , \ A_{s}) \qquad (\ell = 0,1,\ldots k)$$

Case where there is no covariance, but only a variogram.

The system (4-2-5) also depends linearly on the $f^{\ell}(x_o) - f^{\ell}(y_o)$. Thus a unique matrix λ_{ℓ}^{α} ($\ell = 1,2,\ldots k$) exists, verifying :

$$(4\text{-}2\text{-}10) \qquad \begin{cases} \lambda_{\ell}^{\beta} \ \gamma_{\alpha\beta} = - \mu_{\ell s} \ f_{\alpha}^{s} - \mu_{o\ell} \\ \\ \lambda_{\ell}^{\alpha} \ f_{\alpha}^{s} = \delta_{\ell}^{s} \qquad (\ell = 1,2,\ldots k) \\ \\ \sum_{\alpha} \lambda_{\ell}^{\alpha} = 0 \end{cases}$$

The $\lambda_\ell^\alpha \; Z_\alpha$, $\ell = 1,2,\ldots k$; are the optimal estimators A_1, A_2, A_k of the coefficients a_1, a_2,...a_k (except a_o) and the matrix $\mu_{\ell s}$ is equal to the covariance matrix of these estimators.

$$\mu_{\ell s} = \mathrm{Cov}\,(A_\ell\ ,\ A_s) \qquad (\ell = 1,2,\ldots k) \tag{4-2-11}$$

Note : To complete the system (4-2-10), it is possible to introduce a vector λ_o^α and the function $f^o(x) = 1$, and to determine λ_o^α by the conditions $\sum_\alpha \lambda_o^\alpha = 1$, $\lambda_o^\beta\ \gamma_{\alpha\beta} = -\ \mu_{os}\ f_\alpha^s - \mu_{oo}$. The system thus completed is identical to (4-2-7).

$$\begin{cases} \lambda_\ell^\beta\ \gamma_{\alpha\beta} = -\ \mu_{\ell s}\ f_\alpha^s \\ \\ \lambda_\ell^\alpha\ f_\alpha^s = \delta_\ell^s \end{cases} \qquad (\ell = 0,1,\ldots k)$$

We shall return in paragraph 4-6 to the interpretation that can be given to the λ_o^α and to the linear combination

$$A_o = \lambda_o^\alpha\ Z_\alpha$$

which is not a permitted linear combination, and hence has in general, no finite variance (relationship (4-2-11) is only valid for ℓ and $s \neq 0$).

Continuous case. - There is no difficulty in transposing the system (4-2-7) or (4-2-10) to the case where the data set S is infinite. With the usual reservations about the capability of representing the estimators A_ℓ (which always exist in the sense of Hilbertian spaces) by means of the measures λ_ℓ :

$$A_\ell = \int_S \lambda_\ell\,(dx)\ Z(x)$$

the system (4-2-10) for instance, becomes :

$$\begin{cases} \int_S \lambda_\ell(dx)\ \gamma(x,y) = -\ \mu_{\ell s}\ f^s(y) - \mu_{o\ell} & (y \in S) \\ \int \lambda_\ell(dx)\ f^s(x) = \delta_\ell^s & (\ell = 1,2,\ldots k) \\ \int \lambda_\ell(dx) = 0 \end{cases} \tag{4-2-12}$$

$$(\mu_{\ell s} = \mathrm{Cov}\,(A_\ell\ , A_s))$$

4-2-3 Tensorial Invariance.

The optimal estimator of the drift is unchanged when a linear transformation is performed on the functions f^ℓ. Consider, for example, the case where a covariance σ exists (the argument would be the same in the case of a variogram). Let $B = (B_s^\ell)$ be a regular matrix, $B' = (B'^\ell_s)$ its inverse. Put

$$\varphi^\ell(x) = B_s^\ell f^s(x)$$

the drift $m(x)$ takes the form

$$\begin{cases} m(x) = a'_\ell \ \varphi^\ell(x) \\ a'_\ell = B'^s_\ell \ a_s \end{cases}$$

With these new functions φ^ℓ, the optimal estimator of $m^*(x_o)$ is

$$m'^*(x_o) = \lambda'^\alpha Z_\alpha$$

with coefficients λ'^α the solution of the system :

$$(4\text{-}2\text{-}2') \qquad \begin{cases} \lambda'^\beta \sigma_{\alpha\beta} = \mu'_\ell \ f_\alpha^\ell \\ \lambda'^\alpha \varphi_\alpha^\ell = \varphi^\ell(x_o) \end{cases}$$

But $\varphi_\alpha^\ell = B_s^\ell f_\alpha^s$, $\varphi^\ell(x_o) = B_s^\ell f^s(x_o)$ and the matrix B is regular . Hence, the second relationship (4-2-2') is equivalent to

$$\lambda'^\alpha f_\alpha^\ell = f^\ell(x_o)$$

The system (4-2-2') itself is thus equivalent to (4-2-2) and $\lambda'^\alpha = \lambda^\alpha$, as the solution is unique, hence :

$$m'^*(x_o) = m^*(x_o)$$

The optimal estimator does not change for linear transformations of the functions f^ℓ. Concerning the optimal estimator A'_ℓ of the new coefficients a'_ℓ, they verify :

$$A'_\ell \ \varphi^\ell(x_o) = A_\ell \ f^\ell(x_o)$$

for any x_o, thus :

$$A'_s \ B^s_\ell = A_\ell \quad \text{or} \quad A'_s = B'^\ell_s \ A_\ell$$

The optimal estimators are transformed according to the same law as the real coefficients, i.e., in a contravariant manner with respect to the f^ℓ.

Note - In applications, it often happens that too large a zone is to be studied, so that the representation of $m(x)$ in the form $a_\ell \ f^\ell(x)$ is not admissible. This zone is then cut into domains V_1, V_2, ... in which the approximation is legitimate, but with different coefficients a_ℓ on V_1, V_2 etc....

It is more convenient to place the origin of the coordinates at a point $x_i \in V_i$ to study the domain V_i, and this amounts to changing $\varphi^\ell(x)$ to $f^\ell(x-x_i)$ in the basic functions. If the functions f^ℓ are polynomials, trigonometric functions, or, more generally, multinomial exponentials, the new functions φ^ℓ are linearly expressed by means of the original ones : hence, from the tensorial invariance, this change of coordinates does not affect the optimal estimator of the drift.

In fact, it can be shown that the multinomial exponentials are the only functions having this property (invariance by translation of the finite dimensional vectorial space generated by the f^ℓ). Therefore, in applications, only this type of functions is used (unless the physical conditions of the problem give a privileged rôle to a particular point of the space, and this is then chosen as the absolute origin of the coordinates).

4-2-4 The variogram of the residuals.

We will call the residual at the experimental point $x \in S$ the difference between the observed value and the (optimally) estimated drift at this point, i.e. :

$$R(x) = Z(x) - A_\ell \ f^\ell(x)$$

When a covariance exists and when S is finite, the two matrices :

$$\Lambda_\alpha^\beta = \lambda^\beta f_\alpha \quad , \qquad U_\alpha^\beta = \delta_\alpha^\beta - \Lambda_\alpha^\beta$$

respectively give the optimal estimator m_α^* of the drift at x_α and the residual at this point :

(4-2-13) $$m_\alpha^* = \Lambda_\alpha^\beta Z_\beta \quad , \qquad R_\alpha = U_\alpha^\beta Z_\beta$$

We shall return later on to the algebraic properties of these two matrices which play an important rôle in the problem of the indeterminability of the underlying variogram (paragraph 4-6).

When there is no covariance but only a variogram, the residuals themselves are indefinable (as A_o no longer exists), but their <u>increments</u>

$$R(x) - R(y) = Z(x) - Z(y) = m^*(x) + m^*(y)$$

are well defined, since they do not depend on A_o. One can then speak of the <u>variogram of the residuals</u> which will be, by definition, the function :

$$\gamma_R(x,y) = \frac{1}{2} E[R(x) - R(y)]^2 \qquad (x,y \in S)$$

Let us compute this expectation (assuming that the $m^*(x)$ are effectively the optimal estimators of the drift). We have :

$$\gamma_R(x,y) = \frac{1}{2} D^2(Z_x - Z_y) - \mathrm{Cov}\,(Z_x - Z_y)(m_x^* - m_y^*) + \frac{1}{2} D^2(m_x^* - m_y^*)$$

By definition, $\frac{1}{2} D^2(Z_x - Z_y) = \gamma(x,y)$. From (4-2-6) and (4-2-10), we have also :

$$\frac{1}{2} D^2(m_x^* - m_y^*) = \frac{1}{2} \mu_{\ell s} [f_x^\ell - f_y^\ell][f_x^s - f_y^s]$$

Let us compute the cross term. Firstly, we get :

$$(Z_x - Z_y)(m_x^* - m_y^*) = [f_x^\ell - f_y^\ell][Z_x - Z_y] \int_S \lambda_\ell \,(dz)\, Z(z)$$

then :

$$E[(Z_x - Z_y) \int_S \lambda_\ell(dz)\ Z(z)] = \int_S \lambda_\ell(dz)\ [-\gamma(x,z) + \gamma(y,z)]$$

as these are permitted linear combinations. As x and y belong to S, the first relationship (4-2-12) then gives

$$\int_S \lambda_\ell(dz)\ [-\gamma(x,z) + \gamma(y,z)] = \mu\ell_s\ (f_x^\ell - f_y^s)$$

hence :

$$E(Z_x - Z_y)(m_x^* - m_y^*) = \mu\ell_s\ (f_x^\ell - f_y^\ell)(f_x^s - f_y^s)$$

As the estimators are optimal, the usual simplification occurs, the cross term is equal to the quadratic term, and the variogram of the residuals is :

(4-2-14) $$\gamma_R(x,y) = \gamma(x,y) - \tfrac{1}{2}\ \mu\ell_s\ [f_x^\ell - f_y^\ell][f_x^s - f_y^s]$$

As $\mu\ell_s$ is positive definite, the variogram of the residuals is <u>always less</u> (and even much less) <u>than the true variogram</u> (or, as we will call it, the <u>underlying variogram</u>).

Let then k(x) be the indicator of the experimental data set S, and K(h) its geometric covariogram (in the continuous case. If S is discrete, we would get analogous results). Put

$$\gamma^*(h) = \frac{1}{K(h)} \int k(x)\ k(x+h)\ [Z_{x+h} - Z_x - m_{x+h}^* + m_x^*]^2\ dx$$

$\gamma^*(h)$ is an unbiased estimator of the <u>average variogram of the residuals</u>:

$$E[\gamma^*(h)] = \bar{\gamma}_R(h) = \frac{1}{K(h)} \int k(x)\ k(x+h)\ \gamma_R(x,x+h)\ dx$$

But $\bar{\gamma}_R(h)$ will be very different from the true average variogram :

$$\bar{\gamma}(h) = \frac{1}{K(h)} \int k(x)\ k(x+h)\ \gamma(x,x+h)\ dx$$

From relationship (4-2-14), we get :

(4-2-15)
$$\begin{cases} \bar{\gamma}_R(h) = \bar{\gamma}(h) - \frac{1}{2}\ \mu\ell_s\ T^{\ell s} \\ T^{\ell s} = \frac{1}{K(h)} \int k(x)\ k(x+h)\ [f_{x+h}^\ell - f_x^\ell][f_{x+h}^s - f_x^s] \end{cases}$$

hence $\overline{\gamma}_R(h) < \overline{\gamma}(h)$: it can be seen that $\gamma^*(h)$ gives a heavily biased estimation of the true mean variogram.

In Exercise 1, paragraph c/, it will be seen that this bias is considerable. The bias can be sufficiently large as to persuade the experimenter to draw false conclusions about the independence of the residuals. This effect is of course general, and also occurs when the drift is estimated by any other procedure (non-optimal) for instance by least squares methods : it gives to the experimenter the (wrong) impression that the phenomenon he is studying is reduced to the simple sum of a drift and a white noise.

4-2-5 Comparison with the method of maximum likelihood.

In the case when the set S is finite and the R.F. Z(x) is Gaussian, the optimal estimator of the drift formed in paragraph 4-2-1 is identical to the classical estimator of the maximum likelihood. Indeed, let Z_α be N Gaussian variables, $\sigma_{\alpha\beta}$ their covariance matrix, $B^{\alpha\beta}$ the inverse matrix and :

$$E(Z_\alpha) = m_\alpha = a_\ell \, f_\alpha^\ell$$

the expectations of the Z_α, depending on k+1 parameters a_ℓ to be estimated. The maximum likelihood estimator is obtained by choosing for the a_ℓ the values which maximise the expression for the probability density of the r.v. Z_α at the points z_α corresponding to the numerical values of the realization of the Z_α. In other words, the following quadratic form is minimized :

$$B^{\alpha\beta}(z_\alpha - m_\alpha)(z_\beta - m_\beta) = B^{\alpha\beta}(z_\alpha - a_\ell \, f_\alpha^\ell)(z_\beta - a_s \, f_\beta^s)$$

with respect to the a_ℓ . Let A'_ℓ be the values of the a_ℓ which minimise this expression. They are obtained by solving the system

(4-2-16) $$A'_\ell \; f_\beta^\ell \, f_\alpha^s \, B^{\alpha\beta} = f_\alpha^s \, B^{\alpha\beta} \, z_\beta$$

But the system (4-2-7) gives (by reversing the first relationship)

(4-2-16') $$\lambda_\ell^\beta = B^{\alpha\beta} \, f_\alpha^s \, \mu_{s\ell}$$

and (substituting the second relationship) :

$$f_\beta^\ell \, f_\alpha^s \, B_{\alpha\beta} \, \mu_{si} = \delta_i^\ell$$

Thus, the Lagrange parameters matrix $\mu\ell_s$ of (4-2-7) is the inverse of the matrix f B f of the system (4-2-16). The solution of this system is thus :

$$A'_\ell = \mu\ell_s \, f_\alpha^s \, B^{\alpha\beta} \, z_\beta$$

But from (4-2-16')the optimal estimator A_ℓ can be put in the form :

$$A_\ell = \lambda_\ell^\beta \, Z_\beta = \mu\ell_s \, f_\alpha^s \, B^{\alpha\beta} \, Z_\beta$$

We have thus the numerical equality $A'_\ell = A_\ell$ for $Z_\beta = z_\beta$ fixed, and hence the identity of the two estimators, as stated.

Note.- Our optimal estimator is more general than the maximum likelihood estimator, for it is not related to a Gaussian hypothesis and is generalized to the case where S is infinite. In applications, the system (4-2-7) is easier to solve, as it requires only one matrix inversion.

4-2-6 Case of a drift given in an implicit form.

In certain applications, particularly in geodesy, it may happen that a given drift is to be estimated, not in an explicit form $m_\alpha = a_\ell \, f_\alpha^\ell$, but in an implicit form, by a system of linear relationships to be verified by the m_α. The typical problem is the following : N parameters m_α are to be estimated, these parameters being linked by $N-k < N$ physical or geometrical relationships (for example the 3 sides and the 3 angles of a Euclidean triangle, verifying 3 geometrical relationships). The available data Z_α contain the errors Y_α, i.e. :

$$Z_\alpha = Y_\alpha + m_\alpha$$

We suppose that $E[Y_\alpha] = 0$, and the matrix $\sigma_{\alpha\beta}$ of the covariances of the errors Y_α is known; using the Z_α, we try to form the best possible estimator m^*_α of the unknown parameters m_α (which implies, in

particular, that the m_α^* themselves follow the constraints imposed to the m_α). In general, a solution which is sufficiently approximate is known at the start so that it is possible to linearize the constraints. In other words, a matrix M can be found, such that the constraints are in the form :

$$M_u^\alpha \, m_\alpha = 0 \qquad (u = 1, 2, \ldots, N-k) \tag{4-2-17}$$

and so that M is of rank N-k. It is obvious that (4-2-17) expresses that the vector m belongs to a k-dimensional subspace of $\mathbb{R}^N$. If we choose a base f^ℓ ($\ell = 1,2,\ldots k$) of this subspace, the conditions (4-2-17) are equivalent to the existence of k coefficients a_ℓ such that :

$$m_\alpha = a_\ell \, f_\alpha^\ell \tag{4-2-17'}$$

The problem can then be reduced to the optimal estimation of the drift m_α^* . But it is not necessary to define the f_α^ℓ more closely, and this problem can be treated directly starting from the conditions (4-2-17) given in implicit form.

