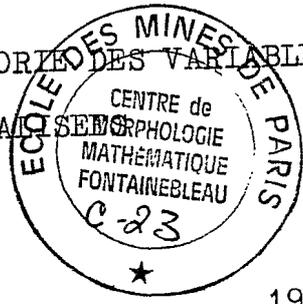


BIBLIOTHEQUE

C-23

ECOLE D'ETE 1968

RESUME DE LA THEORIE DES VARIABLES
REGIONALES



G. MATHERON

1968

BIBLIOTHEQUE

Résumé de la théorie des Variables Régionalisées.

Table des matières

<u>INTRODUCTION</u>	1
Notations	1
Produit de convolution (moyenne mobile)	3
Géostatistique et théorie des VR	4
<u>LES METHODES TRANSITIVES</u>	7
I - Exemple introductif	7
II - Le covariogramme transitif	9
III - Régularisation d'une VR, Montée	10
IV - L'estimation des VR	12
V - Application à l'estimation d'une surface	15
<u>THEORIE DES F.A. INTRINSEQUES</u>	17
I - Définitions générales	17
II - Propriétés de la covariance et du demi-variogramme	22
III - Régularisation d'une FA intrinsèque Montée	25
IV - Variance d'extension, variance d'estimation	29
variance de v dans V	31
Mailles aléatoires et aléatoires stratifiées	33
V - Effet de pépite et phénomènes de transition	35
VI - Calcul des variances d'estimation	38
1 ^o / Cas de l'espace à une dimension	38
2 ^o / Cas de l'espace à deux dimensions	40
3 ^o / Cas de l'espace à trois dimensions	43
VII - Le schéma de de Wijs	44
Equivalents linéaires	45
Schéma de de Wijs à deux dimensions	46
Schéma de de Wijs à trois dimensions	48
VIII - Le schéma sphérique	49
Exercices sur les méthodes transitives	51
Exercices sur les FA intrinsèques	53
Figure 2 (Zitterbewegung)	57
Abaques pour le schéma de de Wijs	58
Abaques pour le schéma sphérique	66

ECOLE D'ETE DE FONTAINEBLEAU

GEOSTATISTIQUE

Résumé de la théorie des variables régionalisées

INTRODUCTION

NOTATIONS

Nous désignerons par x , y , ... des points de l'espace à n dimensions ($n = 1, 2$, ou 3), par dx , dy ... des éléments de longueur ($n = 1$), de surface ($n = 2$) ou de volumes ($n = 3$) centrés en ces points, par $f(x)$, $g(y)$ etc.. des fonctions de ces points.

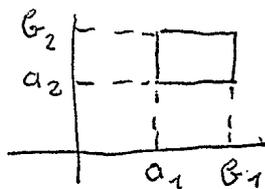
L'intégrale (étendue à tout l'espace) d'une fonction $f(x)$ se note :

$$\int f(x) dx$$

Par exemple, pour $n = 3$ dimensions, et en désignant par (x_1, x_2, x_3) les trois coordonnées du point x , cette notation s'explique ainsi :

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

De même, nous noterons $\int_A f(x) dx$ l'intégrale de la fonction $f(x)$ étendue à un domaine A de l'espace à n dimensions.



Pour $n = 2$, par exemple, et si A est le rectangle représenté ci-joint, on a explicitement :

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

Ces notations sont à la fois très condensées et très parlantes : en première lecture, on peut les interpréter dans le cas de l'espace à une seule dimension, où leur signification est, en général, très claire. La généralisation aux espaces à deux ou trois dimensions est ainsi rendue plus facile.

Exemple. Soit V un volume, et $g(h) = g(h_1, h_2, h_3)$ une fonction du vecteur h de coordonnées h_1, h_2, h_3 . Soient $x = (x_1, x_2, x_3)$ et $y = (y_1, y_2, y_3)$ l'origine et l'extrémité de ce vecteur h , c'est à dire $h = y - x$. La valeur moyenne de la fonction $g(h)$ lorsque les deux extrémités x et y du vecteur h décrivent (chacune pour son propre compte) le volume V se note :

$$\frac{1}{V^2} \int_V dx \int_V g(y - x) dy$$

En notation explicite, ceci s'écrirait :

$$\frac{1}{V^2} \iiint_V dx_1 dx_2 dx_3 \iiint_V g(y_1 - x_1; y_2 - x_2; y_3 - x_3) dx_1 dy_2 dy_3$$

La première notation est un symbole, qui désigne directement le concept de valeur moyenne; la deuxième est un algorithme qui indique la marche à suivre pour calculer cette quantité, mais dont la signification n'apparaît plus au premier coup d'oeil.

PRODUIT DE CONVOLUTION (moyenne mobile)

Le produit de convolution de deux fonctions $f_1(x)$ et $f_2(x)$ est la fonction $g(x)$ définie par :

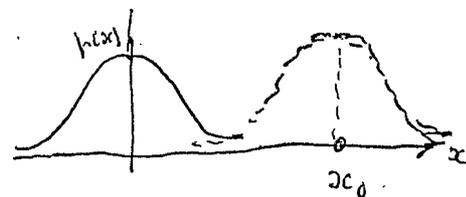
$$g(x) = \int f_1(y) f_2(x - y) dy = \int f_2(y) f_1(x - y) dy$$

On écrit en notation symbolique :

$$g = f_1 * f_2$$

Cette opération de convolution joue un rôle fondamental en géostatistique (comme aussi en probabilités et dans toute la physique théorique). On peut la rattacher à la notion intuitive de "moyenne mobile" :

Régularisée d'une fonction f (moyenne mobile pondérée de cette fonction)



Soit $p(y)$ une fonction de pondération. La valeur au point x_0 de la moyenne mobile de la fonction f pondérée par p est :

$$f_p(x_0) = \int p(y) f(x_0 + y) dy = \int p(-y) f(x_0 - y) dy$$

(on attribue le poids $p(y) dy$ à la valeur prise par f au point $x_0 + y$, et on obtient la deuxième expression en changeant y en $-y$).

Soit $\overset{\vee}{p}$ la transposée de la fonction p (définition : $\overset{\vee}{p}(x) = p(-x)$).

$f_p(x_0)$ est la valeur en x_0 de $f * \overset{\vee}{p}$:

$$f_p = f * \overset{\vee}{p}$$

Cette moyenne mobile f_p de la fonction f (pondérée par p) s'appelle régularisée (de f par p)

Exemple 1. (prélèvement d'un échantillon v implanté au point x)

Soit v l'échantillon implanté à l'origine des coordonnées, et $k(x)$ sa fonction indicatrice ($k(x) = 1$ si $x \in v$ et $k(x) = 0$ si $x \notin v$). Prenons comme fonction de pondération $p(x) = \frac{1}{\sigma} k(x)$. La moyenne mobile correspondante

$$f_{\sigma} = \frac{1}{\sigma} f * k$$

représente la teneur moyenne de l'échantillon v prélevé au point x . Explicitement, on a :

$$f_{\sigma}(x) = \frac{1}{\sigma} \int_{\sigma} f(x+y) dy$$

Exemple 2. (radioactivités)

Si l'on place à l'origine 0 des coordonnées une masse unité de substance radioactive, on observe à la distance d une radioactivité $A e^{-\lambda d}/d$. (A est une constante, λ le coefficient d'absorption du milieu). Soit alors $f(x)$ la teneur au point x en cette substance radioactive. Désignons par $d(x - x_0)$ la distance entre deux points x et x_0 . La masse $f(x) dx$ placée en x entraîne en x_0 une radioactivité $A e^{-\lambda d(x_0-x)}/d(x_0-x)$. Au total, on observe en x la radioactivité :

$$A \int f(x) \frac{e^{-\lambda d(x_0-x)}}{d(x_0-x)} dx$$

Ce n'est pas autre chose que la produit de convolution $A f * p$ (avec une fonction de pondération $p(x) = e^{-\lambda d}/d$, d'ailleurs symétrique : $p = \check{p}$)

GEOSTATISTIQUE ET THEORIE DES VARIABLES REGIONALISEES.

La Géostatistique est l'application de la théorie des variables régionalisées à l'estimation des gisements miniers (avec tous les tours de main que cela implique). Plus généralement, nous dirons d'un phénomène qui se déploie dans l'espace et y manifeste une certaine structure, qu'il.

est régionalisé. Les sciences de la terre, entre autres, en fournissent de nombreux exemples. Si $f(x)$ désigne la valeur au point x d'une caractéristique f de ce phénomène, nous dirons que $f(x)$ est une variable régionalisée, en abrégé V.R. . C'est là un terme neutre, purement descriptif, antérieur, en particulier, à toute interprétation probabiliste.

Du point de vue mathématique, une V.R. est donc simplement une fonction $f(x)$ du point x , mais c'est, en général, une fonction fort irrégulière :
ex. : une teneur dans un gisement minier. Elle se présente sous deux aspects contradictoires (ou complémentaires) :

- un aspect aléatoire (haute irrégularité, et variations imprévisibles d'un point à l'autre)
- un aspect structuré (elle doit cependant refléter à sa manière les caractéristiques structurales du phénomène régionalisé)

La théorie des V.R. se propose donc deux objectifs principaux :

- sur le plan théorique, exprimer ces caractéristiques structurales sous une forme mathématique adéquate;
- sur le plan pratique, résoudre la problème de l'estimation d'une V.R. à partir d'un échantillonnage fragmentaire.

Ces deux objectifs sont liés : pour un même réseau de prélèvements, l'erreur d'estimation dépend des caractéristiques structurales; elle est, par exemple, d'autant plus élevée que la V.R. est plus irrégulière et plus discontinue dans sa variation spatiale.

Champ et support d'une V.R.

Le champ V d'une V.R. est le domaine où celle-ci est différente de 0.
Un panneau est un sous-ensemble V' de V .

Support. Souvent, on ne connaît pas $f(x)$ elle-même, mais seulement sa valeur moyenne $f_v(x)$ dans l'échantillon v prélevé au point x . Cette \bar{X} régularisée $f_v(x)$ est effectivement plus régulière que la V.R. $f(x)$. Le volume v s'appelle le support de la V.R. f_v , régularisée de f .

Une autre tâche importante de la théorie des V.R. consistera donc à déterminer les caractéristiques de f_v connaissant celles de f . Ex. : dans un gisement, prévoir les caractéristiques des panneaux V' (variable $f_{V'}$), connaissant celles de f ou de f_v (échantillons).

Plus généralement, on cherchera à relier les caractéristiques de f à celles d'une régularisée f_h par une fonction p donnée (exemple des radioactivités).

METHODES TRANSITIVES ET THEORIE INTRINSEQUE

Pour atteindre ces objectifs, nous disposons de deux groupes de méthodes :

- méthodes transitives : absolument générales, et ne nécessitant en particulier aucune hypothèse de nature probabiliste, et a fortiori aucune hypothèse de stationnarité

- théorie intrinsèque : c'est une application de la théorie des fonctions aléatoires; on introduit donc des interprétations probabilistes, et même une certaine hypothèse de stationnarité (hypothèse intrinsèque)

Du point de vue théorique, ces deux groupes de méthodes conduisent à des résultats équivalents : c'est important pour la méthodologie, car cela montre que les résultats de la théorie intrinsèque ne sont pas réellement liés à l'hypothèse de la stationnarité (on peut d'ailleurs construire une théorie probabiliste affranchie de cette hypothèse et permettant de retrouver les principaux résultats de la géostatistique)

En pratique, la théorie intrinsèque est plus facile à mettre en oeuvre, et c'est toujours elle que l'on utilise, sauf toutefois dans le cas particulier très important de l'estimation d'une surface ou d'un volume (probleme géométrique)

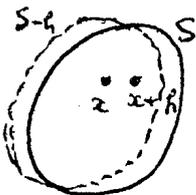
PREMIERE PARTIE - LES METHODES TRANSITIVES

I - Exemple introductif

Considérons le phénomène de "transition" le plus simple que l'on puisse imaginer : présence ou absence d'un caractère. Soit, par exemple, une formation géologique S d'extension limitée. Un sondage foré au point x la rencontre, ou ne la rencontre pas. Posons :

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

Il s'agit là d'un phénomène unique, pour lequel aucune formulation probabiliste n'est possible : parler de la probabilité pour qu'un point x donné appartienne à S n'aurait pas grand sens. On peut noter aussi que tout l'intérêt se concentre ici sur la frontière de S. En effet, k(x) est constante à l'intérieur comme à l'extérieur de S, et c'est au franchissement de cette frontière seulement que k(x) varie, passant de 1 à 0 ou de 0 à 1. De là le nom de phénomène de transition, et de méthodes transitives.



L'aire S de notre formation est évidemment donnée par :

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

La valeur de S constitue un renseignement fort intéressant du point de vue pratique : le plus souvent, c'est elle que l'on cherche à estimer à partir d'un réseau de sondages. Toutefois, ce paramètre scalaire ne nous apporte encore aucune information de nature réellement structurale. On peut, en effet, définir la structure d'un ensemble comme le système des relations existant entre les éléments ou les parties de cet ensemble. Nous n'aurons donc d'information de nature structurale sur notre surface S qu'à la condition de faire intervenir simultanément au moins deux points.