Before establishing this result, notice first of all that the optimal estimator will be (implicitly or explicitly) of the form :

$$m_\alpha^* = A_\ell \, f_\alpha^\ell$$

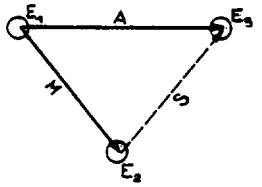
and consequently, from the equivalence of relationships (4-2-17) and (4-2-17'), will verify by itself the imposed conditions, i.e. :

$$M_u^\alpha \, m_\alpha^* = 0$$

without the necessity of introducing supplementary constraints.

To be able to perform the computations in implicit form, we need the following simple lemma :

Algebraic Lemma : Let E_1, E_2, E_3 be linear spaces with respectively n_1, n_2, n_3 dimensions, and $n_1 \geq n_2$; let M be a linear mapping from E_1 onto E_2, A a linear mapping from E_1 into E_3. A is factorized by M (i.e. there exists a linear mapping S from E_2 into E_3 such that $A = S \circ M$) if and only if the kernel of A contains the kernel of M (thus that $Mx_1 = 0$, $x_1 \in E_1$ involves $Ax_1 = 0$).



This condition is obviously necessary. Conversely, suppose that it is true and let x_2 be a point of E_2, $x_1 \in M^{-1}(x_2)$ a point of the inverse image of x_2 into E_1. Any other point of this inverse image is of the form $x_1 + y_1$ with $My_1 = 0$. We have then $A(x_1+y_1) = Ax_1$, as $My_1 = 0$ involves $Ay_1 = 0$. The mapping S of E_2 within E_3 defined by $Sx_2 = Ax_1$ for an $x_1 \in M^{-1}(x_2)$ is hence defined without ambiguity, is obviously linear, and verifies by construction $Ax_1 = Sx_2 = S\,M\,x_1$ for any $x_1 \in E_1$, thus $A = S_o\,M$, which establishes the lemma.

Let us return to our implicit drift. To estimate the m_α, we are going to form the linear estimators

$$m^*_\alpha = \Lambda^\beta_\alpha\, Z_\beta$$

verifying the two usual conditions of optimality and universality :

Universality condition : $E[m^*_\alpha] = m_\alpha$, or $\Lambda^\beta_\alpha\, m_\beta = m_\alpha$ whatever the unknown parameters m_α verifying (4-2-17). In other words $M^\alpha_u\, m_\alpha = 0$ must entail $(\delta^\beta_\alpha - \Lambda^\beta_\alpha)\, m_\alpha = 0$. From the lemma, this is so if and only if the linear mapping $I - \Lambda$ of $\mathbb{R}^N$ into $\mathbb{R}^N$ is factorized by the mapping M of $\mathbb{R}^N$ onto $\mathbb{R}^{N-k}$, hence if and only if there exists a matrix S^u_α such that :

$$\delta^\beta_\alpha - \Lambda^\beta_\alpha = S^u_\alpha\, M^\beta_u \tag{4-2-18}$$

Optimality condition : for any given α, the N-k coefficients S^u_α must minimize the variance $D^2(m^*_\alpha)$. From (4-2-18) we have

$$m^*_\alpha = Z_\alpha - S^u_\alpha\, M^\beta_u\, Z_\beta$$

hence

$$D^2(m^*_\alpha) = \sigma_{\alpha\alpha} - 2\, S^u_{(\alpha)}\, M^\beta_u\, \sigma_{(\alpha)\beta} + S^u_{(\alpha)}\, S^v_{(\alpha)}\, M^\beta_u\, \sigma_{\beta\gamma}\, M^\gamma_v$$

(without summation on α). Equating to zero the partial derivatives with respect to $S^u_{(\alpha)}$, we obtain the system

$$S^v_\alpha\, M^\beta_u\, \sigma_{\beta\gamma}\, M^\gamma_v = M^\beta_u\, \sigma_{\alpha\beta} \tag{4-2-19}$$

This system is necessarily regular (as the algebraically equivalent problem formulated with the f^ℓ_α has a unique solution). This can be verified directly ($\sigma_{\alpha\beta}$ is regular and M^β_u is of rank N-k). The

corresponding estimation variance is :

$$D^2(m^*_\alpha) = \sigma_{\alpha\alpha} - S^u_{(\alpha)}\, M^\beta_u\, \sigma_{(\alpha)\beta} \qquad (4\text{-}2\text{-}20)$$

4-2-7 Comparison with least squares methods.

To estimate the coefficients a_ℓ of the drift $a_\ell f^\ell(x)$, where the known functions $f^\ell(x)$ are (for example) polynomials, we have sometimes recourse to the least squares method. This method will be studied and the efficiency of the estimators (non-optimal in general) that it gives will be examined through a simple case. Finally, we shall see in which case the least suqares method leads to the same optimal estimator as universal kriging.

a/ Theory of least squares. - Let Z(x) be a R.F. having a drift

$$m(x) = a_0 + \sum_{\ell=1}^{k} a_\ell\, f^\ell(x)$$

with unknown coefficients a_ℓ . The realization of Z(x) is known on a set S with indicator k(x) (= 1 if $x \in S$, = 0 if $x \notin S$). From this realization, k+1 numbers $B_0, B_1, \ldots, B_k$ minimizing the integral are determined:

$$I = \int k(x)\, [Z(x) - B_\ell\, f^\ell(x)]^2\, dx$$

These numbers verify the system obtained by equating to zero the partial derivatives $\frac{\partial I}{\partial B_\ell}$, i.e.

$$B_j \int k(x)\, f^j(x)\, f^\ell(x)\, dx = \int K(x)\, Z(x)\, f^\ell(x)\, dx$$

Put

$$T^{\ell j} = \frac{1}{S} \int k(x)\, f^\ell(x)\, f^j(x)\, dx$$

$S = \int k(x)\, dx$ being the measure of S. The preceding system can be written as (with summation from j = 0 to j = k) :

$$B_j\, T^{j\ell} = \frac{1}{S} \int k(x)\, Z(x)\, f^\ell(x)\, dx \qquad (4\text{-}2\text{-}21)$$

In applications, the functions $f^{\ell}(x)$ are always chosen so that the matrix $T^{j\ell}$ has an inverse $S_{j\ell}$. The B_j are therefore given (by summing from $\ell = 0$ to $\ell = k$) by :

$$B_j = \frac{1}{S} S_{j\ell} \int k(x)\ Z(x)\ f^{\ell}(x)\ dx \qquad (4\text{-}2\text{-}22)$$

It will be seen that this can be written :

$$B_j = \int \lambda_j(dx)\ Z(x)$$

with measures λ_j having support in S and which have densities :

$$\frac{\lambda_j(dx)}{dx} = \frac{1}{S} k(x)\ S_{j\ell}\ f^{\ell}(x)$$

These measures verify the universality conditions (i.e. those expressing the unbiased nature of the estimator of the drift). In fact, it can be shown that :

$$\int \lambda_j(dx)\ f^{s}(x) = \frac{1}{S} S_{j\ell} \int k(x)\ f^{\ell}(x)\ f^{s}(x)\ dx = S_{j\ell}\ T^{\ell s}$$

As the matrices S and T are the inverses of each other, we have :

$$\int \lambda_j(dx)\ f^{s}(x) = \delta_j^{s} \qquad (j,s = 0,1,\ldots k) \qquad (4\text{-}2\text{-}23)$$

b/ Expectations and covariances of the B_j.- We shall begin with the simplest case, in which Z(x) has a covariance function C(x,y). The B_j, defined by (4-2-22), which are numbers when Z(x) represents the realization of a R.F., become random variables when Z(x) is considered to be the R.F. itself (the stochastic integrals which appear in the second term of (4-2-22) do have meaning, as a covariance exists). Relationship (4-2-23) expresses, as we have already noticed, that the B_j are the unbiased estimators of the true unknown coefficients a_j : in this sense, it can be said that the least squares estimators are universal estimators (even though they are not, in general, optimum ones). Thus, we have :

$$E(B_j) = a_j \qquad (j = 0,1,\ldots k)$$

Let us now compute the covariance matrix

$$\beta_{ij} = E[B_i\ B_j] - a_i\ a_j$$

As the B_j are without bias, we can perform the computation as if $a_j = 0$ (which amounts to replacing $Z(x)$ by $Z(x) - a_\ell\ f^\ell(x)$). From (4-2-22) we get :

$$B_i\ B_j = \frac{1}{S^2}\ S_{i\ell}\ S_{js} \iint k(x)\ k(y)\ Z(x)\ Z(y)\ f^\ell(x)\ f^s(y)\ dx\ dy$$

hence, passing on to the expectations :

$$(4\text{-}2\text{-}24) \qquad \beta_{ij} = \frac{1}{S^2}\ S_{i\ell}\ S_{js} \iint k(x)\ k(y)\ f^\ell(x)\ f^s(y)\ C(x,y)\ dx\ dy \qquad (i,j = 0,1,\ldots k)$$

We shall now examine the case where only a <u>semi-variogram</u> $\gamma(x,y)$ exists, $Z(x)$ itself no longer has moments of order two, nor even, in general, an expectation. For $\ell = 0$, the stochastic integral on the right hand side of (4-2-22) no longer exists (in the sense of the convergence in the mean square) for, from (4-2-23), we have $\int \lambda_0(dx) = 1$. This amounts to saying that the variance of $D^2(B_0)$ is infinite, and that the expectation $E(B_0)$ also does not exist in general : we find here again a circumstance already encountered several times. On the other hand, when $j \neq 0$, relationship (4-2-23), written with $s = 0$, gives $\int \lambda_j(dx) = 0$ $(j = 1,2,\ldots k)$. This condition ensures the existence of the stochastic integral (4-2-22) (in the sense of the convergence in the mean square) which has then a finite variance. The <u>covariance matrix</u> β_{ij}, limited to the indices $i,j \neq 0$, is then given by the following relationship, similar to (4-2-24) :

$$(4\text{-}2\text{-}25) \qquad \beta_{ij} = -\frac{1}{S^2}\ S_{i\ell}\ S_{js} \iint k(x)\ k(y)\ f^\ell(x)\ f^s(y)\ \gamma(x,y)\ dx\ dy \qquad (i,j = 1,2,\ldots k)$$

c/ <u>Example of the linear variogram</u>.- We shall take $\gamma(x,y) = |x-y|$, and try to estimate a quadratic drift $a_0 + a_1x + a_2x^2$ from a realization of the R.F. $Z(x)$ given in the interval $(-R, +R)$ of the line $\mathbb{R}$.

There is no problem in computing the matrix T^{ij} and its inverse S_{ij} $(i,j = 0,1,2)$. We get :

$$T = \begin{bmatrix} 1 & 0 & \frac{R^2}{3} \\ 0 & \frac{R^2}{3} & 0 \\ \frac{R^2}{3} & 0 & \frac{R^4}{5} \end{bmatrix} \qquad S = \begin{bmatrix} \frac{9}{4} & 0 & -\frac{15}{4R^2} \\ 0 & \frac{3}{R^2} & 0 \\ -\frac{15}{4R^2} & 0 & \frac{45}{4R^4} \end{bmatrix}$$

Hence the least squares estimators :

$$\begin{cases} B_o = \frac{9}{4}\frac{1}{2R}\int_{-R}^{R} Z(x)\,dx - \frac{15}{4R^2}\frac{1}{2R}\int_{-R}^{R} x^2 Z(x)\,dx \\ B_1 = \frac{3}{R^2}\frac{1}{2R}\int_{-R}^{R} x\,Z(x)\,dx \\ B_2 = -\frac{15}{4R^2}\frac{1}{2R}\int_{-R}^{R} Z(x)\,dx + \frac{45}{4R^4}\frac{1}{2R}\int_{-R}^{R} x^2 Z(x)\,dx \end{cases}$$

For the covariance matrix limited to the indices $i,j = 1,2$ (β_{oo} would be infinite here) we get

$$\beta_{11} = \frac{6}{5R}\ ,\ \beta_{12} = \beta_{21} = 0\ ,\ \beta_{22} = \frac{15}{14R^3}$$

Comparing this with exercise 1 ($\frac{6}{5R}$ against $\frac{1}{R}$ for the optimal estimator, and $\frac{15}{14R^3}$ against $\frac{3}{4R^3}$) , we see that the least squares estimators are distinctly more variable than the optimal estimators.

Condition of optimality for the least squares estimator :

Among the estimators of the form $B_j = \int \lambda_j(dx)\ Z(x)$ (support of λ_j included in S) the least squares estimators are characterized by the two following conditions :

~ they are without bias ($E(B_j) = a_j\ ,\ \forall a_j$)

~ the measures λ_j have densities which are linear combinations of the functions $k(x)\ f^{\ell}(x)$.

We have already seen that the meast squares estimator verifies these conditions. Conversely, let us suppose that the measures λ_j verify the universality conditions (4-2-23). If the λ_j's have densities of the form $\frac{k(x)}{S}\ S'_{j\ell}\ f^{\ell}(x)$, these conditions (4-2-23) express that the matrix $S'_{j\ell}$ coïncides with the inverse $S_{j\ell}$ of the matrix T^{ij}, so that the estimators B_j are indeed given by (4-2-22).

An estimator verifying these two properties verifies by definition the universality conditions. It is optimal if and only if it verifies the first relationships (4-2-6), i.e. we can find constants μ_{js} such that :

(4-2-26) $$\frac{1}{S}\ S_{j\ell}\int k(x)\ f^{\ell}(x)\ C(x,y)\,dx = \mu_{js}\ f^{s}(y)\ ,\quad \forall\ y \in S$$

It follows from this that the least squares estimator coïncides with the optimal estimator of the drift if and only if the k+1 functions $f^{\ell}(x)$ are linear combinations of the k+1 distinct eigen functions of the kernel $k(x)\ C(x,y)\ k(y)$ considered as an operator acting upon the functions with support in S.

In particular, if the $f^{\ell}(x)$ are themselves eigen functions ($\int k(x)\ f^{\ell}(x)\ C(x,y)\ dx = \mu^{\ell}\ f^{\ell}(y)$ for $y \in S$); relationship (4-2-26) is reduced to :

$$\mu_{js} = \frac{1}{S}\ S_{js}\ \mu^{s} \qquad \text{(without summation in S)} \tag{4-2-27}$$

[In the case where no covariance C(x,y) exists, but only a variogram $\gamma(x,y)$, the condition is slightly less simple. If $\overline{Z} = \frac{1}{S} \int k(x)\ Z(x)\ dx$ represents the (spatial) average of Z(x) on S, $Z(x) - \overline{Z}$ is a R.F. having moments of order 1 and 2. Its covariance $\overline{C}(x,y)$ is given by :

$$\overline{C}(x,y) = -\gamma(x,y) + \frac{1}{S}\int k(y)\ \gamma(x,y)\ dx + \frac{1}{S}\int k(x)\ \gamma(x,y)\ dx$$

$$- \frac{1}{S^2}\iint k(x)\ k(y)\ \gamma(x,y)\ dx\ dy$$

It can then be shown that the B_j of the least squares ($j = 1,2,\ldots k$) coïncide with the optimal estimator if and only if the k functions $f^{\ell}(x)$ ($\ell = 1,2,\ldots k$) are linear combinations of the k eigen functions of the kernel $\overline{C}(x,y)$ corresponding to non-zero eigen values .

This condition is of quite a complex nature, since it calls in at the same time the functions $f^{\ell}(x)$, the kernel C(x,y) and the geometry of the set S where the available data are. But there is a case where this condition is automatically fulfilled. It is the case where the random part of Z(x) is reduced to a simple white noise (or, in geostatistical terms, a pure nugget effect). In this case, the covariance C(x,y) is reduced to $\varpi\delta(x-y)$, δ being the Dirac measure and ϖ a constant : any function $f^{\ell}(x)$ continuous on S is then a eigen function (for the eigen value ϖ) and this for any set S. In this case, the least squares estimator is always optimal, and the covariances μ_{js} , from (4-2-27) are given by :

$$\mu_{js} = \frac{\varpi}{S}\ S_{js} \tag{4-2-28}$$

In fact, a covariance $\varpi\ \delta(x-y)$ characterizes a measure, and no longer a random function. This presents no problem if the set S is continuous, for then we work only on regularizations (which are

again R.F.'s). of this random measure. In practice, it is sufficient for the least square estimator to be (nearly) optimal that the covariance is of the form C(x,y) with a very small range (for example, $C(h) = e^{-ah}$ with $\frac{1}{a}$ small relative to the dimensions of S). In particular, if S is finite, it is sufficient that this range is less than the smallest distance between two points of S.

In this case (covariance C(h) with a very small range) the least squares estimator is (nearly) optimal, and the covariances μ_{ij} are still given by relationship (4-2-28), provided that we take as a factor ϖ the integral C(h) extended to the whole space, i.e. :

$$\mu_{ij} = \frac{S_{js}}{S} \int C(h)\, dh$$

S_{js} still being the inverse of

$$T^{js} = \frac{1}{S} \int k(x)\, f^j(x)\, f^s(x)\, dx$$

In exercise 6, some indications for a finite set S will be found.

4-3 KRIGING.

4-3-1 Equations of Universal Kriging.

In true kriging problems, we want to estimate a weighted average of the type $Z_o = \int p(dx)\, Z(x)$ (with $\int p(dx) = 1$) from experimental data known on a set S. The equations will be determined for the case where the set S is finite, and, with the usual reservations, will be later transposed to the case where S is infinite.

Let us first assume that a centred (known) covariance $\sigma(x,y)$ and a drift $a_\ell f^\ell(x)$ ($\ell = 0,1,\ldots k$, a_ℓ unknown) exist. A linear estimator of Z_o will be formed, i.e. :

$$Z_U = Z_U^\alpha\ Z_\alpha \qquad (\text{or } Z_U = \int_S \lambda_U(dx)\, Z(x))$$

and we shall impose the usual conditions :

Universality condition : we must have $E[Z_U - Z_o] = 0$ whatever the unknown values of the coefficients a_ℓ are, which gives :

$$\begin{cases} \lambda^\alpha \, f_\alpha^\ell = b^\ell \\ \\ (b^\ell = f^\ell(x) \, p(dx)) \end{cases}$$

Optimality condition : having regards to the preceding condition, the coefficients λ_U^α are chosen in order to minimize the estimation variance

$$\begin{cases} D^2[Z_o - Z_U] = \sigma_{Z_o}^2 - 2 \, \lambda_U^\alpha \; \sigma_{Z_o \alpha} + \lambda^\alpha \, \lambda^\beta \, \sigma_{\alpha\beta} \\ \\ \sigma_{Z_o \alpha} = \int p(dx) \; \sigma(x, x_\alpha) \\ \\ \sigma_{Z_o}^2 = \int\int p(dx) \; \sigma(x,y) \; p(dy) \end{cases}$$

The following system is then obtained, in which there appear Lagrange parameters μ_ℓ associated with the universality conditions, and where the variance σ_U^2 is the optimal value of the quadratic form $D^2[Z_o - Z_U]$, computed when taking the first two equations into account :

$$(4\text{-}3\text{-}1) \qquad \begin{cases} \lambda_U^\beta \, \sigma_{\alpha\beta} = \sigma_{Z_o, \alpha} + \mu_\ell \, f_\alpha^\ell \\ \\ \lambda_U^\alpha \, f_\alpha^\ell = b^\ell \\ \\ \sigma_U^2 = \sigma_{Z_o}^2 - \lambda^\alpha \, \sigma_{Z, \alpha} + \mu_\ell \, b^\ell \end{cases}$$

This system is always regular (provided that $\sigma_{\alpha\beta}$ is strictly positive definite and the f_α^ℓ are linearly independent on S).

This system can be transposed without any difficulty to the case where only a variogram γ exists. In fact, for $\ell = 0$, as usual, $f^o = 1$. The universality condition relative to $\ell = 0$ is thus reduced to :

$$\sum_\alpha \lambda^\alpha = 1$$

But (exactly as in simple kriging, paragraph 3-5), this is exactly the condition for the error :

$$Z_o - Z_U = \int p(dx)\ Z(x) - \lambda^\alpha\ Z_\alpha$$

to be a permitted linear combination (i.e. having a finite variance). As $\int p(dx) = 1$, the sum $1 - \sum_\alpha \lambda^\alpha$ of the coefficients of the linear combination $Z_o - Z_U$ must, as usual, be equal to zero. The constant term a_o of the drift, in which all the difficulty in the optimal estimation of the drift itself was concentrated, is automatically eliminated from the expression $Z_o - Z_U$ and no longer intervenes in kriging : this is precisely the significance of the condition $\sum \lambda^\alpha = 1$.

Let us transpose the system (4-3-1) in terms of the variogram, adopting the notations of the continuous case to lend some variety to our discourse.

$$(4\text{-}3\text{-}2)\qquad \begin{cases} \int_S \lambda(dx)\ \gamma(x,y) = \int p(dx)\ \gamma(x,y) - \mu_\ell\ f^\ell(y) & (\forall\ y \in S) \\ \int_S \lambda(dx)\ f^\ell(x) = \int p(dx)\ f^\ell(x) & (\ell = 0,1,\ldots k) \\ \sigma_U^2 = -\int\int p(dx)\ \gamma(x,y)\ p(dy) + \int\int \lambda(dx)\ p(dy)\ \gamma(x,y) + \mu_\ell \int p(dx)\ f^\ell(x) \end{cases}$$

Punctual Kriging : If we wish to estimate the value $Z(x_o)$ of the realization at a point x_o, the preceding system is simplified and becomes (using the discrete case notations) :

$$(4\text{-}3\text{-}3)\qquad \begin{cases} \lambda^\beta\ \gamma_{\alpha\beta} = \gamma(x_o, x_\alpha) - \mu_\ell\ f_\alpha^\ell \\ \lambda^\alpha\ f_\alpha^\ell = f^\ell(x_o) \\ \sigma_U^2 = \lambda^\alpha\ \gamma(x_\alpha, x_o) + \mu_\ell\ f^\ell(x_o) \end{cases}$$

It will be noted that punctual universal kriging is an exact interpolator. This can be verified directly by taking $x_o = x_{\alpha_o} \in S$ in (4-3-3). It is simpler to note that Z_{α_o} verifies $E(Z_{\alpha_o} - Z_{\alpha_o}) = 0$ and $D^2(Z_{\alpha_o} - Z_{\alpha_o}) = 0$ for any a_ℓ, and thus satisfies the universality and optimality conditions.

4-3-2 Additivity theorem.

Should the drift $m(x)$ be known, the optimal estimator of $Z_o = \int p(dx)\ Z(x)$ would be :

(4-3-4) $$\lambda_K^\alpha (Z_\alpha - m_\alpha) + \int m(x)\ p(dx)$$

with the coefficients λ_K^α of simple kriging (paragraph 3-4-1), i.e. :

(4-3-5) $$\begin{cases} \lambda_K^\beta\ \sigma_{\alpha\beta} = \sigma_{\alpha, Z_o} \\ \\ \sigma_K^2 = \sigma_{Z_o}^2 - \lambda_K^\alpha\ \sigma_{\alpha, Z_o} \end{cases}$$

Let us show that the circumstance already noticed in paragraph (3-4-2) is general : in other words, the universal optimal estimator can be put in the form :

(4-3-6) $$Z_U = \lambda_K^\alpha (Z_\alpha - m_\alpha^*) + \int m^*(x)\ p(dx)$$

obtained by replacing in (4-3-4) the unknown expression of the drift by its optimal estimate $m^*(x)$. This means, if we want to, that we have a right to krige the estimated residuals $Z(x) - m^*(x)$ as if they were true residuals.

To verify (4-3-6), put

$$Z_U = Z_K + Z_D \quad , \quad Z_K = \lambda_K^\alpha\ Z_\alpha$$

The term Z_D represents the drift correction. It is of the form $Z_D = \lambda_D^\alpha\ Z_\alpha$, with $\lambda_D^\alpha = \lambda_U^\alpha - \lambda_K^\alpha$. Let us look for the conditions to be verified by the λ_D^α . If λ_U^α is replaced by $\lambda_K^\alpha + \lambda_D^\alpha$ in the first relationship (4-3-1) it follows from (4-3-5) that the λ_K^α and σ_{α, Z_o} disappear. There finally remains the following system (equivalent to (4-3-1) :

$$\begin{cases} \lambda_D^\beta\ \sigma_{\alpha\beta} = \mu_\ell\ f_\alpha^\ell \\ \\ \lambda_D^\alpha\ f_\alpha^\ell = \int p(dx)\ f^\ell(x) - \lambda_K^\alpha\ f_\alpha^\ell \end{cases}$$

This system is equivalent to (4-2-2) with the exception that $f^\ell(x_o)$ is replaced by $\int [p(dx) - \lambda_K(dx)]\ f^\ell(x)$. Consequently, if we put :

(4-3-7) $$b^\ell = \int p(dx)\ f^\ell(x) - \lambda_K^\alpha\ f_\alpha^\ell$$

the λ_D^α are deduced from the λ_ℓ^α of the delinearized system (4-2-6) by the relationships :

$$\lambda_D^\alpha = \lambda_\ell^\alpha \, b^\ell$$

The term Z_D representing the drift correction is thus $\lambda_\ell^\alpha \, Z_\alpha \, b^\ell = A_\ell b^\ell$, A_ℓ being the optimal estimator of the drift, and, replacing b^ℓ by its expression (4-3-7), this gives exactly :

$$Z_D = \int p(dx) \; m^*(x) - \lambda_K^\alpha \, m_\alpha^*$$

Hence, Z_U itself is truly of the form (4-3-6) :

$$Z_U = Z_K + Z_D = \lambda_K^\alpha \, (Z_\alpha - m_\alpha^*) + \int m^*(x) \; p(dx)$$

Always as in the case of simple kriging, this additivity theorem extends to the variances, i.e. :

$$D^2(Z_o - Z_U) = \sigma_K^2 + \sigma_D^2 \qquad (4\text{-}3\text{-}7)$$

with $\sigma_D^2 = D^2(Z_D)$ and σ_K^2 the simple kriging variance. Indeed the system (4-3-5) expresses exactly that $Z_o - Z_K$ has a zero covariance with the linear combinations of the Z_α , and thus, in particular with Z_D.

It can be verified without difficulty that the additivity theorem remains when there is no covariance, but only a variogram (in fact, for $\ell = 0$, the universality condition relative to λ_D^α is written :

$$\sum_\alpha \lambda_D^\alpha = \int p(dx) - \int \lambda_K(dx) = 0$$

and Z_D is a permitted linear combination).

4-3-3 Kriging considered as an interpolator.

We have already noticed that kriging (simple or universal) constitutes an exact interpolator. This is an important property in certain applications like contouring. From the methodological point of view, it may not be a useless exercise to forget for a moment the probabilistic contents of the theory, and to analyze from a purely pragmatical point of view the manipulations to which the experimental data have been subjected when kriging a point x.

a/ Definition of interpolators. - Let $S \subset \mathbb{R}^n$ be the set of experimental points, and V a domain containing S which may coincide with $\mathbb{R}^n$ itself. We will define an interpolator as a linear operator T mapping a space $\mathfrak{F}(S)$ of functions defined on S into a space $\mathfrak{F}(V)$ of functions defined on V : $f \in \mathfrak{F}(S) \rightarrow T f \in \mathfrak{F}(V)$. T will be said to be an exact interpolator if the restriction to S of Tf is identical to f.

In what follows, we shall limit ourselves to the case where the set $S = \{x_\alpha , \alpha = 1,...N\}$ is finite. A function f on S is then a vector (f_α), $\alpha = 1,...N$, and the operator T is defined by the datum of the $T e^\alpha$, e^α being a basis of the N-dimensional space constituted by these vectors. For example, let us take for e^α the vector of components $e^\alpha_\beta = \delta^\alpha_\beta$, and let $T^\alpha(x)$ be the image of this vector. for any vector f, we have then :

$$T f(x) = f_\alpha T^\alpha(x) \tag{4-3-8}$$

The interpolator T is exact if and only if the matrix $T^\alpha_\beta = T^\alpha(x_\beta)$ verifies $f_\alpha T^\alpha_\beta = f_\beta$ for any $f \in \mathbb{R}^n$, and thus verifies

$$T^\alpha_\beta = \delta^\alpha_\beta \tag{4-3-9}$$

Example : simple kriging is the interpolator λ_K defined by the $\lambda^\alpha_K(x)$ obtained by solving the system :

$$\lambda^\beta_K(x) \, \sigma_{\alpha\beta} = \sigma_{\alpha,x} \tag{4-3-10}$$

Let $B^{\alpha\beta}$ be the inverse matrix of $\sigma_{\alpha\beta}$. We have then

$$\lambda^\beta_K(x) = B^{\alpha\beta} \, \sigma_{\alpha,x} \tag{4-3-11}$$

For $x = x_\gamma$, we get $\lambda^\beta_K(x_\gamma) = B^{\alpha\beta} \, \sigma_{\alpha\gamma} = \delta^\beta_\gamma$, and simple kriging is an exact interpolator. Moreover, the components $\lambda^\beta_K(x)$ are, from (4-3-11), linear combinations of the N functions $\sigma_{\alpha,x} = \sigma(x_\alpha, x)$.

These two properties are characteristic : simple kriging is the only exact interpolator whose components are linear combinations of the functions $\sigma_{\alpha,x}$. Indeed, let T be an interpolator of the form :

$$T^{\beta}(x) = H^{\beta\alpha}\, \sigma_{\alpha,x}$$

$T^{\beta}(x)$ is an exact interpolator if and only if $T^{\beta}_{\alpha} = \delta^{\beta}_{\alpha}$, hence :

$$\delta^{\beta}_{\alpha} = H^{\beta\gamma}\, \sigma_{\alpha\gamma}$$

Thus $H^{\beta\gamma} = B^{\beta\gamma}$, and $T^{\beta}(x) = \lambda^{\beta}_{K}(x)$.

In order to characterize universal kriging, we need a new notion :

b/ Drifts compatible with an interpolator. - A function $f(x)$, $x \in V$ is a drift consistent with an interpolator T if the interpolation of f by T is exact, in other words if :

(4-3-12) $$f_{\beta}\, T^{\beta}(x) = f(x)$$

The familiar condition of universality can be recognized in (4-3-12). For $x = x_{\alpha}$, this condition entails :

(4-3-13) $$T^{\beta}_{\alpha}\, f_{\beta} = f_{\alpha}$$

In other words, f_{α} must be the eigen vector for the eigen value 1. Conversely, let f_{α} be a vector verifying (4-3-13). The function $f(x)$ defined on V by :

$$f(x) = f_{\beta}\, T^{\beta}(x)$$

is by construction a drift consistent with T. So (4-3-12) establishes a biunivocal correspondence between the consistent drifts and the eigen vectors associated with the proper value 1.

T is an exact interpolator if and only if $T^{\alpha}_{\beta} = \delta^{\alpha}_{\beta}$, and this is so if and only if all the vectors f_{α} are proper vectors for the eigen value 1. Consequently, T is an exact interpolator if and only if there exist N drifts consistent with T and linearly independent on S.

c/ Characterization of U.K. - In the case of simple kriging, these N drifts are the functions $\sigma_{\alpha,x}$: this results from the system (4-3-9) itself. Consider universal kriging in the case where the drift is of the form $a_{\ell}\, f^{\ell}(x)$, with $\ell = 1,2,\ldots k$ $(k \leq N)$. It defines an interpolator λ whose

components $\lambda^{\alpha}(x)$ are the solution of :

$$(4\text{-}3\text{-}11) \qquad \begin{cases} \lambda^{\beta}(x)\ \sigma_{\alpha\beta} = \sigma_{\alpha,x} + \mu_{\ell}\ f_{\alpha}^{\ell} \\ \\ \lambda^{\alpha}(x)\ f_{\alpha}^{\ell} = f^{\ell}(x) \qquad (\ell = 1,2,\ldots k) \end{cases}$$

The universality conditions show that the functions f^{ℓ} are drifts which are consistent with this interpolator. It being understood that these k functions f^{ℓ} are linearly independent on S, we recognize therefore a linear space of k dimensions of drifts consistent with λ. Finally, to characterize universal kriging, there remains the determination of a supplementary space of N-k dimensions.