Soient donc x et $x+h$ deux points (c'est à dire le plus petit ensemble structurant que nous puissions imaginer). Considérons alors l'expression $k(x) k(x+h)$: elle vaut 1 si x et $x+h$ appartiennent tous les deux à S , et 0 autrement. Mais dire que $x+h$ appartient à S équivaut à dire que x appartient au translaté $S-h$ de S dans la translation $-h$. On a donc :

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

Intégrons en x cette expression : nous obtenons une fonction de h :

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

qui représente la mesure (l'aire) de l'intersection de S et de son translaté par $-h$. Cette fonction est symétrique, car les deux intersections $S \cap S-h$ et $S \cap S_h$ se déduisent l'une de l'autre par translation. Cette fonction $K(h)$ est le covariogramme géométrique associé à S . Il donne une certaine image de la forme de l'ensemble S :

Propriétés du covariogramme géométrique $K(h)$

a/ Symétrie : $K(h) = K(-h)$; inégalités $0 \leq K(h) \leq K(0)$

Relations $S = K(0)$

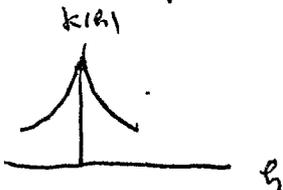
$$S^2 = \int K(h) dh$$

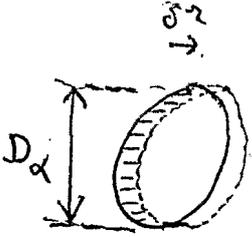
b/ Portées : la portée $a(\alpha)$ dans la direction α est la distance à partir de laquelle $K(h)$ s'annule dans cette direction. C'est donc la plus grande dimension de S dans cette direction.

c/ Pente à l'origine (dérivées à droite et à gauche)

Si le module δ_r de h est petit, on a :

$$K(h) = K(0) - \delta_r D_\alpha$$





δr D_α est la moitié de la petite surface balayée par le vecteur δr dont l'origine décrit le contour de S. D_α est la variation diamétrale de S dans la direction α (si S est convexe c'est le diamètre apparent dans cette direction.)

II - Le covariogramme transitif (cas général)

Soit $f(x)$ une V.R. nulle en dehors d'un champ V borné. Le covariogramme transitif de cette V.R. est la fonction $g(h)$ définie par :

$$(1) \quad g(h) = \int f(x) f(x+h) dx$$

Propriétés du covariogramme transitif $g(h)$

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

Soit $Q = \int f(x) dx$ la quantité de métal. On a :

$$(2) \quad Q^2 = \int g(h) dh$$

(pour démontrer (2), remplacer g par son expression (1), et intégrer d'abord en h).

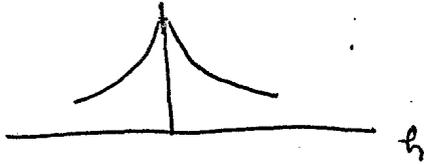
b/ Portée : $a(\alpha)$ est définie par la condition $g(h) = 0$ dès que $|h| > a(\alpha)$ pour le vecteur h de direction α : c'est une propriété du champ de la V.R.

c/ Comportement du $g(h)$ au voisinage de l'origine. La régularité de $g(h)$ au voisinage de l'origine reflète les propriétés de continuité de la V.R. dans sa variation spatiale. Cela résulte de :

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

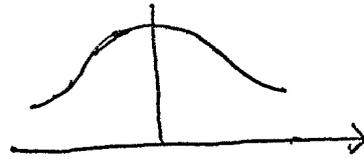
Si $f(x)$ est continue par morceaux, $g(h)$ a un comportement linéaire au voisinage de l'origine. Si $f(x)$ est dérivable, $g(h)$ a un comportement

parabolique :



V.R. continue par morceaux

$g(r)$



V.R. dérivable

Cas isotrope. Si $g(h) = g(r)$ ne dépend que du module $r = |h|$ du vecteur h , son développement au voisinage de l'origine comporte une partie régulière (termes de degré entier pair) et une partie irrégulière (termes en r^λ avec λ différent d'un entier pair, ou encore termes en $r^{2n} \log r$) :

$$g(h) = \underbrace{g(0) + a_2 r^2 + \dots}_{\text{partie régulière}} + \underbrace{\sum_{\lambda} c_{\lambda} r^{\lambda} + C_{2n} r^{2n} \log r}_{\text{partie irrégulière}}$$

Le terme de plus bas degré de la partie irrégulière caractérise l'irrégularité de la V.R. dans sa variation spatiale. Ex. : la V.R. géométrique associée à une surface S a un covariogramme en $K(h) = S - D|h| + \dots$: le terme de plus bas degré de la partie irrégulière est de degré 1.

III - Régularisation d'une V.R., et montées.

Soit $f(x)$ une V.R., $p(x)$ une fonction de pondération, $f_h = f *^v p$ la régularisée de f par p . La relation (1) peut déjà s'écrire :

$$g = f *^v f$$

et montre que le covariogramme transitif est l'autorégularisée de la V.R. f . Le covariogramme $g_h = f *^v p *^v f *^v p = f *^v f *^v p *^v p = f_h *^v f_h$ se met aussi sous la forme :

$$g_h = g *^v P$$

avec $P = p *^v p$: on obtient le covariogramme de la régularisée en régularisant g par le covariogramme transitif P de la fonction de pondération p : c'est

effectivement une fonction plus régulière que g , de même que f_h est plus régulière que la V.R. initiale f .

Montée. (cas limite de la régularisation). Si $f_3(x) = f_3(x_1, x_2, x_3)$ est une V.R. dans l'espace à 3 dimensions, on dit que la V.R. :

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

définie dans l'espace à deux dimensions se déduit de f par montée. Par exemple, si $f_3(x)$ est une teneur ponctuelle, $f_2(x_1, x_2)$ est l'accumulation (quantité de métal au mètre carré) du sondage implanté au point (x_1, x_2) de la surface topographique.

On démontre que le covariogramme $g_2(h_1, h_2)$ de la V.R. f_2 déduite de f_3 par montée s'obtient en effectuant cette même opération de montée sur le covariogramme $g_3(h_1, h_2, h_3)$ de f_3 .

Montée dans le cas isotrope. Si $g_3(h) = g_3(r)$ ne dépend que du module r du vecteur h , on peut effectuer la montée terme à terme sur la partie irrégulière de g_3 (ce principe de correspondance terme à terme détermine donc le covariogramme g_2 à une série entière paire près). A tout terme en r^λ correspond ainsi un terme en $r^{\lambda+1}$. Si λ est un entier impair, ou dans le cas d'un terme logarithmique, on obtient la séquence singulière :

$$\begin{aligned} \log r &\rightarrow \pi r \\ r &\rightarrow -r^2 \log r \end{aligned}$$

où alternent termes impairs et termes logarithmiques. On notera que, dans tous les cas, la montée a un effet régularisant.

IV - L' estimation des V.R.

a/ Expression rigoureuse de la variance d'estimation. Nous raisonnons dans l'espace à une seule dimension, mais les résultats se généralisent sans peine. Soit $f(x)$ une V.R. Pour estimer la quantité de métal :

$$Q = \int f(x) dx$$

on dispose d'un réseau de prélèvements à maille régulière a . Si x_0 est l'implantation de l'un (quelconque) de ces prélèvements, on connaît donc les valeurs numériques des $f(x_0 + pa)$ pour p entier positif ou négatif. En fait, sauf un nombre fini de ces valeurs est $\neq 0$, si le champ est borné, et il suffit donc en réalité que le réseau effectif ait quelque peu débordé ce champ. On prend comme estimateur :

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

L'erreur d'estimation est $Q - Q^*(x_0)$: c'est une fonction périodique de période a de x_0 , car il est manifestement indifférent de prendre comme origine du réseau l'un ou l'autre des prélèvements. Pour faire de cette erreur une variable aléatoire, il suffit donc d'admettre que l'"origine" x_0 du réseau est choisie au hasard avec une densité de probabilité uniforme sur le segment $(0, a)$. On a alors :

$$E[Q^*(x_0)] = a \int_0^a \frac{dx_0}{a} \sum_p f(x_0 + pa) = \int_{-\infty}^{+\infty} f(x) dx = Q$$

est la variance d'estimation (variance de cette erreur aléatoire) est :

$$\sigma^2(a) = \frac{1}{a} \int_0^a [Q^*(x_0) - Q]^2 dx_0 = \frac{1}{a} \int_0^a [Q^*(x_0)]^2 dx_0 - Q^2$$

Evaluons l'intégrale de $[Q^*(x_0)]^2$. On a :

$$\begin{aligned} [Q^*(x_0)]^2 &= a^2 \sum_{p=-\infty}^{+\infty} \sum_{q=-\infty}^{+\infty} f(x_0 + pa) f(x_0 + qa) = \\ &= a^2 \sum_k \sum_h f(x_0 + ka) f(x_0 + ka + ha) \end{aligned}$$

Intégrons de 0 à a. Il vient :

$$\begin{aligned} \int_0^a [\varphi^*(x_0)]^2 dx_0 &= a^2 \sum_k \int_0^a \sum_h f(x_0+ta) f(x_0+ta+ka) dx_0 = \\ &= a^2 \sum_k \int_{-\infty}^{+\infty} f(x_0) f(x_0+ka) dx_0 = a^2 \sum_k g(ka) \end{aligned}$$

Compte tenu aussi de la relation (2), on obtient donc la formule :

$$(3) \quad \sigma^2(a) = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h) dh$$

Dans l'espace à n dimension, on a un résultat tout à fait analogue. Par exemple, pour $n = 3$, et une maille parallélépipédique $a_1 a_2 a_3$, il vient :

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

Remarque. La variance d'estimation (3) est la différence entre une valeur approchée et la valeur exacte de l'intégrale $\int g(h) dh$. Elle est donc d'autant plus petite que :

- la maille a est plus petite ,
- la fonction g, donc aussi la V.R. elle-même est plus

régulière.

Si la maille a est petite vis à vis de la portée, la formule (3) comporte un grand nombre de termes. On est conduit à chercher des formules d'approximation.

b/ Formules d'approximation pour l'espace à une dimension.

Dans l'espace à une seule dimension, et pour une maille a petite vis à vis de la portée, on peut appliquer un principe de correspondance terme à terme entre la partie irrégulière du $g(h)$ et le développement limité de la variance d'estimation $\sigma^2(a)$ au voisinage de $a = 0$. A tout terme

irrégulier en r^λ correspond un terme en $A_\lambda a^{\lambda+1}$ (A_λ constante ne dépendant que de λ). En particulier :

$$\begin{aligned} -r &\rightarrow \frac{1}{6} a^2 \\ r^2 \log r &\rightarrow 0,0609 a^3 \end{aligned}$$

Remarque. Si $L = na$ est la longueur minéralisée, et n le nombre de prélèvements positifs, au covariogramme en r^λ correspond la variance d'estimation :

$$(4) \quad \sigma^2(a) = A_\lambda a^{1+\lambda} = (A_\lambda L^{1+\lambda}) \frac{1}{n^{1+\lambda}}$$

Elle est en raison inverse de $n^{1+\lambda}$ (et non en $1/n$ comme l'aurait suggéré une application maladroite de la statistique classique)

Zitterbewegung, ou terme fluctuant. Le principe de correspondance énoncé ci-dessus ne constitue qu'une approximation. La valeur exacte de $\sigma^2(a)$ calculée à partir de (3) peut en différer notablement (Figure ci-jointe). La différence entre valeur exacte et valeur approchée peut être élevée, mais elle est nulle en moyenne : c'est le terme fluctuant, ou Zitterbewegung. Son ampleur peut aller de 1 à 3 sur la figure ci-jointe.

c/ Formules d'approximation pour l'espace à deux ou trois dimensions.

A deux dimensions (toujours exclusion faite d'un Zitterbewegung) on a des formules d'approximation du type :

$$\sigma^2(a_1, a_2) = B_\lambda a_2^{2+\lambda} + C_\lambda a_1^{1+\lambda} a_2$$

pour un $g(h)$ en $|h|^\lambda$ et avec $a_1 \leq a_2$. Si $a_1 = 0$, on connaît

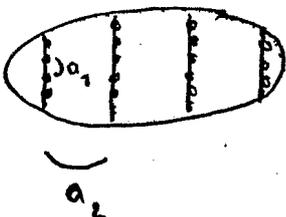
parfaitement les accumulations des lignes, et le terme

$B_\lambda a_2^{2+\lambda}$ est la variance de l'erreur que l'on commet en

estimant Q à partir des résultats (supposés parfaitement

connus) de ces lignes : c'est le terme de tranche

(extension des lignes dans leur



(extension des lignes dans leurs tranches d'influence). On a des résultats analogues dans l'espace à 3 dimensions.

V - Application à l'estimation d'une surface S

Appliquons la formule (4) à la V.R. $f(x) = k(x)$ égale à 1 ou à 0 selon que x appartient ou non à une surface S : il s'agit donc de l'estimation de l'aire de cette surface S à partir d'un réseau de sondages à maille rectangulaire a_1, a_2 ($a_1 \leq a_2$). Le covariogramme transitif $K(h)$ associé à S est linéaire au voisinage de l'origine, et on a :

$$K(h) = K(0) - |h| D_\alpha$$

D_α représentant, comme nous l'avons vu, la demi-variation diamétrale de S dans la direction α du vecteur h .

a/ Plaçons nous, tout d'abord, dans le cas isotrope, c'est à dire dans le cas où $D_\alpha = D$ est à peu près indépendant de la direction α . La formule (4) est directement applicable, et donne :

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

La variance est en $1/n^{3/2}$, n désignant le nombre des sondages positifs. Pour utiliser effectivement cette formule à partir des données disponibles expérimentalement, on peut estimer S en attribuant à chaque sondage son rectangle d'influence, ce qui donne :

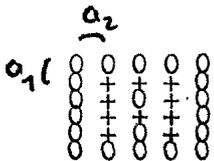
$$S = n a_1 a_2$$

On estimera D à partir du contour de la réunion de ces zones d'influence (voir figure ci-après). On comptera donc les nombres N_1 et N_2 des éléments parallèles à a_1 et a_2 respectivement qui constituent ce périmètre, et on aura $D = N_1 a_1 = N_2 a_2$ (puisque'il y a isotropie). Par suite, il vient :

$$(5) \quad \frac{\sigma_s^2}{S^2} = \frac{1}{n^2} \left[\frac{1}{6} N_2 + 0,061 \frac{(N_1)^2}{N_2} \right] \quad (N_2 \leq N_1)$$

B/ En général, cependant, le contour ne sera pas suffisamment isotrope pour que D_a puisse être regardée comme une constante D . Il présentera, par exemple, une direction principale d'allongement. Si l'un des cotés de la maille est parallèle à cette direction principale (ce qui sera souvent le cas), la formule (5) précédente reste applicable : en effet, prenant cette direction principale comme axe des x , et multipliant les ordonnées par un module convenable, nous obtenons une nouvelle figure, isotrope cette fois (au moins en première approximation) pour laquelle (5) est donc valable. Mais cette transformation linéaire n'a modifié ni N_1 , ni N_2 , ni la variance relative σ^2/S^2 , de sorte que (5) est valable aussi dans le cas de la figure anisotrope initiale.