Let f^{ν} , $\nu = k+1,\ldots N$ be a basis of this supplementary space. The N-k missing drifts will thus be the $f_{\beta}^{\nu}\ \lambda^{\beta}(x)$. To determine the f^{ν} , we shall use the <u>additivity theorem</u> of paragraph 4-3-2. From this theorem, the interpolation of the data Z_{α} leads to the function :

$$Z_U(x) = \lambda_K^{\alpha}(x)\ (Z_{\alpha} - A_{\ell}\ f_{\alpha}^{\ell}) + A_{\ell}\ f^{\ell}(x)$$

with coefficients A_{ℓ} of the form :

$$A_{\ell} = \lambda_{\ell}^{\beta}\ Z_{\beta}$$

where the matrix λ_{ℓ}^{β} is the solution of the system (4-2-6). Making the expression of $Z_U(x)$ explicit, we find :

$$Z_U(x) = \lambda_K^{\beta}(x)\ (\delta_{\beta}^{\alpha} - \lambda_{\ell}^{\alpha}\ f_{\beta}^{\ell})\ Z_{\alpha} + \lambda_{\ell}^{\alpha}\ f^{\ell}(x)\ Z_{\alpha}$$

Hence, the coefficients $\lambda^{\beta}(x)$ of the system (4-3-11) have the following expressions :

$$\lambda^{\alpha}(x) = \lambda_K^{\beta}(x)\ (\delta_{\beta}^{\alpha} - \lambda_{\ell}^{\alpha}\ f_{\beta}^{\ell}) + \lambda_{\ell}^{\alpha}\ f^{\ell}(x)$$

We recognize the expression of the matrix U :

$$U_{\beta}^{\alpha} = \delta_{\beta}^{\alpha} - \lambda_{\ell}^{\alpha}\ f_{\beta}^{\ell}$$

introduced in paragraph 4-2-4 (the residual $Z_{\alpha} - m_{\alpha}^{*}$ is $U_{\alpha}^{\beta}\ Z_{\beta}$). Hence :

(4-3-12) $$\lambda^{\alpha}(x) = \lambda^{\beta}_{K}(x)\ U^{\alpha}_{\beta} + \lambda^{\alpha}_{\ell}\ f^{\ell}(x)$$

From the form of the universality condition ($\lambda^{\alpha}_{\ell}\ f^{s}_{\alpha} = \delta^{s}_{\ell}$) it is possible to choose the basis f^{v} of the missing drifts in the <u>orthogonal</u> of the λ^{α}_{ℓ} ($\ell = 1,2,\ldots k$) i.e. to give ourselves the relationships :

(4-3-13) $$\lambda^{\alpha}_{\ell}\ f^{v}_{\alpha} = 0$$

This condition fulfilled, it follows from (4-3-13) that the N-k missing drifts are the functions :

$$f^{v}(x) = \lambda^{\beta}(x)\ f^{v}_{\beta} = \lambda^{\beta}_{K}(x)\ f^{v}_{\beta}$$

i.e. :

$$f^{v}(x) = f^{v}_{\beta}\ B^{\alpha\beta}\ \sigma_{\alpha,x}$$

They are <u>linearly</u> expressed with the help of the functions $\sigma_{\alpha,x}$, i.e. :

(4-3-14) $$\begin{cases} f^{v}(x) = D^{\alpha}\ \sigma_{\alpha,x} \\ \\ D^{\alpha} = f^{v}_{\beta}\ B^{\alpha\beta} \end{cases}$$

We still have to characterize the N-k dimensional space of the vectors D^{α} such that the $D^{\alpha}\ \sigma_{\alpha,x}$ are consistent drifts. To do this, consider the conditions (4-3-13), noticing that the λ^{α} of the system (4-2-6) are given by :

$$\lambda^{\alpha}_{\ell} = \mu_{\ell s}\ f^{s}_{\beta}\ B^{\alpha\beta}$$

The relationship $f^{v}_{\alpha}\ \lambda^{\alpha}_{\ell} = 0$ is then equivalent (since the matrix $\mu_{\ell s}$ of the Lagrange parameters is strictly positive definite) to the relationship :

$$f^{v}_{\alpha}\ B^{\alpha\beta}\ f^{s}_{\beta} = 0$$

From (4-3-14), this is also equivalent to :

(4-3-15) $$D^{\alpha}\ f^{\ell}_{\alpha} = 0$$

Consequently, the supplementary N-k dimensional space is constituted by functions of the form $D^\alpha \sigma_{\alpha,x}$ for vectors D^α orthogonal to the k vectors f^ℓ, i.e. verifying (4-3-15).

d/ Probabilistic contents of universal kriging. - As always for any exact interpolator, universal kriging is characterized by the (N-dimensional)space of the drifts consistent with it. This space contains

1 - An imposed space of dimension k (that of the functions $a_\ell f^\ell$) the significance of which is purely interpolatory (non probabilistic).

2 - A supplementary space of N-k dimensions (that of the functions $D^\alpha \sigma_{\alpha,x}$, with the orthogonality condition $D^\alpha f^\ell_\alpha = 0$ expressing that the vector D describes the orthogonal of the former space) with a purely probabilistic significance.

In other words, when the number of imposed drifts increases, the probabilistic contents of universal kriging decreases while its purely interpolatory contents increases.

In particular, for k = 0, simple kriging is purely probabilistic. For k = N, on the contrary the probabilistic contents are null, and universal kriging is reduced to the purely interpolatory process which consists of fitting a linear combination of N functions chosen beforehand to N experimental points.

4-4 UNIVERSAL KRIGING FOR A RANDOM DRIFT.

4-4-1 Hypotheses.

Instead of considering the drift as a function m(x),unknown but not random, we may consider the case where m(x) is a random function $m(\underline{x},\omega)$, in other words, study a R.F. $Z(x,\omega)$ of the form :

(4-4-1) $$Z(x,\omega) = Y(x,\omega) + m(x,\omega)$$

with $E[Y(x)]= 0$. Put :

$$\begin{cases} E[Y(x)\ Y(y)] = \sigma(x,y) \\ E[m(x)\ m(y)] = K(x,y) \\ E[Y(x)\ m(y)] = R(x,y) \end{cases}$$

(the second covariance is not centered).

It is not necessary that the R.F.'s Y and m be independent. But, in order that the dichotomy written in (4-3-1) should have a real physical significance, and not be a pure artefact, each of the two components Y and m has to take into account structural features relating to distinctly differentiated scales (gigogne structures of J. SERRA). The R.F. m must have a very great (or infinite) range with respect to that of Y, and moreover, at the scale of the range of Y, the variability of m must be very small. At the scale of Y, the realization $m(x,\omega)$ must thus be assimilated to a very regular and continuous function, so that - on a certain domain V - we may write :

$$m(x,\omega) \simeq a_\ell(\omega)\ f^\ell(x) \qquad (x \in V) \tag{4-4-2}$$

the $a_\ell(\omega)$ designating r.v.'s (which depend on the domain V), and the f^ℓ the usual basic functions (with $f^0 = 1$) : for example polynomials or, more generally, continuous and sufficiently regular functions, linearly independent on V.

Let us examine the <u>significance</u> of the hypothesis (4-4-2). We shall put :

$$K_{\ell s} = E(a_\ell\ a_s) \tag{4-4-3}$$

so that the covariance K(x,y) of the R.F. m(x) has, in the domain V where the approximation (4-4-2) is valid, the following representation :

$$K(x,y) \simeq K_{\ell s}\ f^\ell(x)\ f^s(y) \qquad (x,y \in V) \tag{4-4-4}$$

For the covariance R(x,y), we find

$$R(x,y) \simeq E[Y(x)\ a_\ell f^\ell(y)] = R_\ell(x)\ f^\ell(y) \qquad (x,y \in V)$$

If the R.F. Y can be considered as stationary on V, $R_\ell(x)$ is a constant R_ℓ . More generally, $R_\ell(x)$

will have on V a good approximate representation by means of the same functions $f^s(x)$, i.e. :

$$R(x) \simeq R_{\ell s}\, f^s(x) \qquad (x \in V)$$

(this assumption is not absolutely required, but it makes the argument much simpler) and so, finally, we have for R(x,y) a result analogous to (4-4-4) :

(4-4-4') $$R(x,y) \simeq R_{\ell s}\, f^{\ell}(x)\, f^s(y) \qquad (x,y \in V)$$

Furthermore, the two relationships (4-4-4) and (4-4-4') appear to be the only assumptions which are really required for what follows, and relationship (4-4-2) is only a more obvious translation of them. From a purely pragmatical point of view, they only mean that the functions K(x,y) and R(x,y) vary slowly enough in space at the scale of the domain V of reference to be replaced on V by the first terms of their limited expansion with respect to $f^{\ell}(x)\, f^s(y)$. The following example will show the significance of this hypothesis.

Example of a Gaussian exponential.- For easier comprehension, consider the one-dimensional space $\mathbb{R}^1$ (we would, of course, get similar results in $\mathbb{R}^n$) and suppose that the covariance K has the form :

$$K(x,y) = B\, e^{-\frac{(x-y)^2}{2}}$$

with a range b greater than the length L of the reference interval V. Develop this Gaussian exponential :

$$K(x,y) = B\left[1 - \frac{(x-y)^2}{2\,b^2} + \frac{1}{8}\frac{(x-y)^4}{b^4} - \frac{1}{48}\frac{(x-y)^6}{b^6} + \dots\right]$$

each of the terms $(x-y)^{2n}$ can be expanded in its turn in monomials $x^{\ell}\, y^s$, and we see that K(x,y) has effectively a limited expansion of the type (4-4-4) with $f^{\ell}(x) = x^{\ell}$.

In the case of a "quadratic drift", i.e. $\ell = 0, 1, 2$, this expansion is taken to the fourth order term, and the error is bounded by

$$\frac{B}{48}\,\frac{(x-y)^6}{b^6} \leq \frac{4}{3}\, B\,\frac{L^6}{b^6}$$

To appreciate this upper bound, notice that B can be very great (even, at the limit, infinite if m has no covariance but only a variogram), but B/b^2 is necessarily finite for it represents the coefficient of the term in h^2 of the variogram of m. Writing $B/b^2 = B_2$, we see that the upper bound is in $B_2 L^2 (\frac{L}{b})^4$, and thus of order at least 4 in L/b. If the range b is effectively great with respect to L, the approximation made will be absolutely legitimate.

Note. - From (4-4-1), the covariance of Z would be $\sigma + 2R + K$. We may wonder whether this dichotomy would be really useful, and if it would not be simpler to perform ordinary kriging or cokriging directly on this covariance. The answer is that (in the conditions of our hypotheses) the estimation of this covariance is not possible in acceptable conditions. The covariance K(x,y) is not, in general, a stationary covariance. Even if, locally (that is at the scale of V) it can be put in the form K(x-y), statistical inference starting from data covering a domain the size of V will not be possible in general with a covariance which is too regular.(see par. 2-10).

4-4-2 Estimation of a drift.

Let Z_α be the value of Z(x) at the experimental points $x_\alpha \in S \subset V$. S will be supposed to be finite, but what follows can be transposed without difficulty to the case where S is infinite.

The drift m(x) is to be estimated at a point $x \in V$, and for this we use an estimator of the form :

$$M^*(x) = \lambda^\alpha Z_\alpha = \lambda^\alpha Y_\alpha + \lambda^\alpha m_\alpha$$

The expectation of the square of the error is thus

$$E[M^*(x) - m(x)]^2 = E[\lambda^\alpha Y_\alpha + (\lambda^\beta m_\beta - m_x)]^2 =$$

$$= \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} + 2 \lambda^\alpha (\lambda^\beta R_{\alpha\beta} - R_{\alpha x}) + \lambda^\alpha \lambda^\beta K_{\alpha\beta} - 2 \lambda^\alpha K_{\alpha x} + K_{xx}$$

Hypotheses (4-4-4) and (4-4-4') give :

$$\lambda^\beta R_{\alpha\beta} - R_{\alpha x} = R_{\ell s} (\lambda^\beta f_\beta^s - f_x^s) f_\alpha^\ell$$

$$\lambda^\alpha \lambda^\beta K_{\alpha\beta} - 2 \lambda^\alpha K_{\alpha x} + K_{xx} = K_{\ell s} (\lambda^\alpha \lambda^\beta f_\alpha^\ell f_x^s - \lambda^\alpha f_\alpha^\ell f_x^s -$$

$$- \lambda^\beta f_x^\ell f_\beta^s + K_{xx}) = K_{\ell s} (\lambda^\alpha f_\alpha^\ell - f_x^\ell) (\lambda^\beta f_\beta^s - f_x^s)$$

Hence :

(4-4-5)
$$\begin{cases} E(M_x^* - m_x)^2 = \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} + 2 \lambda^\alpha f_\alpha^\ell R_{\ell s} (\lambda^\beta f_\beta^s - f_x^s) \\ \\ \qquad + K_{\ell s} (\lambda^\alpha f_\alpha^\ell - f_x^\ell) (\lambda^\beta f_\beta^s - f_x^s) \end{cases}$$

<u>The significance of the universality conditions</u> appear clearly here. We cannot hope to minimize the quadratic form (4-4-5) for any λ^α, for we do not know the $K_{\ell s}$ ans $R_{\ell s}$; hence we have to impose on the λ^α supplementary conditions by which (4-4-5) will no longer depend on $R_{\ell s}$ and $K_{\ell s}$. The matrix $K_{\ell s}$ always being positive definite, these conditions are :

(4-4-6)
$$\lambda^\alpha f_\alpha^\ell = f_x^\ell$$

i.e. the usual universality conditions. Taking these conditions into account, there remains :

(4-4-5')
$$E(M_x^* - m_x)^2 = \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta}$$

It is enough then to minimize (4-4-5'), taking (4-4-6) into account, to find again <u>the same system as in the usual theory</u> - <u>and the same estimation variance</u> (it can be easily verified that $E(M_x^* - m_x) = E(\lambda^\alpha m_\alpha - m_x) = 0$ when $m(x,\omega)$ - without being obligatorily stationary - has itself a drift $E\, m(x)$ representable on V by an expression of the form $a_\ell f^\ell(x)$). Given the assumptions (4-4-4) and(4-4-4') <u>the random drift is amenable to the same treatment as the functional drift</u>.

By delinearizing the system with respect to the right hand side of (4-4-6), we get, as in the classical case, the estimators

$$A_\ell = \lambda_\ell^\alpha Z_\alpha$$

of the coefficients $a_\ell = a_\ell(\omega)$ of the drift $m(x) = a_\ell f^\ell(x)$ $(x \in V)$, and the Lagrange factors $\mu_{\ell s}$ keep their significance. More precisely - the a_ℓ being r.v.'s, we have :

$$\begin{cases} E(A_\ell - a_\ell) = 0 \\ E(A_\ell - a_\ell)(A_s - a_s) = \mu_{\ell s} \end{cases}$$

4-4-3 An equivocal example.

Let $Z(x,\omega)$ be an I.R.F. without drift, i.e. :

$$\begin{cases} E[Z(x) - Z(y)] = 0 \\ \frac{1}{2} E[Z(x+h) - Z(x)]^2 = \gamma(h) \end{cases}$$

and let p be a very regular weighting function with a sum equal to 1 (for example a Gaussian exponential). Put :

$$\begin{cases} m(x) = Z * \check{p} = \int Z(x+y)\ p(y)\ dy \\ Y(x) = Z - Z * \check{p} \end{cases}$$

In the expansion

$$Z = Y + m$$

m is a moving average, Y the "residual" of Z around this moving average. The variogram of m is $\gamma * p * \check{p} - < \gamma,\ p * \check{p} >$; because of the regularity of p, it has an expansion of the type (4-4-4) in a domain V with small dimension in respect to the range of p. We are thus in the required conditions to estimate the moving average m(x), considered as a random drift, by the methods of Universal Kriging.

Y is a stationary R.F. of order two. Its covariance σ is deduced from the variogram γ by the relationship

$$\sigma = - \gamma * (\delta - p - \check{p} + p * \check{p})$$

which is explicitly (with $P = p * \check{p}$) :

(4-4-7) $$\sigma(h) = -\gamma(h) + \int\gamma(y)[p(y+h) + p(y-h)]\,dy - \int\gamma(y)\,P(y+h)\,dy$$

In particular :

(4-4-7') $$\sigma^2 = \sigma(0) = 2\int\gamma(x)\,p(x)\,dx - \int\gamma(x)\,P(x)\,dx$$

This covariance (4-4-7) is the one which must appear in the left hand side of the optimality conditions of the system of U.K.. But the conditions imposed on p (regularity, and large range with respect to the neighbourhood V) entail that $\sigma(x-y)$ itself has an expansion of the type (4-4-4), except for the term $\gamma(h)$, i.e. :

(4-4-8) $$\sigma(x-y) \simeq -\gamma(x-y) + P_{\ell s}\, f^{\ell}(x)\, f^{s}(y) \quad (x,y \in V)$$

The quantity to be minimized in order to estimate $m(x)$ becomes :

$$\lambda^{\alpha}\lambda^{\beta}\sigma_{\alpha\beta} = -\lambda^{\alpha}\lambda^{\beta}\gamma_{\alpha\beta} + P_{\ell s}\,\lambda^{\alpha} f_{\alpha}^{\ell}\,\lambda^{\beta} f_{\beta}^{s}$$

But the universality conditions (4-4-6) then require :

$$\lambda^{\beta}\lambda^{\alpha}\sigma_{\alpha\beta} = -\lambda^{\alpha}\lambda^{\beta}\gamma_{\alpha\beta} + P_{\ell s}\, f_{x}^{\ell}\, f_{x}^{s}$$

From (4-4-8), we also have :

$$\sigma(0) = P_{\ell s}\, f_{x}^{\ell}\, f_{x}^{s}$$

Hence finally - the conditions (4-4-6) supposedly being verified - we have only to minimize the expression :

$$\lambda^{\alpha}\lambda^{\beta}\sigma_{\alpha\beta} = \sigma(0) - \lambda^{\alpha}\lambda^{\beta}\gamma_{\alpha\beta}$$

Hence the system :

(4-4-9) $$\begin{cases} -\lambda^{\beta}\gamma_{\alpha\beta} = \mu_{\ell}\, f_{\alpha}^{\ell} \\ \lambda^{\alpha} f_{\alpha}^{\ell} = f_{x}^{\ell} \end{cases}$$

and the corresponding variance is :

$$D^2(M_x^* - m_x) = \sigma(0) + \mu\rho\, f_x^\rho \qquad (4\text{-}4\text{-}9')$$

It is noted that the system (4-4-9) does not depend on the weighting function p (provided that it is regular and has a large range) : the same estimator $\lambda^\alpha Z_\alpha$ applies to all the moving averages, and it is only on the variance (4-4-9') that the choice of the function p has any influence by means of the variance $\sigma(0)$ defined in (4-4-7)- and which is nothing else but the a priori variance of the residual $Y(x) - m(x)$.