Exemple. Sur la figure ci-après, l'aire minéralisée est estimée à 10 fois le rectangle de maille $a_1 a_2$. Elle comporte un trou (une lacune). Dans le décompte de N_1 et de N_2 doivent figurer aussi bien les éléments extérieurs que les éléments intérieurs. On lit donc sur la figure :



$$2 D_1 = 12 a_1 \quad \text{soit} \quad N_1 = 6$$

$$2 D_2 = 8 a_2 \quad \text{soit} \quad N_2 = 4$$

D'où, par conséquent :

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

soit un écart type relatif $\sigma_s/S = 11/100$, et une fourchette d'erreur relative de $\pm 22\%$.

On se gardera toutefois d'oublier que ce calcul fait abstraction du terme fluctuant, ou Zitterbewegung, dont nous savons que l'amplitude peut être énorme.

DEUXIEME PARTIE - THEORIE DES FONCTIONS ALEATOIRES INTRINSEQUESI - Définitions générales.

Notion de fonction aléatoire. En théorie des probabilités, on définit la notion de variable aléatoire (V.A.) vectorielle $Y = (Y_1, Y_2, \dots, Y_k)$ à k composantes : c'est une famille de k V.A. ordinaires Y_1, Y_2, \dots, Y_k (en général non indépendantes). Lorsque le nombre k de ces composantes devient infini, on obtient une famille infinie de variables aléatoires : c'est une fonction aléatoire. En particulier, si x est un point décrivant l'espace à n dimensions R^n , on peut définir une famille infinie $(Y_x)_{x \in R^n}$. A tout point x_0 de l'espace correspond ainsi une V.A. ordinaire Y_{x_0} , que l'on peut aussi bien noter $Y(x_0)$. $Y(x)$ est alors une fonction du point x , dont la "valeur" en x_0 n'est pas un nombre, mais une V.A. (à savoir $Y(x_0)$). On dit que $Y(x)$ est une fonction aléatoire (en abrégé, F.A.). On remarquera bien qu'en général les V.A. correspondant à deux points d'appui x_1 et x_2 , soient $Y(x_1)$ et $Y(x_2)$ ne sont pas indépendantes.

Si Y est une V.A. ordinaire, le résultat d'un tirage au sort effectué selon la loi de probabilité de Y est une valeur numérique particulière y . De même, si Y est une V.A. vectorielle (Y_1, Y_2, \dots, Y_k) , un tirage au sort selon la loi (à k variables) de Y donne un vecteur $y = (y_1, y_2, \dots, y_k)$ c'est à dire k valeurs numériques particulières. Enfin, si $Y(x)$ est une F.A. - c'est à dire une V.A. vectorielle à une infinité de composantes - un tirage au sort, effectué selon la loi (à une infinité de variables) de $Y(x)$ est une fonction numérique particulière $y(x)$, en général extraordinairement irrégulière. On dit que $y(x)$ est une réalisation de la F.A. $Y(x)$.

On peut toujours considérer une réalisation $y(x)$ d'une F.A. $Y(x)$ comme une variable régionalisée. Inversement, on peut interpréter une V.R. donnée $y(x)$ comme une réalisation d'une certaine F.A. $Y(x)$: cette interprétation permet d'appliquer aux V.R. les résultats de la théorie probabiliste des F.A.

Remarques. 1/ On ne peut jamais dire qu'une V.R. donnée $y(x)$ est une F.A. .
 Cela n'aurait pas plus de sens que de dire : le nombre 98 est une V.A.
 L'énoncé correct de l'hypothèse probabiliste de base que nous désirons introduire est : $y(x)$ est une réalisation d'une F.A. $Y(x)$.

2/ Pour que cette hypothèse probabiliste ait une signification réelle, il faut que l'on puisse reconstituer au moins en partie la loi de la F.A. $Y(x)$ dont la V.R. $y(x)$ est supposée être une réalisation, et cela suppose que l'inférence statistique soit possible. Or, l'inférence statistique n'est en général pas possible si l'on ne dispose que d'une seule réalisation $y(x)$ de $Y(x)$ (de même, on ne peut pas reconstituer la loi d'une V.A. Y à partir du résultat numérique $y = 98$ d'une épreuve unique). Pour que l'inférence statistique soit possible, il est nécessaire d'introduire des hypothèses supplémentaires sur la F.A. $Y(x)$, de manière à réduire le nombre des "paramètres" dont dépend sa loi. Tel est le but de l'hypothèse stationnaire que nous allons définir : une F.A. stationnaire se répète en quelque sorte elle-même dans l'espace, et cette répétition rend à nouveau possible l'inférence statistique à partir d'une réalisation unique. Précisons donc cette hypothèse.

F.A. stationnaire. On dit qu'une F.A. $Y(x)$ est stationnaire si sa loi est invariante par translation : autrement dit, si x_1, x_2, \dots, x_k sont k points d'appui arbitraires (k entier quelconque) et si h est un vecteur quelconque, les k V.A. $Y(x_1), Y(x_2), \dots, Y(x_k)$ ont la même loi (à k variables) que les k V.A. $Y(x_1 + h), Y(x_2 + h), \dots, Y(x_k + h)$. Dans ce qui suit, $Y(x)$ désignera une F.A. stationnaire.

Espérance mathématique. Considérons un point d'appui x_0 . Si la V.A. ordinaire $Y(x_0)$ admet une espérance mathématique, celle-ci est une fonction $m(x_0) = E[Y(x_0)]$ du point d'appui x_0 . Mais $Y(x)$ est stationnaire, et on a par suite $m(x_0 + h) = m(x_0)$ pour tout vecteur h , et $m(x_0)$ est donc une constante m indépendante de x_0 :

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

Quitte à remplacer $Y(x)$ par $Y(x) - m$, nous supposons toujours $m = 0$,

(pourvu toutefois que cette espérance existe).

La covariance $K(h)$. Considérons maintenant deux points d'appui x_0 et $x_0 + h$. Si les deux V.A $Y(x_0)$ et $Y(x_0 + h)$ admettent des variances finies (donc aussi une espérance m que nous supposerons nulle), elles admettent aussi une covariance $K(x_0; h)$, qui dépend en principe du point d'appui x_0 et du vecteur h . Mais, $Y(x)$ étant stationnaire, on a $K(x_0 + a; h) = K(x_0; h)$ pour tout vecteur a : $K(x_0; h)$ ne dépend donc pas de x_0 , et nous écrirons simplement $K(h)$:

$$(1) \quad K(h) = E [Y(x) Y(x + h)]$$

On comparera cette définition de la covariance avec celle du covariogramme transitif : $K(h)$ est la transposition probabiliste de $g(h)$.

Pour $h = 0$, on a $K(0) = E([Y(x)]^2)$: c'est la variance de la V.A. $Y(x_0)$. Pour que la F.A. stationnaire $Y(x)$ admette une fonction de covariance $K(h)$, il faut et il suffit qu'elle admette une variance finie $K(0)$.

Hypothèse stationnaire d'ordre 2. Nous dirons qu'une F.A. $Y(x)$ est stationnaire d'ordre 2 si la V.A. $Y(x_0)$ admet une espérance m indépendante du point d'appui x_0 , et si pour tout vecteur h la covariance :

$$K(h) = E [Y(x_0 + h) Y(x_0)] - m^2$$

existe et ne dépend pas de x_0 . Cette hypothèse (qui n'entraîne pas la stationnarité au sens strict, telle que nous l'avons définie ci-dessus) est suffisante pour la théorie des V.R. . Mais elle suppose l'existence d'une variance a priori finie $K(0)$

Variance a priori infinie. Or d'assez nombreux phénomènes présentent une capacité de dispersion illimitée, et ne peuvent être décrits correctement si on leur attribue une variance a priori finie : cette affirmation surprendra peut-être, mais il faut bien voir que la nature nous tend ici une sorte de piège. Lorsque l'on prélève des échantillons v dans un

champ V , on obtient un histogramme à partir duquel on peut toujours calculer numériquement une variance, qui prend ainsi une valeur parfaitement définie. Mais cette variance expérimentale est en réalité une fonction $\sigma^2(v|V)$ du support v et du champ V . Elle augmente, en particulier, lorsque le champ V augmente. Si les échantillons de taille v possèdent une variance a priori finie, celle-ci doit apparaître comme la limite pour V infini de la variance expérimentale $\sigma^2(v|V)$.

C'est ainsi que les auteurs d'Afrique du Sud (D.G. Krige etc.), à partir de centaines de milliers d'échantillons prélevés dans le grand gisement d'or du Rand, ont pu calculer la variance de ces échantillons dans des panneaux de plus en plus grand, puis dans une concession entière, puis dans le gisement du Rand dans son ensemble : ils ont ainsi obtenu une relation expérimentale de la forme :

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

La croissance de la variance se poursuit sans défaillance selon cette loi logarithmique (formule de De Wijs) jusqu'au dernier point expérimental, pour lequel V/v est de l'ordre de la dizaine de milliards. On peut conclure en toute certitude qu'il n'existe pas ici de variance a priori finie.

On est donc conduit à remplacer l'hypothèse stationnaire d'ordre 2 par une hypothèse plus faible (mais de signification analogue) :

Hypothèse intrinsèque. Même lorsque la variance a priori $K(0)$ n'existe pas (est infinie), il peut arriver que les accroissements $Y(x_0 + h) - Y(x_0)$ aient une variance finie. Nous dirons donc que la F.A. $Y(x)$ vérifie l'hypothèse intrinsèque si, pour tout vecteur h , l'accroissement $Y(x_0 + h) - Y(x_0)$ admet une espérance et une variance indépendantes du point d'appui x (mais dépendant de h), soit :

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

$$E [(Y(x+h) - Y(x))^2] = 2 \gamma(h)$$

La fonction $m(h)$ est la dérivée linéaire. Pour montrer qu'elle est linéaire en h , on part de la relation évidente :

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

et on passe aux espérances, d'où : $m(h' + h'') = m(h') + m(h'')$. On peut toujours supposer que cette dérivée linéaire $m(h)$ est nulle, quitte à remplacer $Y(x)$ par $Y(x) - m(x)$.

La fonction $\gamma(h)$:

$$(2) \quad \gamma(h) = \frac{1}{2} E [(Y(x+h) - Y(x))^2]$$

s'appelle le demi-variogramme, ou fonction intrinsèque. Une F.A. vérifiant l'hypothèse intrinsèque constitue ce que l'on appelle un schéma intrinsèque, caractérisé par son demi-variogramme.

Remarque. Si $Y(x)$ vérifie l'hypothèse stationnaire d'ordre 2, elle vérifie aussi l'hypothèse intrinsèque, et l'on a dans ce cas :

$$(3) \quad \gamma(h) = K(0) - K(h)$$

Comme $|K(h)| \leq K(0)$, on a $\gamma(h) \leq 2K(0)$, de sorte que le demi-variogramme d'une F.A. stationnaire d'ordre 2 est nécessairement borné.

Il existe des schémas intrinsèques d'utilisation très courante dont le $\gamma(h)$ n'est pas borné, et qui par conséquent ne peuvent pas vérifier l'hypothèse stationnaire d'ordre 2 (ont une variance a priori infinie).

EX; : schéma de De Wijs ($\gamma(h) = 3\alpha \log|h|$), schéma linéaire ($\gamma(h) = A|h|$)

II - Propriétés de la covariance $K(h)$ et du demi-variogramme $\gamma(h)$

a/ Symétrie $K(h) = K(-h)$; $\gamma(h) = \gamma(-h)$

Inégalités $|K(h)| \leq K(0)$; $\gamma(h) \geq 0$ et $\gamma(0) = 0$

Ces conditions sont nécessaires, mais il ne suffit pas qu'une fonction K ou γ les vérifie pour qu'il existe une F.A. stationnaire ou intrinsèque admettant cette covariance ou ce demi-variogramme. En fait, il faut et il suffit que K appartienne à la classe des fonctions "de type positif" et γ à celle des fonctions de type positif conditionnel. Par exemple, les fonctions $\log r$ et r^λ avec $\lambda < 2$ peuvent servir de demi-variogramme, mais non r^λ pour $\lambda \geq 2$. Ces conditions expriment, entre autres, que les formules que nous établirons ci-dessous pour les variances d'extension ou d'estimation conduisent nécessairement à des valeurs positives (ce qui ne serait pas toujours le cas si l'on prenait une fonction quelconque comme demi-variogramme)

b/ On peut remarquer que le variogramme donne un contenu précis à la notion traditionnelle de zone d'influence d'un échantillon : sa croissance plus ou moins rapide reflète, en effet, la manière plus ou moins rapide dont se détériore l'influence d'un échantillon sur des zones de plus en plus lointaines du gisement.

c/ Les anisotropies se manifestent par le comportement différent du variogramme dans les différentes directions de l'espace. En l'absence d'anisotropie, $\gamma(h) = \gamma(r)$ ne dépend que du module r de h , et non de la direction de ce vecteur. On dit qu'il y a anisotropie géométrique lorsqu'une simple transformation linéaire des coordonnées suffit à rétablir l'isotropie.

Il y a des types plus complexes d'anisotropies. Par exemple, dans l'espace à trois dimensions, il peut arriver que $\gamma(x)$ ne dépende que de la troisième coordonnée x , et reste par suite constante dans les plans parallèles aux deux premiers axes de coordonnées. Le demi-variogramme

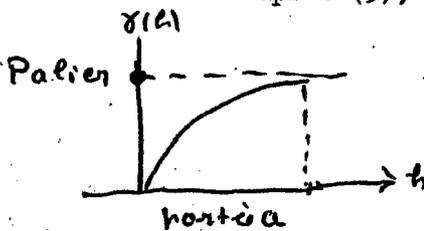
$\gamma(h) = \gamma(h_3)$ ne dépend que de la troisième composante de h . Le plus souvent, $Y(x)$ ne sera pas réellement constante dans les plans horizontaux, mais y variera moins vite ou plus régulièrement que dans la direction verticale.