This ambiguous result is worthy of comment :

Note 1 - The variogram γ being known; the estimation of the moving average $m(x)$ is in fact subject to ordinary kriging. Instead of the system (4-4-9), we should use the following system (4-4-10) :

$$\begin{cases} \lambda^\beta \gamma_{\alpha\beta} = \int \gamma(y)\, p(y-x+x_\alpha)\, dy - \mu_o \\ \sum_\alpha \lambda^\alpha = 1 \end{cases} \qquad (4\text{-}4\text{-}10)$$

which includes only one universality condition ($\sum_\alpha \lambda^\alpha = 1$), but where, on the other hand, the right hand side of the optimality condition contains the variogram γ, and involves explicitly the values taken by $\gamma(h)$ for distances $|h|$ considerably larger than the dimensions of the set S of experimental data x_α . This system (4-4-10) hence refers to a situation where we know very well the true behaviour of $\gamma(h)$ up to fairly large distances. On the contrary, in the system (4-4-9), the variogram γ intervenes only by its terms $\gamma_{\alpha\beta}$, and implies a knowledge of γ only over small distances.

This question can then be seen in a different way. Let us assume (and the statistical inference theory shows that this is a fairly general circumstance) that we are sure of our variogram $\gamma(h)$ only at small distances (of the order of the dimensions of S) so that we have a good knowledge of $\gamma_{\alpha\beta}$ but not of the right hand side of (4-4-10) : in this case, the system (4-4-9) gives us the best possible estimation of a large moving average $m(x)$, taking the information actually available into account.

Note 2 - In the absence of an actual drift, the optimal estimation of the drift is not rendered meaningless : it measures a large moving average - without it being possible to specify which one -

for it applies to any. The variance (4-4-9') of this estimator is all the larger as it is supposed to represent a more extended moving average. If we put :

$$m(x) = a_\ell f^\ell(x) \quad , \quad M^*(x) = A_\ell f^\ell(x)$$

we obtain the $A_\ell = \lambda_\ell^\alpha Z_\alpha$ by delinearizing the system (4-4-9) :

$$(4\text{-}4\text{-}11) \qquad \begin{cases} -\lambda_\ell^\beta \gamma_{\alpha\beta} = \mu_{\ell s} f_\alpha^\ell \\ \\ \lambda_\ell^\alpha f_\alpha^s = \delta_\ell^s \end{cases} \qquad (\ell, s = 0, 1, \ldots n)$$

From (4-4-9'), the variance is then :

$$D^2(M_x^* - m_x) = \mu_{\ell s} f_x^\ell f_x^s + \sigma(0)$$

Hence :

$$(4\text{-}4\text{-}11') \qquad \begin{cases} E(A_o - a_o)^2 = \sigma(0) + \mu_{oo} \\ \\ E(A_\ell - a_\ell)(A_s - a_s) = \mu_{\ell s} \end{cases} \qquad (\ell s \neq 0)$$

Hence the Lagrange parameters - except for μ_{oo} - still represent the covariances of the $(A_\ell - a_\ell)$. On the other hand, the term μ_{oo} - relative to the coefficients λ_o^α of sum equal to 1 - does not give the variance of $(A_o - a_o)$: it has to be overvalued by the term $\sigma(0)$, which is the only term depending on p. Thus, this remark suggests a solution to the apparently unsolvable problem posed by the constant term A_o of a drift when there is no covariance.

4-4-4 The problem of the constant term A_o .

Let $Z(x) = Y_1(x) + m_1(x)$ a R.F. having a (functional or random) drift $m_1(x)$, and a component $Y_1(x)$ without covariance and with a variogram $\gamma(h)$. We can write :

$$Y(x) = Y_1(x) - \int Y_1(x+y)\, p(y)\, dy \qquad Yn = Y_1 - Y_1 * p$$

$$m(x) = m_1(x) + \int Y_1(x+y)\, p(y)\, dy \qquad m = m_1 + Y_1 * p$$

In this new expansion, $Z = Y + m$, the component Y has a covariance $\sigma(x,y)$, and the(random) drift verifies the hypotheses (4-4-4) and (4-4-4'), provided that the weighting function is sufficiently regular and has a great enough range. The optimal estimation $M^*(x) = A_\ell f^\ell(x)$, $A_\ell = \lambda_\ell^\alpha Z_\alpha$ of the drift m(x) is obtained by solving the same system (4-4-11) as above, and leads in particular to an estimator A_o : the corresponding variance $E(A_o - a_o)^2 = \sigma(0) + \mu_{oo}$ depends explicitly on the $\sigma(0)$ defined in (4-4-7'), that is on the choice of the weighting function p which we used to give a sense to this expression (subtraction of a moving average). The numerical value does not depend on it.

But the system (4-4-11) is nothing else but the system (4-2-5) completed by the equations relative to $\ell = 0$. Thus, to use this completed system and to deduce an estimation A_o of the constant term a_o (which, properly speaking, did not exist within our hypotheses) is not absurd. The expression $A_o + \sum_{\ell=1}^{n} A_\ell f^\ell(x)$ gives an estimation of the true drift corrected by a moving average, and its variance is the estimation variance of the drift so corrected. This variance has only a relative significance (since it depends on the choice of the moving average). But, as the estimator itself

$$M^*(x) = A_o + \sum_{\ell=1}^{n} A_\ell f^\ell(x)$$

does not depend on this choice, it can be used to contour the drift without any problem of coupling between the different working domains V_1 , V_2 ,.... More precisely, a neighbourhood V_x (translate of V centred at x) can be assigned to each point x, and then the set $S_x \subset V_x$ of the experimental points contained in V_x . It is then sufficient to solve the system (4-4-9) with the help of the $x_\alpha \in S_x$ to form the estimator $M^*(x)$ of the drift at x. Putting the origin of coordinates at this point x, and with $f^\ell(o) = 0$ ($\ell = 1,2,...n$) (in the case of polynomials) it is in fact $M^*(0) = A_o$ which is estimated for any V_x, and this method of the sliding neighbourhood allows the drift to be mapped.

4-4-5 Kriging.

The case of kriging will appear to be very similar. To shorten the notations, we shall limit ourselves to the case of kriging a point $x \in V$ (although the general case may be studied in the same way). We suppose that the hypotheses of paragraph 4-4-1 are still verified. We want to form an estimator $Z_x^* = \lambda^\alpha Z_\alpha$ of $Z_x = Z(x)$. Let us consider the expression :

$$Z_x^* - Z_x = \lambda^\alpha Z_\alpha - Z_x = \lambda^\alpha Y_\alpha - Y_x + \lambda^\alpha m_\alpha - m_x$$

and take the expectation of its square :

$$E(Z_x^* - Z_x)^2 = \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} - 2 \lambda^\alpha \sigma_{\alpha x} + \sigma_{xx} + 2 E(\lambda^\alpha Y_\alpha - Y_x)(\lambda^\beta m_\beta - m_x) + E(\lambda^\alpha m_\alpha - m_x)^2$$

Taking (4-4-4) and (4-4-4') into account, we get :

$$\begin{cases} E(\lambda^\alpha Y_\alpha - Y_x)(\lambda^\beta m_\beta - m_x) = R_{\ell s} (\lambda^\alpha f_\alpha^\ell - f_x^\ell)(\lambda^\beta f_\beta^s - f_x^s) \\ \\ E(\lambda^\alpha m_\alpha - m_x)^2 = K_{\ell s} (\lambda^\alpha f_\alpha^\ell - f_x^\ell)(\lambda^\beta f_\beta^s - f_x^s) \end{cases}$$

Thus, in this case also, $E(Z_x^* - Z_x)^2$ is independent of the (unknown) $K_{\ell s}$ if and only if the universality conditions (4-4-6) are verified. Provided these conditions hold, we get $E(Z_x^* - Z_x) = 0$ and :

$$D^2(Z_x^* - Z_x) = \lambda^\alpha \lambda^\beta \sigma_{\alpha\beta} - 2 \lambda^\alpha \sigma_{\alpha x} + \sigma_{xx}$$

By minimizing this quadratic form, taking the universality conditions (4-4-6) into account, we get once again the usual U.K. system for functional drifts.

4-5 COKRIGING.

In certain problems, several regionalized variables have to be studied andestimated simultaneously. In [4] or [5], a systematic study of coregionalizations will be found, with the two usual aspects (transitive, i.e. non probabilistic methods, and intrinsic theory). Hence, we shall limit ourselves to seeing what the problem of Universal Kriging becomes when we deal simultaneously with several non-stationary R.F.'s. It will be seen that from a conceptual point of view, no fresh element appears. All the difficulties (which are real) come from the notation, which is necessarily more complex. To avoid handling three systems of indices, we shall use the notation of the continuous case where measures intervene instead of coefficients. This in fact is the most general notation, equally valid for the infinite case and for the finite case.

4-5-1 Notation.

Let $Z_i(x)$, $i = 1,2,\ldots d$, be d random functions (non stationary in general) defined on the space $\mathbb{R}^n$. Each one of them possesses a drift $m_i(x)$ representable on a certain domain V_i by means of a linear combination (with unknown coefficients) of given functions $f_i^\ell(x)$, $\ell = 0,1,\ldots k_i$. We shall limit ourselves to the case where these R.F.'s have covariance functions, and the task to proceed to the case where some of them have only a variogram will be left to the reader. In addition to the (centred) covariances of the d R.F.'s Z_i , it is advisable to introduce the cross-covariances. We shall put

$$\sigma_{ij}(x,y) = E[Z_i(x)\ Z_j(y)] - m_i(x)\ m_j(y) \tag{4-5-1}$$

From a theoretical point of view, the problem of estimating these drifts and of kriging these R.F.'s when their realizations are known on sets S_1 , S_2 , $\ldots S_d$ (distinct in general) is not different from the similar problems already solved in the case of a single R.F. Indeed, we have never once used the fact that the definition space of this single R.F. was the Euclidean space $\mathbb{R}^n$, and all the results obtained remain valid for a R.F. defined on any abstract space E. In particular, we can take as a definition space the product space

$$E = \mathbb{R}^n \times D$$

of $\mathbb{R}^n$ by the set $D = \{1,2,\ldots d\}$ of the indices. An element of E will be a couple constituted by a point $x \in \mathbb{R}^n$ and an index $i \in D$. The d random functions $Z_i(x)$ then constitute a unique R.F. in this product space, and it will be denoted by $Z(x,i)$. In the same way, the set of the d drifts $m_i(x)$ constitutes a single drift $m(x,i)$ in the product space, and this drift has - in a certain domain V of $\mathbb{R}^n \times D$ defined by the d domains V_i - a representation of the form :

$$m(x,i) = a_\ell\ f^\ell(x,i) \qquad (x,i) \in V \tag{4-5-2}$$

The function $f^\ell(x,i)$ on $\mathbb{R}^n \times D$ is defined in d functions $f_i^\ell(x)$ on $\mathbb{R}^n$ and $(x,i) \in V$ denotes the set of the couples (x,i) such that $x \in V_i$. In the same way, instead of $\sigma_{ij}(x,y)$ we should write

$$\sigma[(x,i)(y,j)] = E[Z(x,i)\ Z(y,j)] - m(x,i)\ m(y,j)$$

For convenience, we shall keep the notation $\sigma_{ij}(x,y)$.

As all the previous results remain valid for the R.F. $Z(x,i)$, there is little point in going through the proofs again, as we only have to retranscribe these results in the notation of the new system. However, it is necessary to take care concerning the linear independence condition of the f^{ℓ} on S, as this is the sole condition which guarantees the existence and uniqueness of the solutions of our systems.

Let S_i , $i \in D$ be the set of experimental data on which the realization of the R.F. $Z_i(x)$ is known, with in general $S_i \neq S_j$ when $i \neq j$. Some of these sets can be void if we have no experimental information about the corresponding R.F.'s. In the product space notation, the set S of experimental points (x,i) where the realization of $Z(x,i)$ is known is the set :

$$S = \{(x,i) : x \in \mathbb{R}^n , i \in D , x \in S_i\}$$

of the couples (x,i) such that the point x belongs to the set S_i .

The linear independence condition on S of the functions $f^{\ell}(x,i)$ is thus formulated :

$$c_{\ell} f^{\ell}(x,i) = 0 \text{ for any } i \in D \text{ and any } x \in \mathbb{R}^n \text{ such that } x \in S_i$$

entails $c_{\ell} = 0$. A simple example will show how natural this condition is.

Example : Suppose that $D = \{1,2\}$, in other words, we only have two R.F.'s at hand. Take as functions $f(x,i)$ $k_1 + k_2$ functions such that

$$(4\text{-}5\text{-}3) \qquad \begin{cases} f^1(x,2) = \ldots = f^{k_1}(x,2) = 0 \\ \\ f^{k_1+1}(x,1) = \ldots = f^{k_1+k_2}(x,1) = 0 \end{cases}$$

These relations describe a situation where the two drifts $m_1(x)$ and $m_2(x)$ are algebraically independent. The k_1 first functions, null on V_2, serve only to represent on V_1 the drift $m_1(x)$ of the first R.F., and the k_2 other functions, null on V_1, represent on V_2 the drift $m_2(x)$ of the second R.F.

Let us suppose that S_2 is void, so that we have no experimental information about $Z_2(x)$. As the drifts $m_1(x)$ and $m_2(x)$ are, by hypothesis, algebraically independent, it is clear that the knowledge of $Z_1(x)$ on S_1 will bring no further information about the drift $m_2(x)$ of the second R.F. In fact, the functions $f^{\ell}(x,i)$ are not linearly independent on S. Indeed, as S_2 is void, S is reduced to

$\{S_1\} \times \{1\}$, that is to the set (x,i) of the couples such that $i = 1$ and $x \in S_1$. But, precisely, we have $f^{\ell}(x,1)$ identically equal to zero for $\ell > k_1$, and consequently these functions are not linearly independent on S_1 .

Let us pass on now to the systems themselves. It will be noted carefully that the notation

$$\forall\ (y,j) \in S$$

means precisely : <u>for any index $j \in D$ and for any point $y \in \mathbb{R}^n$ such that $y \in S_j$</u> .