On prendra alors un demi-variogramme de la forme $\gamma(h) = \gamma_0(h_1 + h_2 + h_3) + \gamma_1(h_3)$

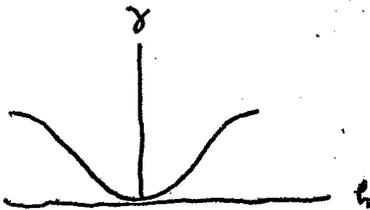
d/ Portée. Dans le cas stationnaire d'ordre 2, la portée $a(\alpha)$ dans une direction α est la valeur de la distance au delà de laquelle, dans cette direction, $Y(x)$ et $Y(x+h)$ sont sans corrélation (ou en corrélation négligeable) : $K(h) = 0$ (ou $\neq 0$) pour $|h| > a(\alpha)$.

D'après (3), $K(h) = 0$ équivaut à $\gamma(h) = K(0) = \gamma(\infty)$:

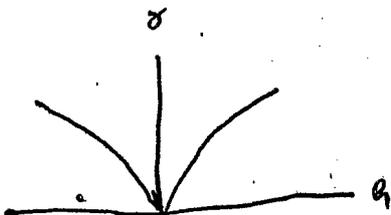
la portée est aussi la distance à partir de laquelle le demi-variogramme atteint sa valeur limite $\gamma(\infty)$ ou palier. Ainsi, une F.A. intrinsèque dont le variogramme n'est pas borné ne peut pas avoir de portée finie.



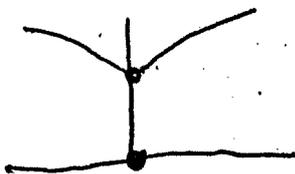
e/ Comportement au voisinage de l'origine. La continuité et la régularité dans l'espace de la F.A. $Y(x)$ s'expriment dans le comportement du $\gamma(h)$ au voisinage de l'origine. Par ordre de régularité décroissante, on peut distinguer quatre types



Comportement parabolique : $\gamma(h)$ est deux fois dérivable en $h = 0$, $Y(x)$ est alors elle-même dérivable (en moyenne quadratique), donc présente un haut degré de régularité dans l'espace.



Comportement linéaire (tangente oblique à l'origine) : $\gamma(h)$ est continu en $h = 0$, mais non dérivable, $Y(x)$ est continue en moyenne quadratique, mais non dérivable, donc déjà moins régulière.



Effet de pépite : $\gamma(h)$ ne tend pas vers 0 lorsque h tend vers 0 (discontinuité à l'origine). $Y(x)$ n'est même pas continue en moyenne quadratique, donc

extraordinairement irrégulière.



Cas limite complètement aléatoire : $Y(x)$ et $Y(x')$ sont indépendantes pour deux points distincts quelconques, si rapprochés soient-ils.

Partie irrégulière. Dans le cas isotrope ($\gamma(h) = \gamma(r)$) et en l'absence d'effet de pépite, on caractérisera le comportement du $\gamma(h)$ par un développement limité de la forme :

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

Exactement comme pour les covariogrammes transitifs, on distinguera une partie régulière (termes de degré entier pair), et une partie irrégulière (termes en r^{λ} avec λ différent d'un entier pair, et aussi termes logarithmiques du type $r^{2n} \log r$). En l'absence de partie irrégulière, la F.A. serait indéfiniment dérivable, donc parfaitement régulière. C'est donc la partie irrégulière seule qui représente le degré d'irrégularité de la F.A., et dans cette partie irrégulière c'est le terme de plus bas degré qui joue le rôle principal : on peut définir le degré de régularité de la F.A. comme le degré λ du terme irrégulier principal. Donnons quelques indications mathématiques complémentaires :

Continuité et dérivation en moyenne quadratique.

On dit que la F.A. $Y(x)$ est continue en moyenne quadratique (continue m.q.) si l'on a :

$$E \left([Y(x+h) - Y(x)]^2 \right) \rightarrow 0 \quad \text{pour } |h| \rightarrow 0$$

Il en est ainsi (par définition) si et seulement si $\gamma(h)$ est continu en $h = 0$, c'est à dire en l'absence d'effet de pépite.

Dans l'espace à une dimension, on dit que la F.A. $Y(x)$ est la dérivée en moyenne quadratique (dérivée m.q.) de la F.A. $Y(x)$ si :

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

On a des définitions analogues dans l'espace à $n \neq 1$ dimensions.

On démontre qu'une F.A. intrinsèque $Y(x)$ admet une dérivée m.q. $Y'(x)$ si et seulement si $\gamma(h)$ est deux fois dérivable en $h = 0$. La dérivée seconde $\gamma''(h)$ existe alors pour tout h , $Y'(x)$ est stationnaire d'ordre 2 (même si $Y(x)$ est seulement intrinsèque) et admet comme covariance la fonction $-\gamma''(h)$. De même, $Y(x)$ est n fois dérivable m.q. si et seulement si la dérivée $\gamma^{(2n)}(h)$ existe en $h = 0$ (elle existe alors pour tout h).

Si λ est le degré du terme irrégulier principal de $\gamma(h)$, $Y(x)$ admet donc une dérivée m.q. d'ordre n si et seulement si $\lambda > 2n$ (si ce terme irrégulier principal est $r^{2n} \log r$, $Y(x)$ est $n-1$ fois m.q. dérivable, mais non n fois).

III - Régularisation d'une F.A. intrinsèque.

Pour simplifier l'exposé, nous ferons les raisonnements dans le cas d'une F.A. stationnaire d'ordre 2 admettant une fonction de covariance $K(h)$, mais tous les résultats que nous exprimerons à l'aide de la fonction intrinsèque $\gamma(h) = K(0) - K(h)$ resteront valables dans le cas d'une F.A. intrinsèque (donc, même si $\gamma(h)$ n'est pas borné, et si par suite la covariance $K(h)$ n'existe pas).

a/ Intégrale stochastique $I = \int_{\sigma} Y(x) p(x) dx$

On définit l'intégrale I comme la limite en moyenne quadratique (si elle existe) des sommes discrètes :

$$I_n = \sum_{i=1}^n \gamma(x_i) f(x_i) \Delta x_i$$

(les Δx_i sont des petits éléments de volume disjoints dont la réunion est v , et x_i est un point de Δx_i). L'intégrale stochastique I est donc une variable aléatoire, comme les I_n .

On démontre que cette V.A. I existe si et seulement si l'intégrale :

$$(4) \quad D^2(I) = \int_{\mathcal{U}} p(x) dx \int_{\mathcal{U}} K(x-y) p(y) dy$$

est finie, et $D^2(I)$ est alors la variance (finie) de cette V.A. On retrouve facilement (4) par le calcul (non rigoureux) suivant :

$$\begin{aligned} I^2 &= \int_{\mathcal{U}} Y(x) p(x) dx \int_{\mathcal{U}} Y(y) p(y) dy \\ E(I^2) &= \int_{\mathcal{U}} p(x) dx \int_{\mathcal{U}} E[Y(x) Y(y)] p(y) dy = \\ &= \int_{\mathcal{U}} p(x) dx \int_{\mathcal{U}} K(x-y) p(y) dy \end{aligned}$$

(le calcul n'est pas rigoureux, car il n'est pas tout à fait évident au départ que l'on a le droit d'intervertir les symboles E et \int : en fait, on montre que cette interversion est bien légitime).

b/ Convolution stochastique. Le produit de convolution $Y * f$ de la F.A. $Y(x)$ par une fonction ordinaire $f(x)$ est l'intégrale stochastique (si elle existe) :

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

On peut ainsi définir la régularisée $Y_h = Y * p_h$ de la F.A. $Y(x)$ par une fonction de pondération $p(x)$. C'est la "moyenne mobile pondérée" :

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

Attention : la régularisée $Y_h(x)$ est encore une fonction aléatoire (plus régulière que $Y(x)$, mais toujours aléatoire) : il ne suffit pas de lisser ou régulariser une F.A. par un procédé de moyenne mobile pour faire disparaître comme par magie le caractère aléatoire de cette fonction.

La variance de la régularisée Y_h est donnée par la formule (4) ci-dessus, et $Y_h(x)$ existe au sens de l'intégration m.q. si et seulement si $D^2(I) < \infty$.
Calculons la covariance $E[Y(x_0) Y(x_0+h)]$ de $Y(x_0)$ et $Y(x_0+h)$. On a :

$$Y_{\mu}(x) Y_{\mu}(x_0 + h) = \iint p(x') p(x'') Y(x_0 + x') Y(x_0 + h + x'') dx' dx''$$

Passons aux espérances, en intervertissant E et \int . Il vient :

$$E[Y_{\mu}(x_0) Y_{\mu}(x_0 + h)] = \iint K(h + x'' - x') p(x') p(x'') dx' dx''$$

Cette covariance ne dépend pas de x_0 , mais seulement de h . Donc, la régularisée Y est stationnaire d'ordre 2, et admet la covariance :

$$(5) \quad K_{\mu}(h) = \int p(x) dx \int K(h + x - y) p(y) dy$$

Cette formule généralise (4). Pour la présenter sous forme plus synthétique, faisons le changement de variable $x = y + z$. Il vient :

$$K_{\mu}(h) = \int K(h + z) dz \int p(y + z) p(y) dy$$

Soit P le covariogramme transitif ($P = p \ast p$) de la fonction de pondération p . On a par définition $P(z) = \int p(y + z) p(y) dy$, et par suite :

$$K_{\mu}(h) = \int K(h + z) P(z) dz$$

c'est à dire, sous forme de produit de convolution (car $P = p \ast p$) :

$$(5') \quad K = K \ast P$$

On obtient la covariance K_{μ} de la régularisée Y_{μ} en régularisant la covariance $K(h)$ de Y par le covariogramme transitif de la fonction de pondération. On comparera ce résultat avec celui que l'on a obtenu dans le cas des méthodes transitives.

Le demi-variogramme γ_{μ} de la régularisée Y_{μ} est $K_{\mu}(0) - K_{\mu}(h)$. En remplaçant $K(h)$ par $K(0) - \gamma(h)$ dans (4) et (5), on constate que $K(0)$ s'élimine, et il reste :

$$(6) \quad \gamma_r(r) = \int \gamma(r+z) P(z) dz - \int \gamma(z) P(z) dz$$

On peut encore écrire :

$$(6') \quad \gamma_r = \gamma * P - A$$

avec une constante A déterminée par la condition $\gamma_r(0) = 0$. Ces relations (6) et (6') restent vraies pour toute F.A. intrinsèque (même si la covariance n'existe pas).

Montée. La montée constitue un cas particulier de la régularisation.

Dans les méthodes transitives, nous pouvions intégrer la V.R. de $-\infty$ à $+\infty$ par rapport à l'une des coordonnées, par exemple x_3 , parce que cette V.R. était nulle à l'extérieur d'un champ borné. Ici, nous ne pouvons plus intégrer que sur une longueur finie P . Nous appellerons donc montée sous puissance constante P l'opération qui nous fait passer de la F.A.

$Y_3(x_1, x_2, x_3)$ définie dans l'espace à trois dimensions à la F.A. $Y_2(x_1, x_2)$ définie dans l'espace à deux dimensions par :

$$Y_2(x_1, x_2) = \int_0^P Y(x_1, x_2, x_3) dx_3$$

Si $Y_3(x)$ est la teneur ponctuelle dans une formation stratiforme de puissance P , $Y_2(x_1, x_2)$ est encore l'accumulation du sondage implanté au point (x_1, x_2) de la surface topographique.

La montée a un effet régularisant. Dans le cas isotrope, on montre que le terme irrégulier principal du demi-variogramme de Y_2 est en $r^{1+\lambda}$ si celui de Y_3 était en r^λ . Plus précisément, on a les règles :

$$\left. \begin{array}{l} r^\lambda \rightarrow A_\lambda \frac{r^{1+\lambda}}{P} \\ \rho \log r \rightarrow \pi \frac{r}{P} \\ r \rightarrow -r^2 \rho \log r \end{array} \right\} \left(A_\lambda = \sqrt{\pi} \operatorname{tg} \frac{\lambda \pi}{2} \frac{\Gamma(1 + \frac{\lambda}{2})}{\Gamma(1 + \frac{1+\lambda}{2})} \right)$$

IV - Variance d'extension et variance d'estimation

Définissons d'abord la notion capitale de variance d'extension. Soit $Y(x)$ une F.A. que nous supposons pour l'instant stationnaire d'ordre 2, et soit $K(h)$ sa covariance. Désignons par $Z(v)$ et $Z(v')$ les "teneurs moyennes" de deux domaines v et v' de l'espace à n dimensions, c'est à dire les intégrales stochastiques :

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

La formule (5) montre que, si v et v' sont bornés, $Z(v)$ et $Z(v')$ ont des variances finies. La première, par exemple, a pour expression :

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

Calculons de même la covariance $\sigma(v, v')$ de $Z(v)$ et $Z(v')$. De :

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

on déduit, en passant aux espérances mathématiques :

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

Nous appellerons variance d'extension de v à v' (ou de v' à v) la variance de l'erreur $Z(v') - Z(v)$ que l'on commet en attribuant à v' la teneur moyenne $Z(v)$ de v . Cette variance d'extension σ_F^2 est égale à :

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

Compte tenu des valeurs calculées plus haut, on trouve donc :

$$\sigma_F^2 = \frac{1}{\sigma^2} \int_{\sigma} dx \int_{\sigma} K(x-y) dy + \frac{1}{\sigma'^2} \int_{\sigma'} dx \int_{\sigma'} K(x-y) dy - \frac{2}{\sigma\sigma'} \int_{\sigma} dx \int_{\sigma'} K(x-y) dy$$

Remplaçons $K(h)$ par $K(0) - \gamma(h)$: on constate que la constante $K(0)$ disparaît de l'expression de σ_E^2 , et il vient la formule fondamentale :

$$(7) \quad \sigma_E^2 = \frac{2}{\sigma^2 \sigma^2} \int_{\sigma^2} dx \int_{\sigma^2} \gamma(x-y) dy - \frac{1}{\sigma^2} \int_{\sigma^2} dx \int_{\sigma^2} \gamma(x-y) dy \\ - \frac{1}{\sigma^2} \int_{\sigma^2} dx \int_{\sigma^2} \gamma(x-y) dy$$

On peut montrer que (7) reste valable pour toute F.A. intrinsèque, donc même si la covariance $K(h)$ n'existe pas.

Variance d'estimation. Supposons maintenant qu'au lieu de connaître la "teneur moyenne" $Z(v')$ de $Y(x)$ dans un volume v' , nous connaissons la teneur moyenne :

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

de N prélèvements effectués aux N points x_i . Z' est une variable aléatoire, dont les caractéristiques se déduisent sans peine de $K(h)$ ou de $\gamma(h)$.

Nous appellerons variance d'estimation σ_N^2 (de v par les N prélèvements effectués aux N points x_i) la variance de la différence $Z(v) - Z'$. Pour obtenir l'expression de σ_N^2 , on doit, à chacune des étapes du raisonnement qui nous a conduit à (7), remplacer les intégrales étendues à v' par des sommes discrètes étendues aux N points x_i . On trouve ainsi la seconde formule fondamentale :

$$(8) \quad \sigma_N^2 = \frac{2}{N \sigma^2} \sum_i \int_{\sigma^2} \gamma(x_i - x) dx - \frac{1}{\sigma^2} \int_{\sigma^2} dx \int_{\sigma^2} \gamma(x-y) dy \\ - \frac{1}{N^2} \sum_i \sum_j \gamma(x_i - x_j)$$

Cette formule, où l'on voit alterner des expressions exactes et approchées de mêmes intégrales, présente une structure remarquable, analogue, quoique plus complexe, à celle de la formule (3) des méthodes transitives. On note, en particulier, que la variance d'estimation est d'autant plus faible :

- que le réseau des prélèvements x_i est plus serré et plus représentatif de la géométrie du volume v que l'on veut estimer,
- que la fonction $\gamma(h)$ est plus régulière, donc que la F.A. $Y(x)$ est elle-même plus continue dans sa variation spatiale.

En pratique, cependant, si N est grand, la formule (8) conduirait à des calculs assez longs. Nous donnerons plus loin des formules d'approximation beaucoup plus simples.

Variance de v dans V

La notion de variance $\sigma^2(v|V)$ d'un échantillon v dans un champ V semble de prime abord expérimentalement évidente. Elle n'a cependant un sens précis que dans les deux cas a/ et b/ ci-après :

a/ v est réduit à un point : la variance $\sigma^2(0|V)$ de la variable ponctuelle $Y(x)$ est l'espérance de $[Y(x) - Z(V)]^2$ lorsque x est tiré au sort dans V avec une densité de probabilité uniforme ($Z(V) = \frac{1}{V} \int_V Y(x) dx$ est la teneur moyenne de V). C'est donc la valeur moyenne en $x \in V$ de la variance d'extension du point x au volume V , qui est, d'après (7) ou (8) :

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

En intégrant en x dans V , le premier terme donne le double de la valeur du second, et il reste :

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

(valeur moyenne dans V de $\gamma(h)$ lorsque les deux extrémités x et y du vecteur h décrivent, chacune pour son propre compte, le volume V)

b/ Le volume V est réunion de N volumes v_i disjoints, égaux à un même volume v de référence (plus précisément, chacun des v_i se déduit de v par translation). Si Z_i est la "teneur moyenne" de $Y(x)$ dans v_i , on pose cette fois :

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

C'est donc la moyenne arithmétique des variances d'extension des v_i dans V . La formule (7), après un calcul facile, conduit à :

$$(9) \quad \sigma^2(v|V) = \frac{1}{V^2} \int_V dx \int_V \gamma(x-y) dy - \frac{1}{v^2} \int_v dx \int_v \gamma(x-y) dy$$

En particulier, on a :

$$(9') \quad \sigma^2(v|V) = \sigma^2(0|V) - \sigma^2(0|v)$$

c/ Dans le cas général où v et V sont des volumes quelconques, non nécessairement géométriquement compatibles, on définit la variance par cette même formule (9) : cette quantité ne représente alors qu'un simple artifice de calcul. Elle peut même prendre des valeurs négatives; en particulier, on a toujours $\sigma^2(v|V) = -\sigma^2(V|v)$

d/ Relation d'additivité. Soient v , V et V' trois volumes, avec par exemple $v \subset V \subset V'$. De (9'), on tire :

$$\begin{aligned} \sigma^2(v|V) &= \sigma^2(0|V) - \sigma^2(0|v) \\ \sigma^2(v|V') &= \sigma^2(0|V') - \sigma^2(0|v) \end{aligned}$$

et par différence :

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

c'est à dire :

$$(10) \quad \sigma^2(v|V') = \sigma^2(v|V) + \sigma^2(V|V')$$

La variance de l'échantillon v dans le champ V est égale à la somme des variances de v dans le panneau V et du panneau V dans le champ V .

e/ Covariance de v et v' dans V . On définit d'une manière analogue la covariance $\sigma(v, v' | V)$ de deux échantillons v et v' (dont la distance et la disposition mutuelle restent fixes) dans le champ V . On trouve :

$$(11) \quad \sigma(v, v' | V) = \frac{1}{V^2} \int_V dx \int_V \delta(x-y) dy - \frac{1}{V} \int_V dx \int_{V'} \delta(x-y) dy$$

f/ En particulier, il est souvent commode d'exprimer la variance d'estimation (8) ou la variance d'extension (7) par l'intermédiaire des variances et covariances des échantillons dans un champ V arbitraire. On trouve, par exemple, identiquement :

$$(12) \quad \sigma_E^2 = \sigma^2(v | V) + \sigma^2(v' | V) - 2 \sigma(v, v' | V)$$

comme on le voit en substituant (9) et (11) : le terme $\frac{1}{V^2} \int_V \int_V$ disparaît, et les trois termes qui subsistent donnent le second membre de (7). On obtient ainsi une expression peut-être plus intuitive de la variance d'extension.

Application : Mailles aléatoires et aléatoires stratifiées.

Nous traiterons plus loin le cas des mailles régulières, qui est de loin le plus difficile, et nous allons examiner ici le cas des deux types usuels de mailles aléatoires.

a/ Maille aléatoire pure. Pour estimer la teneur $Z(V)$ d'un volume V , on dispose des valeurs $Y(x_i)$ de la F.A. en N points x_i implantés "n'importe où" dans V . Nous admettrons que tout se passe comme si chaque x_i avait été implanté au hasard dans V avec une densité de probabilité uniforme et indépendamment des autres prélèvements. On peut obtenir la variance

d'estimation σ_N^2 en intégrant (8) dans V relativement à chacun des x_i .
Ce calcul est facile, mais il est encore plus simple de remarquer que les erreurs partielles $Y(x_i) - Z(V)$ sont indépendantes les unes des autres et admettent la même variance $\sigma^2(0|V)$: l'erreur résultante, qui est

$$\frac{1}{N} \sum [Y(x_i) - Z(V)] \quad \text{admet donc la variance :}$$

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

La variance d'estimation, dans le cas d'une maille aléatoire pure, est donc égale à la variance d'un prélèvement dans le champ V divisée par le nombre N des prélèvements.

b/ Maille aléatoire stratifiée. Le volume V que l'on veut estimer est ici divisé en N zones d'influence égales et disjointes, v_i , et dans chaque v_i on plante un prélèvement en un point x_i choisi au hasard dans la zone d'influence v_i avec une densité de probabilité uniforme, et indépendamment des autres prélèvements. On peut calculer σ_N^2 en intégrant (8) en x_i dans v_i pour chacune des v_i , mais il est plus simple de remarquer que l'erreur totale est $\frac{1}{N} \sum [Y(x_i) - Z(v_i)]$ et que chacune des erreurs partielles $Y(x_i) - Z(v_i)$ est indépendante des autres et admet la variance $\sigma^2(0|v_i)$. On en déduit aussitôt :

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

La variance d'estimation, dans le cas d'une maille aléatoire stratifiée, est donc égale à la variance d'un prélèvement dans sa zone d'influence divisée par le nombre N des prélèvements.

Remarque. Il est clair que la maille aléatoire stratifiée donne toujours de meilleurs résultats que la maille aléatoire pure. D'après (9'), en effet, on a :

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

V - Effet de pépite et phénomènes de transition

Un demi-variogramme de portée fini a caractérisé ce que l'on appelle un phénomène de transition : au delà de la distance a , l'indépendance est atteinte, et la portée a donne l'échelle des structures élémentaires du phénomène régionalisé correspondant. Il y a souvent, d'ailleurs, superposition de plusieurs structures d'échelles bien différentes, emboîtées les unes dans les autres. Le variogramme expérimental montre alors une succession de seuils et de paliers, dont l'analyse permet de reconstituer la hiérarchie de ces "structures gigogne".

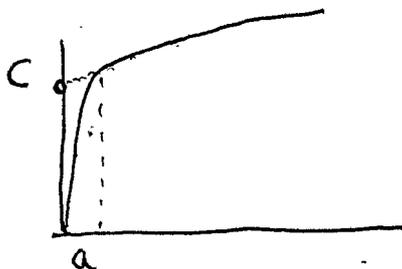
La notion d'échelle joue ici un rôle primordial. A l'échelle de la dizaine ou de la centaine de mètres, un phénomène de transition dont la portée est, par exemple, centimétrique ne se manifeste plus, sur le $\gamma(h)$ expérimental, que comme une discontinuité à l'origine, c'est à dire comme un effet de pépite. D'une manière générale, tout effet de pépite est une réminiscence d'une structure de transition dont les dimensions sont depuis longtemps dépassées à l'échelle à laquelle on travaille : les détails et les caractéristiques qualitatives de cette structure antérieure ont depuis longtemps cessé d'être perceptibles, et l'échelle supérieure n'a guère conservé qu'un paramètre unique - la constante de pépite - qui donne une sorte de mesure globale indifférenciée de l'"intensité" de cette structure dépassée.

Pour analyser la genèse d'un effet de pépite, plaçons nous d'abord au niveau ponctuel, et supposons qu'à une structure primaire de dimension a se superpose une macrorégionalisation, c'est à dire une structure secondaire de dimensions beaucoup plus grandes. Si la structure primaire existait seule, on pourrait décrire la V.R. correspondante comme une réalisation d'une F.A. admettant une covariance $C(h)$ de portée a , ou un demi-variogramme :

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

de portée a et vérifiant $\gamma_1(\infty) = C = C(0)$. Pour tenir compte de la

macrorégionalisation, on doit ajouter une deuxième composante $\gamma_2(h)$, qui représente la structure secondaire, et ne varie qu'avec une extrême lenteur à l'échelle a de la première structure :



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

Ce $\gamma(h)$ présente donc au voisinage de l'origine une zone de croissance très rapide, dont la dimension est de l'ordre de a . A l'échelle de la macrorégionalisation, ce $\gamma(h)$ présentera donc un effet de pépité d'amplitude C .

Examinons les conséquences de cet effet de pépité. Tout d'abord, au niveau macroscopique, les prélèvements ne seront plus ponctuels, mais seront des volumes v déjà grands vis à vis de a . Déterminons donc le demi-variogramme $\gamma_v(h)$ de ces prélèvements v (le seul qui sera expérimentalement accessible). γ_v est la somme de la composante très continue γ_2 (qui n'a pas été sensiblement altérée dans cette régularisation) et de ce que l'on obtient en appliquant la formule (6) à la composante $C - C(h)$: cette composante pépitique est donc :

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

Etudions donc γ_p . On a $\gamma_p(0) = 0$, mais dès que h dépasse les dimensions du volume v , on a $|h+x-y| > a$ si $x \in v$ et $y \in v$, et $C(h+x-y) = 0$. Le premier terme subsiste donc seul dès que h n'est pas très petit, et l'on observe expérimentalement une discontinuité à l'origine, dont la valeur σ_p^2 (variance de pépité) est :

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

D'autre part, les dimensions de v sont grandes, par hypothèse, vis à vis de a , et on a $\int_v C(x-y) dy = \int C(h) dh$ (intégrale étendue à tout l'espace) sauf si x est situé à une distance inférieure à a de la frontière de v . Mais ces points occupent un volume négligeable, et l'on peut aussi

négliger leur influence, qui est de l'ordre de a . Il rest e donc :

$$\sigma_p^2 = \frac{1}{U^2} \int_U dx \int_U C(x) dx = \frac{1}{U} \int_U C(x) dx$$

Ainsi, la constante de pépite σ_p^2 que l'on observe expérimentalement est en raison inverse du volume des prélèvements :

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

et le coefficient $A = \int C(h) dh$, qui est l'intégrale de la covariance $C(h)$ des microstructures, est le seul souvenir de celles-ci qui subsiste à l'échelle des volumes v .

L'effet de pépite majore de la même manière la variance de v dans V , les variances d'extension et les variances d'estimation. Pour $\sigma^2(v|V)$, par exemple, l'apport de l'effet de pépite sera :

$$\frac{1}{U^2} \int_U dx \int_U C(x-y) dy = \frac{1}{V^2} \int_V dx \int_V C(x-y) dy$$

c'est à dire, d'après le calcul précédent :

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

Dans la variance d'estimation de V , par N prélèvements de taille v , on trouverait de même une composante pépitique égale à :

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

Dans tous les cas, la variance due à l'effet de pépite est en raison inverse du volume des prélèvements. Tout se passe donc bien comme si la V.R. elle-même admettait deux composantes indépendantes, l'une très régulière, correspondant au demi-variogramme γ_2 , l'autre complètement aléatoire et discontinue, prenant en charge l'effet de pépite.