<u>4-5-2 Optimal estimation of the drift.</u>

Let us consider the optimal estimator $m^*(x_o,i_o)$ of the drift of the R.F. $Z_{i_o}(x)$ at the point x_o , that is, in product space notation, the optimal estimator of the drift $m(x_o,i_o)$ at the point $(x_o,i_o) \in \mathbb{R}^n \times D$. This estimator must be linear with respect to the $Z(x,i)$, $(x,i) \in S$. Hence it will have the form

$$m^*(x_o,i_o) = \sum_i \int_{S_i} \lambda^i(dx)\ Z_i(x)$$

with d measures λ^i with supports respectively contained in the S_i . The transposition of system (4-2-3) gives then :

$$(4\text{-}5\text{-}4) \qquad \begin{cases} \sum_i \int_{S_i} \lambda^i(dx)\ \sigma_{ij}(x,y) = \mu_\ell\ f^\ell(y,j) & \forall\ (y,j) \in S \\ \sum_i \int_{S_i} \lambda^i(dx)\ f^\ell(x,i) = f^\ell(x_o,i_o) \\ D^2[m^*(x_o,i)] = \mu_\ell\ f^\ell(x_o,i_o) \end{cases}$$

In the same way, by delinearizing system (4-5-4), we get the optimal estimators A_ℓ of the coefficients a_ℓ of the drift by using a set of measures $\lambda^i_\ell(dx)$ with supports respectively in S_i , i.e. :

$$A_\ell = \sum_i \int_{S_i} \lambda^i_\ell(dx)\ Z_i(x)$$

by solving the system obtained by transposing (4-2-7), i.e. :

$$(4\text{-}5\text{-}5)\qquad \begin{cases} \sum_i \int_{S_i} \lambda_\ell^i(dx)\ \sigma_{ij}(x,y) = \mu_{\ell s}\ f^s(y,j) & \forall\ (y,j) \in S \\ \sum_i \int_{S_i} \lambda_\ell^i(dx)\ f^s(x,i) = \delta_\ell^s \\ \mu_{\ell s} = \mathrm{Cov}\ (A_\ell\ ,\ A_s) \end{cases}$$

4-5-3 Punctual Cokriging.

With complete generality, we could form the equations relative to the estimation of a quantity of the form :

$$Z_o = \sum_i \int p^i(dx)\ Z_i(x)$$

with any measures p^i (with supports respectively contained in the V_i). But we have already noticed that the right hand sides of our systems depend linearly on these measures $p^i(dx)$. It is sufficient then to know how to form the estimator $Z^*(x,i)$ of the kriging relative to the point $(x,i) \in \mathbb{R}^n \times D$ to be able to deduce from it the estimator Z_o^* of the kriging of Z_o according to the relationship :

$$Z_o^* = \sum_i \int p^i(dx)\ Z^*(x,i)$$

Thus we will limit ourselves to writing the system for punctual kriging.

$Z_{i_o}^*(x_o)$ is to be estimated, i.e. the value of the R.F. at the point (x_o,i_o) of the product space. The transposition of system (4-3-1) shows then that the optimal universal estimator is of the form :

$$Z_{i_o}^*(x_o) = \sum_i \int_{S_i} \lambda^i(dx)\ Z_i(x)$$

with measures λ^i with support in S_i verifying :

$$(4\text{-}5\text{-}6)\qquad \begin{cases} \sum_i \int_{S_i} \lambda^i(dx)\ \sigma_{ij}(x,y) = C_{ji_o}(y,x_o) + \mu_\ell\ f^\ell(y,j) \quad \forall\ (y,j) \in S \\ \sum_i \int_{S_i} \lambda^i(dx)\ f^\ell(x,i) = f^\ell(x_o,i_o) \\ \sigma_v^2 = \sigma_{i_o i_o}(x_o,x_o) + \mu_\ell\ f^\ell(x_o,i_o) - \sum_i \int_{S_i} \lambda^i(dx)\ \sigma_{ii_o}(x,x_o) \end{cases}$$

It will be noted that the additivity theorem remains unchanged in the case of cokriging, but there is no point in giving the corresponding relationships explicitly here.

4-6 THE INDETERMINABILITY OF THE UNDERLYING VARIOGRAM.

4-6-1 Statement of the problem.

We are now going to deal with the difficult (but obviously essential for the applications) problem of the estimation of the variogram $\gamma(x,y)$ of a non-stationary R.F. $Z(x)$ having a drift $m(x)$. This variogram is the variogram of the true residuals $Z(x) - m(x)$, and the presence of the drift $m(x)$ does not allow it to be estimated directly from the experimental variogram or from the variogram of the estimated residuals $Z(x) - m^*(x)$ (see paragraph 4-2-4). To express the fact that this "true" variogram is hidden (in a partly irremediable manner, as it will be seen), we shall say that it is the underlying variogram, and the fundamental problem which is studied here is that of the identification of the underlying variogram from a unique realization. We cannot deal here with the delicate problems of mathematical statistics concerning the possibility of such an operation. More particularly, it would be proper to examine precisely the type of quasi-stationarity assumption which it should be necessary to make about the true residuals $Z(x) - m(x)$, so as to be able to reconstitute effectively what is theoretically possible to estimate about this underlying variogram.

Indeed, the fact is that we shall see that the problem is theoretically indetermined : at the best, that is assuming that we know the expectation of the estimators that we have a right to use (the universal estimators), and not only the corresponding experimental values computed on several parts of a single realization, it appears that the reconstitution of the underlying variogram remains fundamentally partially indeterminate, this being of an algebraic nature, and it is essentially to

the algebraic study of this indeterminability that we will devote ourselves in this last paragraph.

Notice that this is an absolutely general problem, which is encountered in the most common applications of classical statistics (see Ex. 16). Statistics avoid this problem of fundamental indeterminability by introducing supplementary assumptions which are in general very strong (for example, the independence of different r.v.'s of which we know the realization). Here also, it would be useful to examine to what extent these too strong conditions can be weakened without rendering the problem unsolvable. For physical reasons, we may think that an assumption of quasi-stationarity should be fitting. We shall only verify that indeed it is possible to considerably reduce this indeterminability by choosing from all the variograms compatible with the data, the one which is "the most stationary possible" (in the sense of the least squares, for example). (cf. Ex. 17).

On the other hand, a remarkable (and quite reassuring on methodological grounds) circumstance will appear : this fundamental indeterminability will only affect the variances of our optimal estimators and not the estimators themselves. In other words, all the theoretically possible underlying variograms lead to the same optimal estimators for the drift and for kriging, and these estimators depend in no way on the supplementary physical assumptions we may introduce to remove the indeterminability (on the contrary, the variance to be ascribed to these estimators will depend on them).

This result will be compared with the similar conclusion concerning the constant term a_0 of the drift, a conclusion to which the interpretation of the drift as a very regular R.F. has allowed us to come in paragraph 4-4-4. Here also, the theory of random drift allows us to foresee from the beginning the type of indeterminability we shall meet with. We shall examine this now, before passing on, in the following paragraphs, to the proper algebraic study of the problem.

Let $Z(x)$ be a R.F. having a functional drift of the type $a_\ell f^\ell(x)$ and a covariance $\sigma(x,y)$, and let also $m(x)$ be a random function, very regular and such that its covariance $K(x,y)$ has, on the working domain V, a good approximate representation of the form

$$K(x,y) = K\rho_s f^\ell(x) f^s(y)$$

The random function

$$Z'(x) = Z(x) + m(x)$$

has the underlying covariance

$$\sigma'(x,y) = \sigma(x,y) + K_{\ell s}\, f^{\ell}(x)\, f^{s}(y) \qquad (4\text{-}6\text{-}1)$$

(Assuming $m(x)$ and $Y(x)$ to be independent) and the drift

$$a_\ell f^\ell(x) + m(x) \simeq (a_\ell + b_\ell)\, f^\ell(x)$$

with a_ℓ definite but unknown, whilst the b_ℓ are r.v.'s (the covariance matrix of which is exactly $K_{\ell s}$).

But, when we have only a single realization of the random function at our disposal, only the corresponding realizations of the r.v.'s b_ℓ , which are simple numbers, have an experimental significance. But, from the epistemological point of view, it would be illusory to think that we really gain information when we decide to interpret a given numerical value as a realization of a r.v. whose distribution is unknown. As the $K_{\ell s}$ are experimentally inaccessible (and we have seen that it is the general case) the two parameters a_ℓ and b_ℓ are not really distinguishable. We may as well decide to interpret the b_ℓ as numbers : instead of $\sigma'(x,y)$, our R.F. $Z'(x)$ has then the same covariance $\sigma(x,y)$ as $Z(x)$. Hence there is nothing in a single realization of $Z'(x)$ which allows us to choose between these two covariances. The algebraic indeterminability, concerning the reconstitution of the underlying covariance, will hence be at least (and we shall see that it is in fact exactly) that expressed by relationship (4-6-1).

But on the other hand - reassuring circumstance - the random drift theory stated in paragraph 4-4 shows that the equations giving the optimal estimators are the same for the random drifts and the functional drifts. Consequently, we should expect - and this will be verified - that this fundamental indeterminability (4-6-1) does not have any repercussion on the optimal estimators themselves (but only on the variances to be given to them).

Let us pass on now to the algebraic study of the problem, limiting ourselves to the case where a covariance $\sigma(x,y)$ exists, and where the experimental data set S is finite.

4-6-2 Universal quadratic estimators.

Within a theory of order two (i.e. : where we try to estimate the covariance, without knowing anything a priori about the law of the R.F.), the only estimators we can use to estimate this covariance are quadratic estimators of the type

$$Q^* = Q^{\alpha\beta}\, Z_\alpha\, Z_\beta$$

A quadratic estimator Q^* will be said to be universal if its mathematical expectation does not depend on the coefficients a_ℓ of the drift $m(x) = a_\ell\, f^\ell(x)$. From :

$$E(Q^*) = Q^{\alpha\beta}\, \sigma_{\alpha\beta} + a_\ell\, a_s\, Q^{\alpha\beta}\, f^\ell_\alpha\, f^s_\beta$$

it follows that a quadratic estimator Q^* is universal if and only if it obeys the relationship

(4-6-2) $$Q^{\alpha\beta}\, f^\ell_\alpha\, f^s_\beta = 0$$

When the coefficients a_ℓ of the drift are unknown, these universal estimators are the only utilizable ones for the statistical inference of the covariance function. It follows from this remark that the maximum of information which it is theoretically possible to obtain concerning the matrix $\sigma_{\alpha\beta}$ of the underlying covariances is contained in the set of the numerical values $E(Q^*)$, Q^* describing the class of the universal quadratic estimators, thus in the set of the $Q^{\alpha\beta}\, \sigma_{\alpha\beta}$ for the matrices $Q^{\alpha\beta}$ verifying (4-6-2).

To characterize the class of the universal quadratic estimators, we need some preliminary algebraic considerations. Let N be the number of experimental points $x_\alpha \in S$, k be that of the functions f^ℓ ($\ell = 1,2,\ldots k$) assumed to be linearly independent on S. We can find N-k other functions f^u ($u = k+1,\ldots N$) such that the N vectors $f^i = (f^i_\alpha)$, $i = 1,2,\ldots N$ constitute a basis for the space $\mathbb{R}^n$, and we may associate with this basis the dual basis formed by N co-vectors $\lambda_i = (\lambda^\alpha_i)$ verifying the conditions

(4-6-3) $$\lambda^\alpha_i\, f^j_\alpha = \delta^j_i \quad , \quad \lambda^\alpha_i\, f^i_\beta = \delta^\alpha_\beta$$

(in other words, the matrices λ^α_i and f^i_α are inverse). Among these co-vectors λ_i, $i = 1,2,\ldots N$, the k first ones, that is λ_ℓ , $\ell = 1,2,\ldots k$ verify the universality condition :

(4-6-4) $$\lambda_\ell^\alpha \, f_\alpha^s = \delta_\ell^s$$

In other words, for any matrix $\sigma_{\alpha\beta}$, the estimator

(4-6-4') $$A_\ell = \lambda_\ell^\alpha \, Z_\alpha$$

is such that $E(A_\ell) = a_\ell$ whatever the unknown a_ℓ are : it is a <u>universal estimator</u> (but <u>not optimal in general</u>). We shall write :

$$m_\alpha^* = A_\ell \, f_\alpha^\ell$$

and m_α^* will be a universal estimator (non optimal) of the value of the drift at x_α . The N-k other vectors λ_i, i.e. λ_u , u = k+1,...N constitute a <u>basis</u> for the orthogonal of the f^ℓ . They verify :

$$\lambda_u^\alpha \, f_\alpha^\ell = 0$$

Thus the subspace generated by the λ_u is independent of the choice of the functions f^u , since it is identical to the orthogonal of the f^ℓ . Let us put :

$$A_u = \lambda_u^\alpha \, Z_\alpha$$

These random variables verify $E(A_u) = 0$ whatever the coefficients a_ℓ are. As the matrix λ_i^α is regular, we can change the variables, and replace the Z_α by :

$$A_i = \lambda_i^\alpha \, Z_\alpha \qquad (i = 1,2,...N)$$

Any quadratic estimator Q^* is then of the form :

$$Q^* = Q^{ij} \, A_i \, A_j$$

We shall compute $E(Q^*)$. Since $E(A_u) = 0$ (u = k+1,...N) whatever the a_ℓ are, we have :

$$E(Q^*) = Q^{ij} \, \mathrm{Cov}(A_i \, A_j) + Q^{\ell s} \, a_\ell \, a_s$$

Consequently, Q^* is a quadratic universal estimator if and only if we have $Q^{\ell s} = 0$ ($\ell, s = 1,2,...k$).

Thus, all the information that it is theoretically possible to reconstitute concerning the covariances matrix is contained in the following matrix (N-k) × N :

(4-6-10) $$S_{ui} = \mathrm{Cov}\ (A_u\ A_i) \qquad (u = k+1,\ldots N\ ;\ i = 1,2,\ldots N)$$

4-6-3 General form of the admissible covariance matrices.

Conversely, suppose that we know all the information which it is theoretically possible to reconstitute, i.e. the N(N-k) numbers S_{ui} of relationship (4-6-10). The matrix :

$$S_{ij} = \mathrm{Cov}\ (A_i\ A_j) = \lambda_i^\alpha\ \lambda_j^\beta\ \sigma_{\alpha\beta}$$

is reversible, and theoretically we have :

(4-6-11) $$\sigma_{\alpha\beta} = S_{ij}\ f_\alpha^i\ f_\beta^j$$

In this equation, the $S_{\ell s} = \mathrm{Cov}\ (A_\ell\ A_s)$ $(\ell, s = 1,2,\ldots k)$ represent the inaccessible part of the covariance. Hence we can write :

(4-6-12) $$\sigma_{\alpha\beta} = C_{\alpha\beta} + S_{\ell s}\ f_\alpha^\ell\ f_\beta^s$$

The first term represents the cognizable part of the covariance matrix $\sigma_{\alpha\beta}$, while the undetermined part is represented by the quadratic form $S_{\ell s}\ f_\alpha^\ell\ f_\beta^s$. However, we have :

$$S_{\ell s}\ f_\alpha^\ell\ f_\beta^s = \lambda_\ell^{\alpha'}\ \lambda_s^{\beta'}\ f_\alpha^\ell\ f_\beta^s\ \mathrm{Cov}\ (Z_{\alpha'},\ Z_{\beta'}) = E(m_\alpha^*\ m_\beta^*) - m_\alpha\ m_\beta$$

and hence, from (4-6-12) :

$$C_{\alpha\beta} = \sigma_{\alpha\beta} + m_\alpha\ m_\beta - E(m_\alpha^*\ m_\beta^*)$$

i.e. :

(4-6-13) $$C_{\alpha\beta} = E(Z_\alpha\ Z_\beta - m_\alpha^*\ m_\beta^*)$$

Such is the cognizable part of the covariance. Any matrix of the form :

$$\tau_{\alpha\beta} = C_{\alpha\beta} + D_{\ell s}\, f_\alpha^\ell\, f_\beta^s \tag{4-6-14}$$

where $D_{\ell s}$ is an arbitrary symmetric matrix is quite as admissible (provided that it is positive) as the "true" covariance $\sigma_{\alpha\beta}$. These admissible covariances are deduced from each other according to the following relationship, where $T_{\ell s}$ is an arbitrary symmetric matrix :

$$\tau_{\alpha\beta} = \sigma_{\alpha\beta} + T_{\ell s}\, f_\alpha^\ell\, f_\beta^s \tag{4-6-15}$$

4-6-4 Consequences for the optimal estimators.

Consider for example, the system of equations of the optimal estimation of the drift, taking as a covariance $\tau_{\alpha\beta}$ the general solution (4-6-15) :

$$\begin{cases} \tilde{\lambda}_\ell^\beta\, \tau_{\alpha\beta} = \tilde{\mu}_{\ell s}\, f_\alpha^s \\ \tilde{\lambda}_\ell^\beta\, f_\beta^s = \delta_\ell^s \end{cases}$$

The universality condition does not make $\tau_{\alpha\beta}$ intervene, hence it does not depend on the indeterminate matrix. the first relationship can be put in the following equivalent form :

$$\tilde{\lambda}_\ell^\beta\, \sigma_{\alpha\beta} = \tilde{\mu}_{\ell s}\, f_\alpha^s - T_{\ell' s'}\, f_\alpha^{\ell'}\, f_\beta^{s'}\, \tilde{\lambda}_\ell^\beta$$

i.e., taking the universality condition into account :

$$\begin{cases} \tilde{\lambda}_\ell^\beta\, \sigma_{\alpha\beta} = (\tilde{\mu}_{\ell s} - T_{\ell s})\, f_\alpha^s \\ \tilde{\lambda}_\ell^\beta\, f_\beta^s = \delta_\ell^s \end{cases}$$

But this system is identical to the system (4-2-6) written with the "true" unknown underlying covariance, on the condition that we take :

(4-6-20) $$\mu\ell_s = \tilde{\mu}\ell_s - T\ell_s$$

Consequently, the indeterminability of the underlying covariance has no repercussion on the optimal estimation of the drift. But, from (4-6-20), it affects the covariance matrix of the optimal estimators, which remains theoretically wholly indeterminate.