VI - Calcul des variances d'estimation.

La formule générale (8) étant trop lourde à manipuler numériquement, nous allons dégager quelques principes d'approximation qui simplifieront le calcul des variances d'estimation (dans le cas des mailles régulières) et permettront de présenter les résultats sous forme d'abaques faciles à consulter.

1°/ Cas de l'espace à une seule dimension.

a/ Les fonctions intrinsèques auxiliaires. Outre $\gamma(h)$, on utilise constamment dans les applications les fonctions suivantes :

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

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

$F(h)$ est la valeur moyenne de $\gamma(h)$ lorsque les deux extrémités de h décrivent le segment $(0, h)$. Ainsi, la variance du segment h dans le segment L est :

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

$\chi(h)$ permet de même le calcul des covariances : la covariance dans L du segment h avec l'une de ses extrémités (ponctuelle) est

$$\sigma(0, h|L) = F(L) - \chi(h)$$

B/ Variance d'extension élémentaire. On appelle ainsi la variance d'extension au segment h de l'échantillon ponctuel implanté au centre de ce segment. Elle est donnée par :

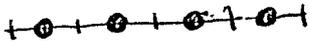
(13)

$$\sigma_E^2 = 2 \chi\left(\frac{h}{2}\right) - F(h)$$



On déduira (13) de (12) et des deux relations établies en a/

c/ Principe de composition des variances d'extension élémentaires.



Soit à estimer la teneur du segment $L = na$ divisé en n zones d'influences de longueur a , au centre de chacune desquelles on a effectué un prélèvement ponctuel. Soit Y_i la teneur du prélèvement i , Z_i celle de sa zone d'influence, et $Z = \frac{1}{n} \sum z_i$ celle de L . L'erreur totale est la moyenne :

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

des erreurs partielles $Y_i - Z_i$. Le principe d'approximation consiste à admettre que ces erreurs partielles sont indépendantes les unes des autres (ce principe est vérifié avec une approximation très raisonnable par les $\gamma(h)$ de type usuel). Comme la variance de $Y_i - Z_i$ est justement la variance d'extension élémentaire calculée en (13), on trouve :

$$(14) \quad \sigma_n^2 = \frac{1}{n} \sigma_E^2 = \frac{1}{n} [2\chi\left(\frac{a}{2}\right) - F(a)]$$

On obtient donc la variance d'estimation en divisant la variance d'extension élémentaire d'un échantillon dans sa zone d'influence par le nombre n des échantillons.

d/ Cas du dispositif fermé. A partir de $n + 1$ prélèvements à maille



régulière a , on veut estimer le segment $L = na$ compris entre le premier et le dernier prélèvement. On montre que ce dispositif fermé est équivalent au dispositif centré examiné

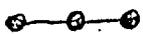
en c/ , et que la variance d'estimation est encore donnée par (14)

(pourvu que n soit supérieur à 1) : le dispositif fermé à $n+1$ échantillons est donc équivalent au dispositif centré à n échantillon.

Remarque. Pour $n = 1$, le dispositif fermé à deux échantillons donne la variance d'estimation :



$$\sigma_{E,1}^2 = 2\chi(a) - F(a) - \frac{1}{2}\gamma(a)$$

qui diffère de la variance d'extension élémentaire (13). On n'a pas le droit de diviser σ_E^2 par n , car les erreurs partielles commises dans l'estimation de deux segments fermés consécutifs  ne sont pas du tout indépendantes (ces deux figures, en particulier, ont en commun l'échantillon central)

2° Cas de l'espace à deux dimensions.

Nous supposons que le demi-variogramme est isotrope (ne dépend que de r)
En cas d'anisotropie géométrique, il est aisé de se ramener à ce cas.

a/ Les fonctions auxiliaires. Il est commode d'introduire les fonctions suivantes :

$\gamma_E(h)$: moyenne de $\gamma(x-y)$ lorsque x et y décrivent les deux cotés parallèles de longueur b du rectangle $b \times h$ (à une constante près, $\gamma_E(h)$ se déduit de $\gamma(h)$ par montée d'ordre 1)



$\chi_E(h)$: moyenne de $\gamma(x-y)$, x décrivant l'un des cotés b du rectangle, et y le rectangle lui-même. On a :

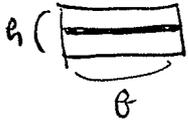
$$\chi_E(h) = \frac{1}{b} \int_0^h \gamma_E(x) dx$$

$F(b, h)$: moyenne de $\gamma(x-y)$ lorsque x et y décrivent le rectangle. Cette fonction est symétrique en b et h , et vérifie :

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

$Q(b, h)$: moyenne de $\gamma(x-y)$, x décrivant un côté b et y un côté h , ou encore, x décrivant le rectangle, y restant fixe en l'un des sommets.

b/ Extension d'un segment médian dans son rectangle d'influence.



$$(15) \quad \sigma_E^2 = 2 \chi_\theta \left(\frac{h}{2} \right) - F(h, \theta) - \gamma_\theta(0)$$

c/ Estimation de S par des traçages parallèles équidistants.

Soient b_1, b_2, \dots, b_n les longueurs des n traçages, h leur équidistance. On assimile la surface S que l'on veut estimer à la réunion des rectangles d'influence des n traçages. La formule

(15) donne l'extension de b_i à son rectangle d'influence, soit

$\sigma_{E_i}^2$. Si Y_i est la teneur de b_i , Z_i celle de sa zone d'influence, on admet que les $(Y_i - Z_i)$ sont indépendantes, et l'erreur totale :

$$\frac{\sum b_i (Y_i - Z_i)}{\sum b_i}$$

admet une variance que l'on peut calculer en pondérant par les carrés des b_i les variances d'extension $\sigma_{E_i}^2$. D'où la variance d'estimation :

$$(14) \quad \sigma_E^2 = \frac{\sum b_i^2 \sigma_{E_i}^2}{(\sum b_i)^2}$$

Remarque. Si les b sont égaux, il reste simplement $\sigma_E^2 = \frac{1}{n} \sigma_{E_i}^2$

d/ Composition d'un terme de ligne et d'un terme de tranche.

Dans le cas c/ précédent, il arrive que l'on ne connaît pas les teneurs réelles des traçages, mais seulement une estimation de ces teneurs à partir de prélèvements ponctuels à maille régulière a ($a \ll h$).

On admet alors que les erreurs commises en estimant les lignes à partir des prélèvements et S elle-même à partir des lignes (supposées connues) peuvent être regardées comme indépendantes. La variance d'estimation est alors :

$$(15) \quad \sigma_E^2 = \frac{1}{N} \sigma^2(a) + \frac{\sum b_i^2 \sigma_{E_i}^2}{(\sum b_i)^2}$$

$\sigma^2(a)$ est la variance d'extension élémentaire (13) d'un prélèvement ponctuel dans son segment a d'influence, N est le nombre total des prélèvements. $\frac{1}{N} \sigma^2(a)$ est le terme de ligne, variance de l'erreur commise en estimant les lignes à partir des prélèvements.

Le deuxième terme, qui figure déjà en (14) est le terme de tranche : variance de l'erreur commise en étendant les teneurs des lignes à leurs tranches d'influence.

Ce principe (15) de composition est valable pourvu que a soit inférieur à h. Il s'applique, en particulier, au cas d'une reconnaissance à maille rectangulaire a h.

e/ Cas d'une maille carrée. La variance d'extension d'un sondage à son carré d'influence, au centre duquel il est implanté, est :



$$(16) \quad \sigma_E^2 = 2 \rho\left(\frac{a}{2}, \frac{a}{2}\right) - F(a, a)$$

Pour une maille carrée, on peut admettre que les erreurs commises en estimant chaque carré à partir de son sondage central sont indépendantes. La variance d'estimation par N sondages est alors :

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

avec un σ_E^2 donné par (16)

f/ Abaques. Pour chaque schéma isotrope, c'est à dire pour chaque fonction $\gamma(r)$, on doit présenter sous forme d'abaques :

1/ La fonction $F(a, b)$ définie en a/, qui sert au calcul de la variance d'un échantillon dans le rectangle a, b.

2/ La variance d'extension (15) d'un traçage médian de longueur b dans le rectangle b, h - avec la cas particulier b = 0, où l'on retrouve la variance d'extension élémentaire (13)

3/ La variance d'extension (16) d'un sondage dans son carré a, a d'influence.

Ces trois documents permettent le calcul rapide de toutes les variances d'estimation.

32/ Cas de l'espace à trois dimension.

On procédera exactement comme dans le cas de l'espace à deux dimensions en composant cette fois trois sortes de termes :

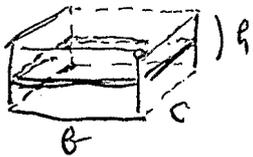
- terme de ligne : extension des prélèvements ponctuels dans les lignes

- terme de section : extension des lignes dans les sections

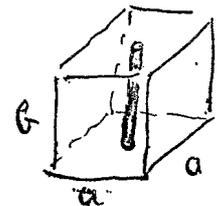
terme de tranche : extension des sections dans les tranches.

X

On devra disposer des abaques précédents (en plus des précédents) :



- variance d'extension d'un plan b, c médian dans son parallélépipède rectangle d'influence



- extension d'un sondage de longueur b dans son prisme droit a, a, b d'influence

- moyenne de $\gamma(h)$ dans le parallélépipède rectangle a, b, c : cette fonction $F(a, b, c)$ est destinée au calcul de la variance d'un échantillon dans ce parallélépipède.

VII - LE SCHEMA DE DE WIJS

Le schéma de de Wijs est défini par la fonction intrinsèque (isotrope)

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

Le coefficient α est la dispersion absolue : elle caractérise la dispersion du phénomène à l'état pur, c'est à dire indépendamment de toute influence de la géométrie du champ ou du support.

1°/ Equivalents linéaires.

Posons :

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

(valeur moyenne de $\log|h|$ dans le volume v : c'est une fonction qui dépend non seulement de la mesure du volume v , mais aussi de sa forme).

D'après la relation (9), on a :

$$(16) \quad \sigma^2(v|v) = 3\alpha [F(v) - F(v)]$$

Si deux échantillons v et v' vérifient $F(v) = F(v')$, ils ont même variance dans n'importe quel champ. Nous dirons qu'ils sont équivalents.

Si nous prenons comme ensemble v un segment de droite de longueur l , nous avons :

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

(le démontrer à l'aide des fonctions auxiliaires de VI, 1°/, a/). Nous appellerons équivalent linéaire d'un échantillon v la longueur du segment l équivalent à v , c'est à dire la longueur l définie par :

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

Soit, de même, L l'équivalent linéaire du champ V :

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

Par différence, il vient :

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

et (16) donne alors :

$$(17) \quad \sigma^2(v|V) = 3\alpha \log\left(\frac{L}{p}\right)$$

Cette formule permet un calcul rapide de la variance. On l'utilise souvent en sens inverse, pour évaluer la dispersion absolue à partir de la variance expérimentale des échantillons dans leur champ :

Si v et V sont géométriquement semblables, on a $V/v = (L/p)^3$, et (17) se réduit à la formule de de Wijs : $\sigma^2(v|V) = \alpha \log(V/v)$

Calcul approché des équivalents linéaires. Avec une excellente précision, l'équivalent linéaire du rectangle de cotés a et b est $a+b$.

Pour le parallélépipède rectangle (a, b, c) , c désignant le plus petit côté, on a, avec une approximation cette fois assez grossière, un équivalent linéaire de $a + b + 0,7 c$. Pour plus de précision, on consultera l'abaque XI - 2.

Application : comportement du demi-variogramme des échantillons v pour h grand devant les dimensions de v . (On n'utilise jamais le schéma de de Wijs au niveau ponctuel, car $\log r$ est infini en $r = 0$, mais toujours ses régularisées). Soit γ_v le demi-variogramme d'échantillons v distants de h . $2\gamma_v(h)$ est la variance d'extension de ces deux échantillons, et la formule (7) nous donne :

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

Dès que h dépasse les dimensions de v , le premier terme ne diffère plus beaucoup de $3\alpha \log r$. Le deuxième terme est $3\alpha(\log P - 3/2)$, P désignant l'équivalent linéaire de v . Pour h assez grand devant v , on a donc :

$$(18) \quad \gamma_r(r) = 3\alpha \log \frac{r}{P} + \frac{9\alpha}{2}$$

Cette formule est précieuse dans les applications expérimentales.

2°/ Le schéma de de Wijs à deux dimensions.

Dans l'application précédente, on peut prendre pour v un segment de longueur P se déplaçant parallèlement à lui-même. D'après (18), on a :



$$\gamma_p(h) = 3\alpha \log \frac{r}{P} + \frac{9\alpha}{2} \quad \text{pour } h \gg P$$

Pour h petit, au contraire, la règle de montée $\log r \rightarrow \pi r/P$ donne

$$(19) \quad \gamma_p(h) = 3\alpha \pi \frac{h}{P}$$

A partir de l'expression rigoureuse de $\gamma_p(h)$ (dédue de $\log r$ par montée) on calcule la variance d'extension d'un segment P dans son rectangle P, h d'influence : Abaque XI-5. Lorsque h est petit devant P (en pratique, il suffit que h soit inférieur ou égal à P), on peut utiliser (19) pour calculer directement cette variance d'extension, ce qui donne :

$$(20) \quad \sigma_E^2 = \alpha \frac{\pi}{2} \frac{h}{P}$$

Application : Gisement stratiforme reconnu par traçages parallèles.



Q/ Supposons d'abord que les teneurs réelles des traçages soient connues, et calculons le terme de tranche. Si h est l'équidistance des traçages, on lit sur XI-5

la variance d'extension σ_E^2 du traçage de longueur l_i dans sa tranche d'influence. Le terme de tranche est alors :

$$\sigma_E^2 = \frac{\sum l_i^2 \sigma_{Ei}^2}{(\sum l_i)^2}$$

Si chacun des traçages a une longueur l_i supérieure à h , σ_{Ei}^2 est donnée par (20), et, en posant $L = \sum l_i$ (longueur tracée totale), on trouve :

$$\sigma_E^2 = \alpha \frac{\pi}{2} \frac{\sum l_i^2 \frac{h}{l_i}}{L^2} = \alpha \frac{\pi}{2} \frac{h}{L}$$

ou encore, avec $S = L h$ (surface minéralisée totale) :

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

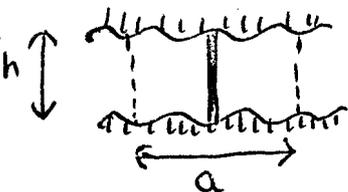
X

Conséquence : un gisement ^{quelque} deux fois plus gros, toutes choses égales d'ailleurs, nécessite, à précision égale, deux fois moins de dépenses de reconnaissance à la tonne de minerai.

4/ Si les teneurs des traçages sont estimées à partir de rainurages implantés à maille régulière a , on doit ajouter un terme de ^{ligne} ~~tranche~~ Pour le calculer, il n'est plus possible de négliger la puissance p de la formation (p est petit, par hypothèse, devant h où les longueurs l_i des traçages, mais non devant a). C'est à nouveau l'abaque XI-5 qui donne

la variance d'extension $\sigma^2(a)$ de la rainure de longueur p dans son rectangle d'influence a, p . Le terme de ligne est alors :

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



Remarque. Dans la plupart des gisements stratiformes ou filoniens, la troisième dimension, celle des puissances, n'est pas équivalente aux deux autres. Il y a anisotropie. La longueur p que l'on doit attribuer au rainurage peut donc différer de la puissance réelle. On détermine cette puissance équivalente p en remarquant, d'après (18), que le demi-variogramme expérimental des rainurages a pour équation :

$$X \quad \gamma(h) = 3\alpha \left[\log \left(\frac{h}{p} \right) + \frac{3}{2} \right] \quad \text{pour } h > p$$

C'est la valeur de p ainsi déterminée qui doit servir au calcul de $\sigma^2(a)$.

Couche mince reconnue par sondages à maille carrée.

Soit p la puissance équivalente de la formation, définie comme dans la remarque qui précède. On doit calculer la variance d'extension d'un sondage dans son carré (a, a) d'influence. En assimilant ce carré au cercle de même surface, on trouve (cf. Exercice 9) :

$$\sigma_E^2 = 3\alpha \left(\log \frac{2a}{p} - \frac{1}{2} \right)$$

La variance d'estimation s'en déduit en divisant par le nombre n des sondages ;

$$\sigma_n^2 = \frac{3\alpha}{n} \left(\log \frac{2a}{p} - \frac{1}{2} \right)$$

3°/ Le schéma de de Wijs dans l'espace à trois dimensions

L'abaque XI-2 donne l'équivalent linéaire du parallélépipède rectangle.

Estimation des amas. On calculera le terme de ligne et le terme de section comme ci-dessus (à partir de l'abaque XI-5). Pour le terme de tranche on utilisera, selon le cas, l'abaque XI-7, XI-8, ou XI-9 (variance d'extension des grandes sections ab , des sections moyennes ac et des petites sections bc , $a \geq b \geq c$).

Amas reconnu par sondages à maille carrée.

On lit sur l'abaque XI-6 la variance d'extension σ_E^2 d'un sondage de puissance h ; dans son prisme droit h_i, a, a d'influence. On a ensuite :

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

(si les h_i sont égaux ou peu différents, : $\frac{1}{n} \sigma_{E_i}^2$)

VII - LE SCHEMA SPHERIQUE

Pour représenter un phénomène de transition, on peut utiliser des schémas de la forme :

$$\gamma(h) = A [K(0) - K(h)]$$

où $K(h)$ est le covariogramme géométrique d'un volume v (cf exercice 7).

Si l'on prend pour v la sphère de diamètre a , on a (cf. ex. 4 des méthodes transitives) :

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

Le schéma sphérique sera donc défini par le demi-variogramme :

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

La portée est a , le palier est $C = \gamma(\infty)$, la pente à l'origine $\frac{3}{2} \frac{C}{a}$

Dans les applications, on déterminera a et C par approximations successives : on part de valeurs a_1 et C_1 , obtenues par simple interpolation sur le demi-variogramme expérimental; on utilise l'abaque XIV-2 pour calculer la variance théorique d'un échantillon dans son champ; si celle-ci est,

par exemple, plus forte que la variance expérimentale des échantillons dans leur champ, on peut soit diminuer C_1 , soit augmenter a_1 . On arrive assez vite à l'ajustement cherché.

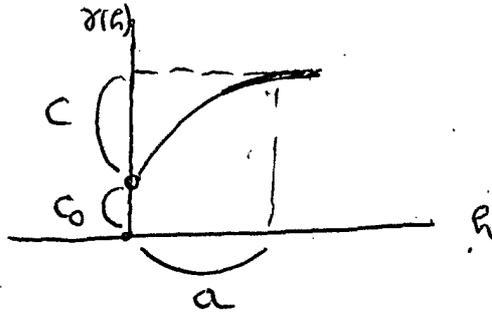
Nous ne disposons que des abaques relatifs aux problèmes à une et à deux dimensions :

Abaque XIV-2 : variance d'un point dans un rectangle

Abaque XIV-4 : variance d'extension élémentaire d'un point dans son segment d'influence (pour le calcul du terme de ligne); variance d'extension d'un point dans son carré d'influence (pour les sondages à maille carrée)

Abaque XIV-3 : variance d'extension d'un segment médian dans son rectangle d'influence (pour le calcul des termes de tranche).

Remarque. Dans les applications, les phénomènes transitifs sont très souvent accompagnés d'un effet de pépité : on les représentera donc par un schéma sphérique à effet de pépité C_0 :



Ecole d'été de Fontainebleau - 1968 - GéostatistiqueExercicesExercices sur les méthodes transitives

Exercice 1. (covariogramme exponentiel) . Soit, dans l'espace à une seule dimension, une V.R. dont le covariogramme transitif est $g(h) = e^{-\lambda|h|}$. Former l'expression exacte de la variance d'estimation $\sigma^2(a)$, trouver sa partie principale, et conclure à l'absence de Zitterbewegung.

(solution : $\sigma^2(a) = a \left(\frac{2}{1-e^{-2a}} - 1 - \frac{2}{\lambda a} \right) \approx \frac{1}{6} \lambda a^2$.)

Exercice 2. (Covariogramme triangulaire)

Soit dans l'espace à une dimension la V.R. $f(x)$ égale à 1 sur l'intervalle $(0, b)$.

1^o/ Former le covariogramme de $f(x)$. (sol. : $g(h) = b - |h|$ pour $|h| \leq b$ et 0 pour $|h| > b$; établir ce résultat par un raisonnement géométrique)

2^o/ Calculer la variance d'estimation pour une maille a petite, par la formule d'approximation (sol. : $1/6 a^2$)

3^o/ On dispose de $n+2$ sondages $S_0, S_1, \dots, S_n, S_{n+1}$ à maille a régulière, dont le premier et le dernier sont négatifs et les autres positifs. On admet que chacune des deux extrémités de l'intervalle dont on veut estimer la longueur b peut tomber n'importe où sur le segment (S_0, S_1) et (S_n, S_{n+1}) respectivement, indépendamment l'une de l'autre et avec une probabilité uniforme. La longueur b devient ainsi une variable aléatoire. Montrer $E(b) = na$; $D^2(b) = a^2/6$: on retrouve bien ainsi la variance d'estimation telle que la donne la formule d'approximation, mais on n'atteint pas le terme fluctuant.

4^o/ Calculer la valeur exacte de la variance d'estimation $\sigma^2(a)$ du 2^o/ (sol. : poser $b = na + \xi a$, avec $0 \leq \xi < 1$. Si le premier prélèvement positif est en x ($0 \leq x < a$), on a l'estimation $(n+1)a$ pour $x \leq \xi a$, et na pour $\xi a < x < a$. Le point x étant implanté au hasard sur le segment

$(0, a)$ avec une densité uniforme de probabilité, cette estimation est une variable aléatoire d'espérance b et de variance $(\xi - \xi^2) a^2$. Cette expression tient compte du Zitterbewegung. Lorsque ξ (qui est toujours inconnu dans la pratique) est assimilé à une variable aléatoire uniformément distribué sur $(0, 1)$, la valeur moyenne de cette expression coïncide avec l'expression approchée du 2^o/ (Zitterbewegung exclu)

Exercice 3 (variance d'estimation dans le cas aléatoire)

Soit, dans l'espace à deux dimensions, une V.R. $f(x)$ de covariogramme $g(h)$ et de support S . Si le rectangle de maille (a_1, a_2) est grand devant S un prélèvement au plus est positif, et tout se passe comme si ce prélèvement unique était implanté au hasard dans un rectangle de côtés a_1, a_2 contenant S . Si ce prélèvement tombe en un point $x \in S$, on prend l'estimation $Q^*(x) = a_1 a_2 f(x)$, et 0 si $x \notin S$. $Q^*(x)$ est ainsi une variable aléatoire. Montrer directement $E(Q^*) = \int f(x) dx = Q$, $E[(Q^*)^2] = \left[\int [f(x)]^2 dx \right] a_1 a_2$ et $D^2(Q^*) = a_1 a_2 g(0) - \int g(h) dh$

Exercice 4. (covariogramme transitif de la sphère)

Dans l'espace à deux dimensions, calculer le covariogramme de la sphère de diamètre a .

(Sol. : remarquer que $-g'(h)$ est l'aire du cercle d'intersection de la sphère et de sa translatée par h , intégrer, en déduire $g(h) = 0$ pour $|h| \geq a$

et $g(h) = \frac{\pi}{6} a^3 \left[1 - \frac{3}{2} \frac{h}{a} + \frac{1}{2} \frac{h^3}{a^3} \right]$ pour $|h| < a$.

Application à l'estimation du volume de cette sphère.

Exercices sur les F.A. intrinsèques

Exercice 1. Une origine x_0 étant choisie au hasard sur $(0, a)$, on divise la droite en segments de longueur a au moyen des points de subdivision $x + k a$ (k entier, positif ou négatif). Soit $Y(x)$ la F.A. prenant sur chacun de ces segments de longueur a une valeur aléatoire constante Y , tirée au sort indépendamment d'un segment à l'autre selon une même loi de moyenne m et de variance σ^2 . Montrer que la probabilité pour que deux points x et $x+h$ appartienne au même segment a est $1 - |h|/a$ pour $|h| \leq a$ et 0 pour $|h| > a$. En déduire le demi-variogramme de $Y(x)$ ($\frac{|h|}{a} \sigma^2$ pour $|h| \leq a$ et σ^2 pour $|h| > a$)

Exercice 2. Même question que dans l'exercice précédent, mais les longueurs des segments successifs sont aléatoires, indépendantes les unes des autres, et obéissent à la même loi exponentielle $e^{-\lambda r}$
(Sol.: la probabilité pour que x et $x+h$ soit dans le même segment est $e^{-\lambda h}$
on en déduit $\gamma(h) = (1 - e^{-\lambda h}) \sigma^2$ -

Exercice 3. On considère le demi-variogramme h^λ ($0 < \lambda < 2$) dans l'espace à une dimension. Calculer les fonctions auxiliaires X et F ; la variance d'extension élémentaire d'un point central à son segment d'influence, la variance d'extension à ce segment de ses deux extrémités (dispositif fermé). Comparer et discuter.

$X(h) = \frac{h^\lambda}{\lambda+1}$; $F(h) = \frac{2 a^\lambda}{(\lambda+1)(\lambda+2)}$; $\frac{2 a^\lambda}{\lambda+1} \left[\frac{1}{2^\lambda} - \frac{1}{\lambda+2} \right]$; $\left(\frac{2}{\lambda+2} - \frac{1}{2} \right) a^\lambda$
On a $\sigma_F^2 > \sigma_C^2$ pour $\lambda > 1$ et inversement, avec égalité pour $\lambda = 1$:
pour $\lambda > 1$, haute continuité, et l'échantillon unique mais bien placé est supérieur aux deux échantillons mal placés; pour $\lambda < 1$, on est proche du cas aléatoire pur, et les deux échantillons, même mal placés, donnent toujours un meilleur résultat que l'échantillon unique; pour $\lambda = 1$ les deux dispositifs sont équivalents)

Exercice 4. Avec le demi-variogramme h^λ dans l'espace à une dimension, calculer la variance d'estimation du segment $L = n a$ par n échantillons ponctuels, dans les trois cas suivants :

maille régulière (dispositif centré) $\left(\frac{1}{n} \frac{2}{\lambda+1} \left[\frac{1}{2^\lambda} - \frac{1}{\lambda+2} \right] a^\lambda \right)$
 maille aléatoire stratifiée $\left(\frac{1}{n} \frac{2a^\lambda}{(\lambda+1)(\lambda+2)} \right)$
 maille aléatoire pure $\left(\frac{1}{n} \frac{2}{(\lambda+1)(\lambda+2)} L^\lambda \right)$

Exercice 5. Soit, dans l'espace à deux dimensions, le demi-variogramme r^λ . Soit un gisement reconnu par traçages de longueurs supérieures à leur équidistance h .

a/ Calculer la variance d'extension d'un traçage de longueur P dans son rectangle P, h d'influence

(la montée sous puissance P transforme r^λ en $A r^{\lambda+1}/P$ (ou $\log r$ en $\pi r/P$)
 d'où $\sigma_E^2 = A \frac{2}{\lambda+2} \left[\frac{1}{2^{1+\lambda}} - \frac{1}{\lambda+3} \right] \frac{P^{1+\lambda}}{P} = B \frac{P^{1+\lambda}}{P}$

b/ En déduire la variance d'estimation (pondérer par les carrés des longueurs des traçages, poser $L = \sum P_i$ et $S = L h$, et mettre la variance d'estimation sous la forme $B \frac{S^{1+\lambda}}{L^{2+\lambda}}$ Cas de Wijsien : $A = \pi$, $B = \frac{\pi}{6}$, $\frac{\pi}{6} \frac{S}{L^2}$

Exercice 6. Même problème qu'en 5, mais on suppose de plus que les traçages sont estimés à partir de rainurages à maille a . Calculer le terme de ligne $\left(\frac{C}{L} a^{1+\lambda} \text{ avec } C = \frac{2}{\lambda+1} \left[\frac{1}{2^\lambda} - \frac{1}{\lambda+2} \right] \right)$

Optimisation économique : soit w le prix du mètre de galerie, p_0 le prix d'un prélèvement, et $l_0 = p_0/w$. Optimiser la précision à prix de revient donné, ou, ce qui revient au même, le prix de revient pour une précision donnée.