When there is no covariance $\sigma_{\alpha\beta}$ but only a variogram, the expectation $E(Z_\alpha Z_\beta - m^*_\alpha m^*_\beta)$ does not exist in general. The expression (4-6-17) will have to be replaced by a similar expression constructed from the increments of the R.F. and of its estimated drift, i.e. :

$$\gamma_{\alpha\beta} = \frac{1}{2} E[(Z_\alpha - Z_\beta)^2 - (m^*_\alpha - m^*_\beta)^2] + \frac{1}{2} D^2 (m^*_\alpha - m^*_\beta)$$

where the first term is the accessible part of the underlying variogram, and the second one the inaccessible part.

In Exercise 17, some indications will be found on how a physically plausible hypothesis like that of local stationarity allows the difficulties associated with indeterminability to be removed for the most part. (It will be noticed that the term $\sigma(o) = \sigma^2$ representing the a priori variance remains, in principle, inaccessible).

4-7 EXERCISES ON CHAPTER 4.

4-7-1 Exercises on drifts.

Exercise 1 (Linear variogram, continuous case)

a/ Let, on a straight line, $f(x)$ be a function having a continuous second derivative. Show (by integration by parts) that we have for $-R \leq y \leq R$:

$$\int_{-R}^{R} |x-y|\, f''(x)\, dx = 2 f(y) - f(R) - f(-R) + R[f'(R) - f'(-R)] - y[f'(R) + f'(-R)]$$

δ_R and δ_{-R} being the Dirac measures set at R and -R respectively, show that, still for

$-R \leq y \leq R$

$$\int \frac{\delta_R+\delta_{-R}}{2} |x-y| = R \ , \quad \int \frac{\delta_R-\delta_{-R}}{2} |x-y| = -y \ , \quad \int_{-R}^{R} |x-y| \, dx = y^2 + R^2$$

b/ On a straight line, consider a R.F. having a variogram $\gamma(x,y) = |x-y|$ and a <u>quadratic drift</u> $a_1 x + a_2 x^2$ (defined except for a constant a_0) with unknown coefficients. The realization is supposed to be known on $(-R, R)$. Show that the solution of system (4-2-11) completed for the value 0 of the index ℓ leads to the estimators :

$$A_0 = \frac{3}{2} \overline{Z} - \frac{1}{4} (Z_R + Z_{-R}) \ , \ A_1 = \frac{1}{2R} (Z_R - Z_{-R}) \ , \ A_2 = \frac{3}{4R^2} (Z_R + Z_{-R}) - \frac{3}{2R^2} \overline{Z}$$

$$(\overline{Z} = \frac{1}{2R} \int_{-R}^{R} Z(x) \, dx) \text{ with } D^2(A_1) = \frac{1}{R} \ , \ D^2(A_2) = \frac{3}{4R^3} \ , \ \text{Cov}(A_1 A_2) = 0$$

[Notice, from a/, that $A_\ell = \int_{-R}^{R} \lambda_\ell(dx) \, Z(x)$ with a measure λ_ℓ which is a linear combination of the measures $\frac{\delta_R+\delta_{-R}}{2}$, $\frac{\delta_R-\delta_{-R}}{2}$ and dx, and use the universality conditions to determine the coefficients of these combinations. Identify the Lagrange parameters in the first relationship (4-2-11) to determine the variances.].

c/ (<u>Variogram of the residuals</u>) - Show that $\gamma^*(h) = \frac{1}{L-h} \int_{-R}^{R-h} \frac{1}{2} [Z(x+h) - m^*(x+h) - Z(x) + m^*(x)]^2$ has, in the above conditions, the expectation

$$E[\gamma^*(h)] = h - \frac{h^2}{R} + \frac{1}{2} \frac{h^3}{R^2} - \frac{1}{8} \frac{h^4}{R^3}$$

[Use the relationship (4-2-15). The graph is a rather flattened curve. The slope at the origin is unchanged, but as soon as h is not very small, the bias is huge. For $h = 2R$, this expectation vanishes. Notice that the experimenter risks concluding that the residuals are independent].

<u>Exercise 2</u> - In the same conditions as above, assume that the R.F. has only a linear drift. Show that the estimator $A_0 + A_1 x$ is the straight line joining the twp points Z_R and Z_{-R} , and that the variogram of the residuals has the expectation

$$E[\gamma^*(h)] = h - \frac{h^2}{2R}$$

Exercise 3 - (Influence of a nugget effect) - On a straight line, consider an I.R.F. with a linear drift and the variogram $\gamma(h) = |h| - C\,\delta$ (nugget effect represented by a Dirac measure). Show that the solution of system (4-2-11) completed for $\ell = 0$ leads to the estimators

$$A_0 = b_0 \int_{-R}^{R} \text{chax } Z(x)\,dx \quad , \quad A_1 = b_1 \int_{-R}^{R} \text{shax } Z(x)\,dx$$

$$\left(b_0 = \frac{a}{2\ \text{sha R}} \quad , \quad b_1 = \frac{a^2}{2(a\ R\ \text{cha R} - \text{sha R})}\right) \quad D^2(A_1) = 2\ b_1\ \text{cha R} \qquad a = \sqrt{\frac{2}{C}}$$

[Use Ex. 7, Chapter 3 to show that the measures λ_0 and λ_1 are of the type b_0 chax dx and b_1 shax dx, identify b_0 and b_1 with the help of the universality conditions, and μ_{11} with the first relationship (4-2-11)].

Exercise 4 - (Linear variogram, discrete case) - Let Z(x) be the same R.F. as in Exercise 1 (quadratic drift and linear variogram), but its realization is known at the n+1 points $x_i = i\,h$, $i = 0,1,\ldots n$. We shall put $L = nh$.

a/ Let ν_1 be the measure $\delta_L - \delta_0$ and ν_2 the measure $\sum_{i=0}^{n} \delta_{ih} - \frac{n+1}{2}(\delta_L + \delta_0)$. Show that these measures verify for $j = 0,1,\ldots n$:

$$\int \nu_1(dx)\ |x-jh| = (n-2j)\ h \ , \quad \int \nu_2(dx)\ |x-jh| = -\ j(n-j)\ h$$

b/ Deduce the optimal estimators of a_1 and a_2 :

$$A_1 = \frac{Z_n - Z_0}{Ah} + \frac{6}{(n-1)h}\left(\overline{Z} - \frac{Z_0+Z_n}{2}\right)$$

$$A_2 = -\frac{6}{h^2\,(n(n-1)}\left(\overline{Z} - \frac{Z_0+Z_n}{2}\right) \qquad \left(\overline{Z} = \frac{1}{n+1}\ \Sigma\ Z_i\right)$$

and the covariance matrix :

$$\mu_{11} = \frac{8n^2-2}{n(n^2-1)}\ \frac{1}{h} \ , \quad \mu_{12} = \mu_{21} = -\ \frac{6}{n^2-1}\ \frac{1}{h^2} \ , \quad \mu_{22} = \frac{6}{n(n^2-1)}\ \frac{1}{h^3}$$

(deduce from a/ that the measures λ_1 and λ_2 are linear combinations of ν_1 and ν_2 ; determine the coefficients using the universality conditions ; identify the $\mu\rho_s$ on the right hand side of the optimality equation).

c/ Compare with Exercise 1,b/. Hence determine the estimators A_0' , A_1' and A_2' of the continuous case when the realization is known on (0,L).

[Put L = 2R , place the origin of coordinates at -R and use tensorial invariance, hence $A_0' = A_0 - A_1 R + A_2 R^2$; $A_1' = A_1 - 2 A_2$, $A_2' = A_2$].

Exercise 5 (Exponential covariance, continuous case) -

a/ Show that, on a straight line, the measures :

$$\nu_0(dx) = \frac{a}{2} dx + \frac{1}{2} (\delta_R + \delta_{-R}) \; ; \; \nu_1(dx) = \frac{a}{2} x \, dx + \frac{1+aR}{2a} (\delta_R - \delta_{-R})$$

verify the integral equations : $(-R \leq y \leq R)$

$$\int_{-R}^{R} \nu_0(dx) \, e^{-a|x-y|} = 1 \; , \quad \int_{-R}^{R} \nu_1(dx) \; e^{-a|x-y|} = y$$

b/ On a straight line, consider a R.F. having a linear drift and the covariance $e^{-a|x-y|}$. Its realization is known on the interval (-R, R). Show that the optimal estimators verify :

$$A_0 = \frac{aR}{1+aR} \, \overline{Z} + \frac{1}{1+aR} \; \frac{Z_R + Z_{-R}}{2}$$

$$A_1 = \frac{a}{R(1+aR + \frac{1}{3} a^2 R^2)} \left[\frac{a}{2} \int_{-R}^{R} x \, Z(x) \, dx + \frac{1+aR}{2a} \; (Z_R - Z_{-R}) \right]$$

$$\mu_{00} = \frac{1}{1+aR} \; , \quad \mu_{10} = \mu_{01} = 0 \; , \quad \mu_{11} = \frac{a}{R(1+aR + \frac{1}{3} a^2 R^2)}$$

Exercise 6 (Case of large grids) -

a/ The notations are those of Ex. 10 of Chapter 3 and of Ex. 13 to 15 of Chapter 2. In $\mathbb{R}^2$, consider a R.F. having a covariance K(x-y) with a small range, and N experimental points x_α , $\alpha = 1,2,\ldots N$ such that their mutual distances apart are greater than the range. We shall suppose that a drift exists, having the form $a_\ell \, f^\ell(x)$, $\ell = 1,\ldots k$. Show that the optimal estimators A_ℓ are identical to those given by the least squares method. Compute their expression and the covariance matrix μ_{ℓ_s} as a function of the matrix

$$T^{\ell s} = \sum_{\alpha=1}^{n} f_\alpha^\ell f_\alpha^s \text{ and of its inverse } S_{\ell s}$$

$(\mu_{\ell s} = C\, S_{\ell s}\ ,\ \lambda_\ell^\alpha = S_{\ell s}\, f_\alpha^s)$

b/ Compute the mathematical expectation of the expression

$$I = \frac{1}{n} \sum_\alpha (Z_\alpha - A_\ell\, f_\alpha^\ell)^2$$

[Make use of the relationship between the cross term and the quadratic term at the optimum. The result is $I = (1 - \frac{k}{N})\, C$].

Exercise 7 (Linear drift on the circumference of a circle). -

a/ In $\mathbb{R}^2$, consider a R.F. with a covariance $C(r)$ which depends only on $r = |x-y|$. The realization is supposed to be known on the circumference of a circle with a centre O and radius R. Show that if there is a quadratic drift, the system giving the optimal estimators is not regular (the f^ℓ are not linearly independent on S, because $x^2 + y^2 = R^2$).

b/ We suppose that the preceding R.F. has a linear drift $a_o + a_1 x + a_2 y$. Form the optimal estimators A_0 , A_1 and A_2 , show that they do not depend on the covariance $C(r)$ and that they coïncide with those of the least squares methods.

[on the circumference, the drift is $a_o + a_1 R \cos\theta + a_2 R \sin\theta$; the covariance develops as a Fourier series $C(2R\,|\sin\frac{\theta-\theta'}{2}|) = \Sigma\, b_p \cos(\theta-\theta')$ with all the b_p different from 0 if $C(r)$ is strictly positive definite. The functions 1, $\sin\theta$ and $\cos\theta$ are then eigen functions for the kernel $C(\theta-\theta')$. Deduce that

$$A_o = \frac{1}{2\pi} \int_0^{2\pi} Z(\theta)\, d\theta\ ,\ A_1 = \frac{1}{\pi R} \int_0^{2\pi} Z(\theta) \cos\theta\, d\theta\ ,\ A_2 = \frac{1}{\pi R} \int_0^{2\pi} Z(\theta) \sin\theta\, d\theta$$

Only the covariance matrix depends on $C(r) = \mu_{oo} = b_o$, $\mu_{11} = \mu_{22} = \frac{b_1}{R^2}$, cross terms are equal to 0].

4-7-2 Kriging and Cokriging.

Exercise 8 - In the conditions of Ex. 1, krige the point $x_o = R + h$ (apply the additivity theorem and the Ex. 6, a/ of Chapter 3 ; the estimator is $Z_U(x_o) = Z_R + A_1\, h + A_2\, h\,(2R + h)$ with the variance

$$4\frac{h^2}{R} + 3\frac{h^3}{R^2} + 3\frac{h^4}{4R^3} + 2h$$

which increases rapidly with h : extrapolation is only possible over small distances).

Exercise 9 - In the conditions of Ex. 4, krige the point $x_o = i\,h + \varepsilon\,h$ $(0 \leq \varepsilon \leq 1)$; $0 < i < n$. (Apply the additivity theorem, and Exercise 6, b/ of Chapter 3. We get

$$Z^*(x_o) = \varepsilon\, Z_{i+1} + (1-\varepsilon)Z_i - A_2\, h^2\, \varepsilon(1-\varepsilon)\ ,\ \sigma_U^2 = 2\,\varepsilon(1-\varepsilon)h + \frac{6\,\varepsilon^2(1-\varepsilon)^2}{n(n^2-1)}\,h\ ).$$

Exercise 10 - In the conditions of Ex. 5, krige $x_o = R + h$. (Apply the additivity theorem and Ex. 4 of Chapter 3).

Exercise 11 - (Case of large grids) - In the conditions of Ex. 6, we want to estimate a domain V large with respect to the range. We suppose that n of the $N = n + n'$ points x_α are inside V and n' outside, and that they are all located at distances greater than the range from the boundary of V. Form the optimal estimator of V in the presence of a drift $a_\ell\, f^\ell(x)$ and compute the associated variance.

(Apply the additivity theorem and Ex. 10 of Chapter 3. To avoid confusion in notation, B will be the term A_1 of that exercise. Putting

$$\lambda = \frac{1}{n+n'} + \frac{n'}{n+n'}\,\pi\,\frac{B}{CV}\ ,\quad \lambda' = \frac{1}{n+n'} - \frac{n}{n+n'}\,\pi\,\frac{B}{CV}$$

i and j' respectively denoting the inside and outside samples, the estimator is :

$$Z_U = \lambda \sum_{i=1}^{n} (Z_i - A_\ell\, f_i^\ell) + \lambda' \sum_{j'=1}^{n'} (Z_{j'} - f_{j'}^\ell\, A_\ell)$$

with the variance

$$\sigma_K^2 + \mu_{\ell s}\, H^{\ell s} \ , \ H^{\ell s} = (\lambda \sum_i f_i^{\ell} + \lambda' \sum_{j'} f_{j'}^{\ell}) \ (\lambda \sum_i f_i^{s} + \lambda' \sum_{j'} f_{j'}^{s})$$

Exercise 12 (multinomial interpolator) - Consider n points $x_1, x_2, \ldots x_n$ on a straight line. The polynomial of degree n-1 such that $P(x_i) = Z_i$ is :

$$P(x) = \sum_i \frac{(x-x_1)(x-x_2)\ldots(x-x_{i-1})(x-x_{i+1})\ldots(x-x_n)}{(x_i-x_1)(x_i-x_2)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\ldots(x_i-x_n)} \ Z_i$$

Form the expression for the cubic interpolator associated with the four points

$x_1 = -\frac{3}{2}\,a$, $x_2 = -\frac{a}{2}$, $x_3 = \frac{a}{2}$, $x_4 = 3\,\frac{a}{2}$, i.e., $b_o + b_1 x + b_2 x^2 + b_3 x^3 = P(x)$

Integrate from $-\frac{a}{2}$ to $\frac{a}{2}$. Show that

$$\frac{1}{a} \int_{-\frac{a}{2}}^{a/2} P(x)\, dx = b_o + \frac{1}{12}\ b_2\, a^2 = \frac{13}{12}\ \frac{Z_2+Z_3}{2} - \frac{1}{12}\ \frac{Z_1+Z_4}{2}$$

Compare with Exercise 1 of chapter 3 for the negative weight of the external data.

Exercise 13 (Cokriging in the presence of a non-systematical error)- Let $Z_o(x)$ be a R.F., $C(x,y)$ its covariance, $a_\ell\, f^\ell(x)$ its drift with unknown coefficients a_ℓ . Let $\varepsilon(x)$ be another R.F. with zero expectation and a covariance $C_\varepsilon(x,y)$: $\varepsilon(x)$ is independent of Z_o . Let $Z_1(x) = Z_o(x) + \varepsilon(x)$.

a/ Show that $C_{oo} = C_{01} = C_{10} = C$, and $C_{11} = C + C_\varepsilon$

b/ The realization of Z_i is supposed to be known on S_1 (S_o is void). Show that the optimal estimators $A_\ell = \int_{S_1} \lambda_\ell(dx)\ Z_1(x)$ are solution of the system :

$$\begin{cases} \int_{S_1} \lambda_\ell(dx)\ [C(x,y) + C_\varepsilon(x,y)] = \mu_{\ell s}\ f^s(y) & (y \in S_1) \\ \int_{S_1} \lambda_\ell(dx)\ f^s(x) = \delta_\ell^s \end{cases}$$

(the drift of Z_1 would be estimated in exactly the same way).

c/ Krige $Z_0(x_0)$ at a point x_0 (belonging or not to S_1). Show that $Z_0^*(x_0) = \int_{S_1} \lambda(dx) Z_1(x)$ with

$$\int_{S_1} \lambda(dx)\ [C(c,y) + C_\varepsilon(x,y)] = C(x_0 y) + \mu_\ell\, f^\ell(y) \qquad (y \in S_1)$$

$$\int_{S_1} \lambda(dx)\ f^\ell(x) = f^\ell(x_0)$$

(This system is not identical to that obtained by kriging $Z_1(x_0)$; in particular, $Z_0^*(x_0)$ is not an exact interpolator for $Z_1(x)$).

In the same way, write the kriging system of $Z_1(x_0)$ ($x_0 \notin S_1$) and compare the two variances using the hypothesis that C_ε is a nugget effect (the estimation of $Z_0(x_0)$ is better than that of $Z_1(x_0)$).

Exercise 14 (Cokriging in the presence of a systematical error) - Let $Z_0(x)$ be a stationary R.F., $C(h)$ its (known) covariance, m_0 its unknown expectation. Let $\varepsilon(x)$ be another stationary R.F., $C_\varepsilon(h)$ its covariance and ε_0 its (unknown) expectation. $\varepsilon(x)$ is independent of $Z_0(x)$. $Z_1(x) = Z_0(x) + \varepsilon(x)$.

a/ Show that $C_{00} = C_{01} = C$; $C_{11} = C + C_\varepsilon$. To represent the drifts, introduce the two functions $f(x,i)$ defined by

$$f^0(x,0) = 1\ ,\ f^0(x,1) = 0\ ,\ f^1(x,0) = 0\ ,\ f^1(x,1) = 1$$

The drift $m(x,i)$ of $Z(x,i)$ is $a_0 f^0(x,i) + a_1 f^1(x,i)$, with $a_0 = m_0$, $a_1 = m_1 = m_0 + \varepsilon_0$. Show that the linear independence condition is realized only if the two sets S_0 and S_1 are not void (if S_0 is void, there is no hope of separating, in m_1, the systematic error ε_0 from the true expectation m_0).

b/ Form the equations of the optimal estimators $A_0 = \int_{S_0} \lambda_0^0(dx)\ Z_0(x) + \int_{S_1} \lambda_0^1(dx)\ Z_1(x)$ and $A_1 = \int_{S_0} \lambda_1^0(dx)\ Z_0(x) + \int_{S_1} \lambda_1^1(dx)\ Z_1(dx)$.

The universality conditions are $\int_{S_i} \lambda_j^i(dx) = \delta_j^i$, those of optimality :

$$\int_{S_o} \lambda_\ell^o \, C(x,y) + \int_{S_1} \lambda_\ell^1 \, C(x,y) = \mu \ell_o \qquad (y \in S_o)$$

$$\int_{S_o} \lambda_\ell^o \, C(x,y) + \int_{S_1} \lambda_\ell^1 \, [C(x,y) + C_\varepsilon(x,y)] = \mu \ell_1 \qquad (y \in S_1)$$

c/ Krige $Z_o(x_o)$ at a point $x_o \notin S_o$.

Universality : $\int_{S_o} \lambda_o(dx) = 1$, $\int_{S_1} \lambda_1(dx) = 0$; Optimality :

$$\int_{S_o} \lambda_o(dx) \, C(x,y) + \int_{S_1} \lambda_1(dx) \, C(x,y) = C(x_o,y) + \mu_o \qquad (y \in S_o)$$

$$\int_{S_o} \lambda_o(dx) \, C(x,y) + \int_{S_1} \lambda_1(dx) \, [C(x,y) + C_\varepsilon(x,y)] = C(x_o,y) + \mu_1 \qquad (y \in S_1)$$

Exercise 15 (Simplified case of Cokriging) -

a/ If $S_1 = S_2 = \ldots = S_d = S$ (the d realizations of the $Z_i(x)$ being known on the same set S) and if the d drifts $m_i(x)$ are expressed by the same functions $f^\ell(x)$ and are linearly independent, it is better to change the notation and put $m_i(x) = a_{i\ell} \, f^\ell(x)$ with an unknown matrix $a_{i\ell}$. Denoting by $\sigma_{ij}(x,y)$ the cross covariances, by $A_{i\ell} = \lambda_{i\ell}^{j\alpha} \, Z_{j\alpha}$ (notation in the finite case $S = \{x_\alpha \, , \, \alpha = 1,2,\ldots N\}$, $i,j = 1,2,\ldots d$, $\ell = 0,1,\ldots k$) the optimal estimator of $a_{i\ell}$, by $Z_i^*(x_o) = \lambda_i^{j\alpha} \, Z_{j\alpha}$ the kriging of $x_o \notin S$, for the systems of the equations for the $\lambda_{i\ell}^{j\alpha}$ and the $\lambda_i^{j\alpha}$. (We get respectively :

$$\begin{cases} \lambda_{i\ell}^{j\alpha} \; f_\alpha^s = \delta_i^j \, \delta_\ell^s \\ \\ \lambda_{i\ell}^{j\beta} \; \sigma_{jj';\beta\alpha} = \mu_{ij';\ell s} \, f_\alpha^s \end{cases} \qquad (\mu_{ij'\ell s} = \operatorname{Cov} A_{i\ell} \, A_{j's})$$

and :

$$\begin{cases} \lambda_i^{j\alpha} \; f_\alpha^s = \delta_i^j \; f^s(x_o) \\ \\ \lambda_i^{j\beta} \; \sigma_{jj';\beta\alpha} = \mu_{ij';s} \; f_\alpha^s \end{cases}$$

b/ Let B_j^i be a regular d × d matrix, and B'^i_j its inverse. Consider the R.F.'s $Z'_i = B_i^j \, Z_j$ deduced from the preceding ones by the matrix B_i^j . Show that the coefficients

$\lambda_{i\ell}^{j\alpha}$ and $\lambda_i^{j\alpha}$ of the preceding optimal estimators vary at one time in a covariant manner and at another time in a contravariant manner, whilst the Lagrange parameters behave in a covariant way (i.e. μ'_{ij} , - = $B_i^{i'}$ $B_j^{j'}$ $\mu_{i'j'}$,- ; $\lambda'^{i'}_{j'}$ =$B'^{i'}_{i}$ $B^j_{j'}$, λ^i_j).

c/ Let $Z_i(x)$ be d R.F.'s as above, independent, having the same covariance function $\sigma(x,y)$. Show that $\sigma_{jj';\beta\alpha}$ = 0 or $\sigma_{\beta\alpha}$, whether $j \neq j'$ or $j = j'$. Deduce from this that the optimal estimators of the drift and of kriging for the R.F. $Z_i(x)$ do not call in the other R.F.'s Z_j, $j \neq i$, i.e. :

$$A_{i\ell} = \lambda_\ell^\alpha \, Z_{i\alpha} \qquad\qquad Z_i^*(x_o) = \lambda^\alpha \, Z_{i\alpha}$$

with coefficients (independent of i) verifying :

$$\begin{cases} \lambda_\ell^\alpha \, f_\alpha^s = \delta_\ell^s \\ \\ \lambda_\ell^\beta \, \sigma_{\alpha\beta} = \mu_{\ell s} \, f_\alpha^s \end{cases} \qquad \text{and} \qquad \begin{cases} \lambda^\alpha \, f_\alpha^s = f^s(x_o) \\ \\ \lambda^\beta \, \sigma_{\alpha\beta} = \mu_\ell \, f_\alpha^\ell \end{cases}$$

d/ Let $Z'_i(x)$ be R.F.'s defined as in a/ such that $\sigma_{ij}(x,y) = K_{ij} \, \sigma(x,y)$ with a strictly positive definite matrix K_{ij} and a covariance function $\sigma(x,y)$. Show that we may find d independent random functions $Z_i(x)$ as in c/ having the same covariance function $\sigma(x,y)$ such that $Z'_i(x) = B_i^j \, Z_j(x)$ with a matrix B_i^j verifying :

$$K_{ii'} = \sum_j \, B_i^j \, B_{i'}^{j'}$$

(diagonalize the matrix K_{ij}).

Deduce that the optimal estimators' coefficients are the same as in paragraph c/, hence that the drift or the kriging of one of these R.F.'s is estimated <u>as if it alone was present</u>. ($B'^{i'}_{i}$ $B^j_{j'}$, $\delta^i_j = \delta^{i'}_{j'}$)

(Apply the variation law of the coefficient found in b/ : in applications, <u>the case where all the cross-covariances are proportional to the same given covariance</u> is particularly interesting, because of the great simplification this allows).

4-7-3 Indeterminability of the underlying covariance.

Exercise 16 (n correlated Gaussian variables) - Let $X_1, X_2, \ldots X_n$ be n Gaussian variables having the same expectation m, the same variance σ^2, and such that the X_i, X_j have, pair by pair, the same correlation coefficient $\rho_{ij} = \rho$ $(i \neq j)$.

a/ Show that the covariance matrix is $\sigma_{ij} = [(1-\rho)\,\delta_{ij} + \rho]\sigma^2$, and as for an inverse :

$$B_{ij} = \frac{1}{(1-\rho)\sigma^2}\left(\delta_{ij} - \frac{\rho}{1 + (n-1)\rho}\right)$$

Compute the determinant of σ_{ij} $(\sigma^{2n}\,(1-\rho)^{n-1}\,(1+(n-1)\rho)$. Deduce that the density $f(x_1, \ldots x_n)$ of these variables verifies

$$-2 \log f = n \log 2\pi + n \log \sigma^2 + (n-1) \log (1-\rho) + [1 + (n-1)\rho]$$

$$+ \sum_{i,j} B_{ij}\,(x_i - m)\,(x_j - m)$$

(it will be noticed that the eigenvalues of σ_{ij} are $(1-\rho)\sigma^2$ of order n-1 (eigenvectors a_i such that $\Sigma\, a_i = 0$) and $1 + (n-1)\rho$ of order 1 (eigenvectors : $a_i = 1$). Notice the condition $1 \geq \rho \geq -\frac{1}{n-1}$).

b/ A realization $(x_1, x_2 \ldots x_n)$ of $X_1, X_2, \ldots X_n$ is supposed to be known. Form the estimators of the maximum likelihood m^*, s^2 and r of m, σ^2 and ρ. To do this, put

$$S_1 = \sum_i (x_i - m) \qquad S_2 = \sum_i (x_i - m)^2$$

and notice that $\frac{\partial S_1}{\partial m} = -n$, $\frac{\partial S_2}{\partial m} = -2\,S_1$, and :

$$\sum_{ij} B_{ij}(x_i - m)(x_j - m) = \frac{1}{(1-\rho)\sigma^2}\left(S_2 - \frac{\sigma}{1 + (n-1)\sigma}\, S_1^2\right)$$

Show that the equation in m gives $S_1 = 0$, i.e. $m^* = \frac{1}{n}\Sigma\, x_i$, that the equation in s^2 leads to :

$$(1-r)\, s^2 = \frac{1}{n}\, S_2$$

and that the equation in r leads to an absurd result, in which the experimental data

no longer appear.

Conclude that $(1-\rho)\sigma^2$ can be estimated (if need be), but not σ^2 and ρ separately.

c/ Explain this result with the help of the theory of quadratic universal estimators. ($Q^{ij} X_i X_j$ is universal if $\sum Q^{ij} = 0$. But $E(Q^{ij} X_i X_j)$ is then equal to $\sum_{ij} Q^{ij} \sigma_{ij}$ $= (1-\rho)\sigma^2 \sum_i Q^{ii} = (1-\rho)\sigma^{2^{ij}}$ represents the inaccessible part of the covariance, and ρ and σ^2 cannot be separated).

d/ Show that these conclusions remain valid for k independent vectorial variables $(X_1^k, \ldots X_n^k)$, all of which have the same covariance σ_{ij} as above, but a different expectation $m_k = E(X_i^k)$.

Exercise 17 - On a straight line, consider a R.F. having a covariance $\sigma(x,y)$ and a linear drift $a_o + a_1 x$. The covariance as well as the coefficients a_o and a_1 are unknown. The realization is supposed to be known in the interval (-R, R).

a/ In order to estimate a_o and a_1, we use the universal (but obviously not optimal in general)estimators

$$A_o = \frac{Z_R + Z_{-R}}{2} \quad , \quad A_1 = \frac{Z_R - Z_{-R}}{2}$$

(cf. Exercise 2). Show that

$$D^2(A_o) = \frac{1}{4}\sigma_{RR} + \sigma_{-R-R} + 2\sigma_{R,-R}) \quad , \quad D^2(A_1) = \frac{1}{4R^2}(\sigma_{RR} + \sigma_{-R,-R} - 2\sigma_{R,-R})$$

and Cov $(A_o\ A_1)$ = Cov $(A_1\ A_o)$ = 0. (These two variances represent the inaccessible part of the covariance).

Deduce the general form of the admissible covariances $\tau(x,y)$ as a function of one of them $\sigma(x,y)$:

$$\tau(x,y) = \sigma(x,y) + D_o + D_1\ x\ y \qquad (x,y \in (-R, R))$$

b/ Suppose that, among all the possible covariances, the "true" unknown covariance is the most stationary possible. To find it, start from a particular solution $\tau(x,y)$, equal for instance to $E(Z_x Z_y - m_x^* m_y^*)$, obtained by experiment. Put :

$$\sigma(x,x+h) = \tau(x,x+h) - D_0 - D_1\, x(x+h) \qquad (0 \le h \le 2R)$$

$$\overline{\sigma}(h) = \frac{1}{2R-h} \int_{-R}^{R-h} \sigma(x,x+h)\, dx \quad , \quad \overline{\tau}(h) = \frac{1}{2R-h} \int_{-R}^{R-h} \tau(x,x+h)$$

hence

$$\overline{\sigma}(h) = \overline{\tau}(h) - D_0 + D_1 \left[\frac{1}{6} (2R-h)^2 - R(R-h) \right]$$

Determine D_1 in order to minimize the integral :

$$\int_0^{2R} dh \int_{-R}^{R-h} [\sigma(x,x+h) - \overline{\sigma}(h)]^2\, dx$$

(notice that D_0 is eliminated and remains indeterminate).

BIBLIOGRAPHY

[1] CARLIER, A. - Contribution aux méthodes d'estimation des gisements d'uranium. (Thesis) - Fontenay-aux-Roses, 1964.

[2] FORMERY, Ph. - Cours de Geostatistics - School of Mines of Nancy (unpublished).

[3] MATHERON, G. - Traité de Geostatistique Appliquée. Tome 1 (1962), Tome 2 (1963). Editions Technip, Paris.

[4] MATHERON, G. - Les variables régionalisées et leur estimation (Thesis) - 1965 Masson et Cie, Paris.

[5] MATHERON, G. - Osnovy Prikladnoĭ geostatistiki.-1968, Editions Mir, Moscow

[6] MATHERON, G. - Le Krigeage Universel - Les Cahiers du Centre de Morphologie Mathématique, Fascicule 1, 1969 - Fontainebleau.

[7] SERRA, J. - Echantillonnage et estimation locale des phénomènes de transition miniers. (Thesis) - Nancy, 1967.

[8] FORMERY, Ph. and MATHERON, G. - Recherche d'optima dans la reconnaissance et la mise en exploitation des gisements miniers. Annales des Mines, May and June 1963.

INDEX

R

S

T

U

V

W

Z