(Tout revient à minimiser $C h a^{1+\lambda} + B h^{2+\lambda}$ sous la condition $\frac{1}{a l_0} + \frac{1}{a a} = C h$
 On obtient la relation suivante $B(2+\lambda) a^{1+\lambda} = \lambda C a^{1+\lambda} + C(1+\lambda) \frac{a^{2+\lambda}}{l_0}$
 qui donne l'équidistance h des niveaux en fonction de la maille a des rainurages.

Exercice 9 Soit dans le plan un schéma de de Wijs $\gamma(r) = \log r$. On rappelle que si g est une fonction harmonique à l'intérieur d'un cercle sa valeur moyenne dans le cercle (ou sur sa circonférence) est égale à sa valeur au centre : $\log r$ est une fonction harmonique sauf en $r = 0$.

a/ On désigne par $C(R)$, $\chi(R)$, $F(R)$, $C_0(R)$ la valeur moyenne de $\log r$ sur la circonférence du cercle de rayon R , entre ce cercle et sa circonférence, dans ce cercle, entre ce cercle et son centre. Montrer

$$C(R) = \chi(R) = \log R \quad ; \quad F(R) = \log R - 1/4 \quad ; \quad C_0(R) = \log R - 1/2$$

(les deux premières relations résultent du théorème sur les fonctions harmoniques. Pour calculer $F(R)$, faire varier $R^4 F(R)$ et montrer

$$\frac{d}{dR} R^4 F(R) = 4 R^3 \chi(R) \quad . \quad \text{Pour } C_0(R), \text{ intégrer } 2\pi r \chi(r).$$

b/ Calculer la variance d'estimation d'un panneau circulaire à partir de sa circonférence ($\sigma_F^2 = 2\chi(R) - F(R) - C(R) = \frac{1}{4}$)

c/ Variance d'extension dans le cercle d'un petit élément central dont l'équivalent linéaire est ρ ($\log R/\rho + 3/4$). Si L est l'équivalent linéaire du cercle, cette variance d'estimation est encore $\log(L/\rho) - 1/2$. En déduire que la variance d'extension d'un échantillon équivalent à ρ dans le carré (a, a) au centre duquel il est implanté est approximativement $\log\left(\frac{2a}{\rho}\right) - 1/2$.

Exercice 10. Schéma sphérique à une dimension Calculer les fonctions auxiliaires du schéma sphérique :

$$\begin{array}{ll} \gamma(\rho) = \frac{3}{2} \frac{\rho}{a} - \frac{1}{2} \frac{\rho^3}{a^3} & \text{pour } \rho \leq a, \quad - \quad \text{pour } \rho \geq a \\ \chi(\rho) = \frac{3}{4} \frac{\rho}{a} - \frac{1}{8} \frac{\rho^3}{a^3} & \text{"} \quad - \quad \frac{3}{8} \frac{a}{\rho} \\ F(\rho) = \frac{1}{2} \frac{\rho}{a} - \frac{1}{20} \frac{\rho^3}{a^3} & \text{"} \quad - \quad \frac{3}{4} \frac{a}{\rho} + \frac{1}{5} \frac{a^2}{\rho^2} \end{array}$$

b/ Variance d'extension élémentaire d'un échantillon dans son segment b d'influence pour $b \leq a$ et $b \geq 2a$. Interpréter.

$$\left(\frac{1}{4} \frac{b}{a} + \frac{3}{160} \frac{b^3}{a^3} \quad \text{et} \quad - \frac{3}{4} \frac{a}{b} - \frac{1}{5} \frac{a^2}{b^2} \right)$$

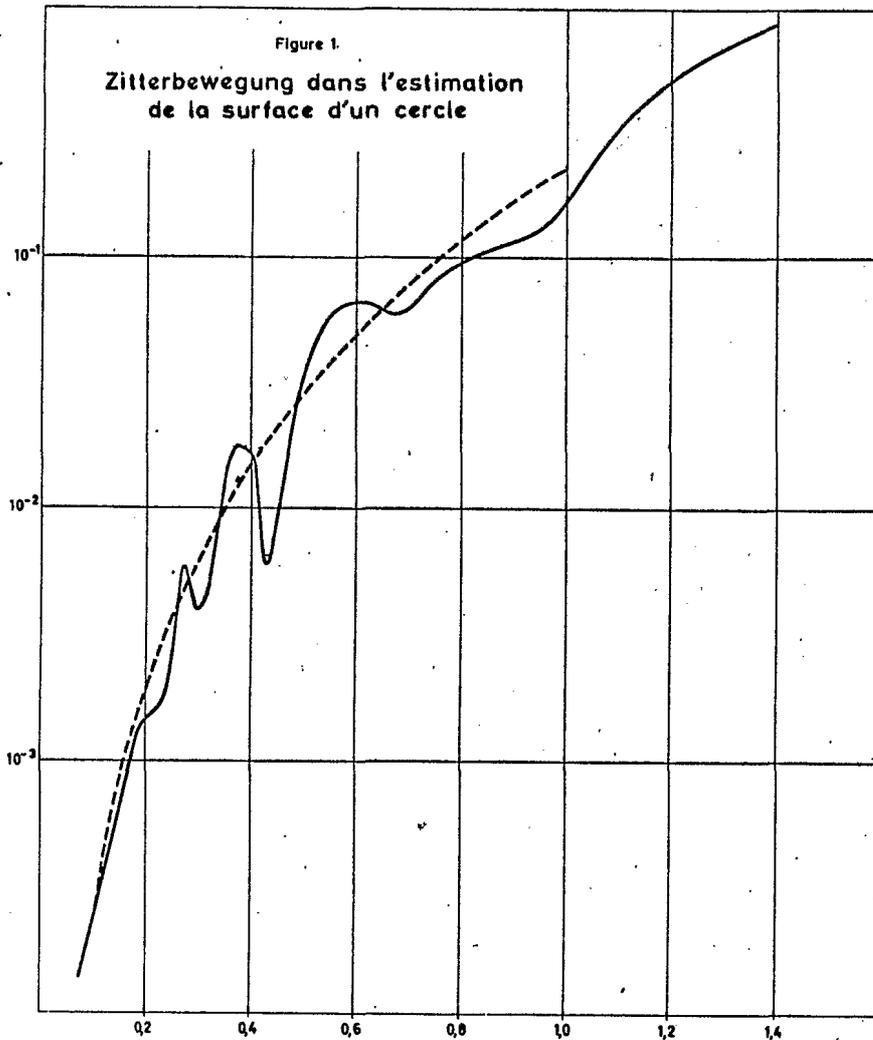


FIG. 1. — Zitterbewegung dans l'estimation de la surface du cercle de diamètre unité à l'aide d'un réseau à maille Carrée. En abscisse la maille a . En ordonnée, la variance d'estimation correspondante $\sigma^2(a)$. La courbe en trait plein représente la valeur exacte, calculée à partir de la formule exacte (IV,1,12). La courbe en pointillé représente la formule $\sigma^2(a) = 0,2276 a^2 + 0,00477 a^4$ obtenue en négligeant le Zitterbewegung et en retenant les deux premiers termes du développement limité donné par le principe de correspondance.

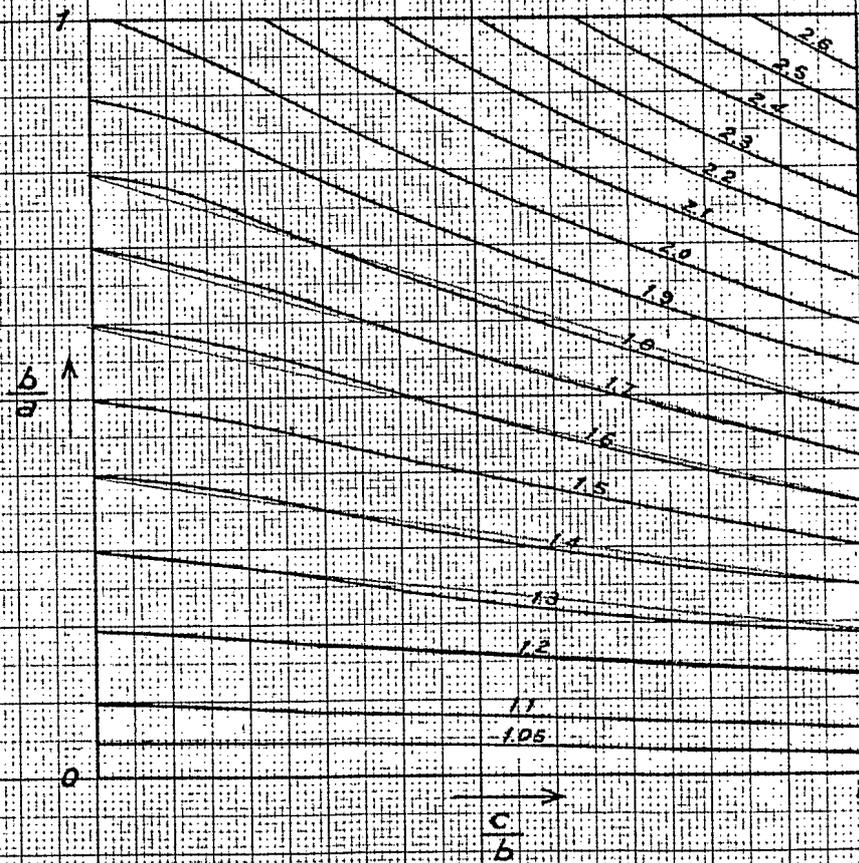
Figure 2

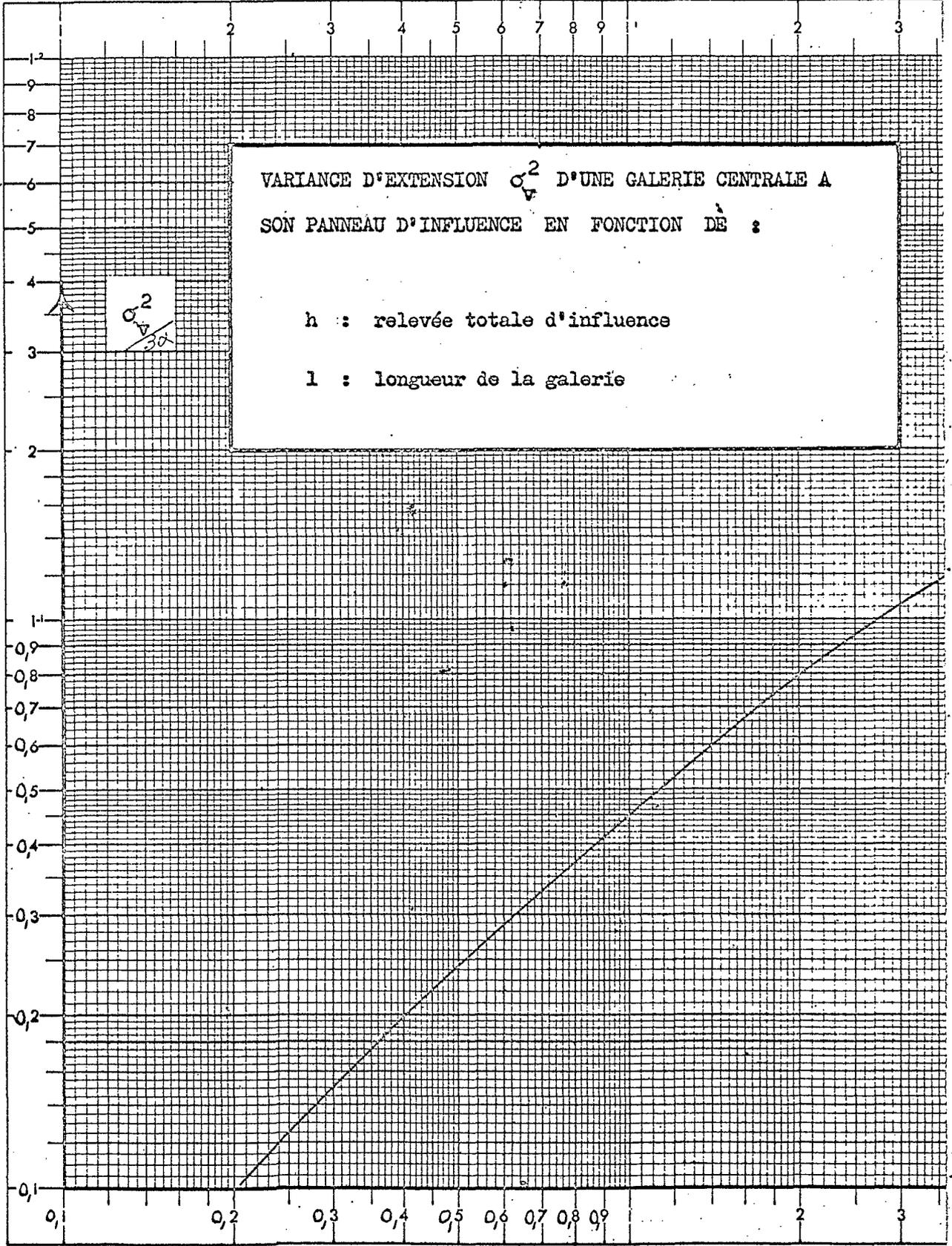
$G(\lambda)$

- Abaque d'équivalence -

Le parallélépipède $a \times b \times c$ ($a \gg b \gg c$) est équivalent
à l'échantillon linéaire λa

Les courbes sont graduées en λ

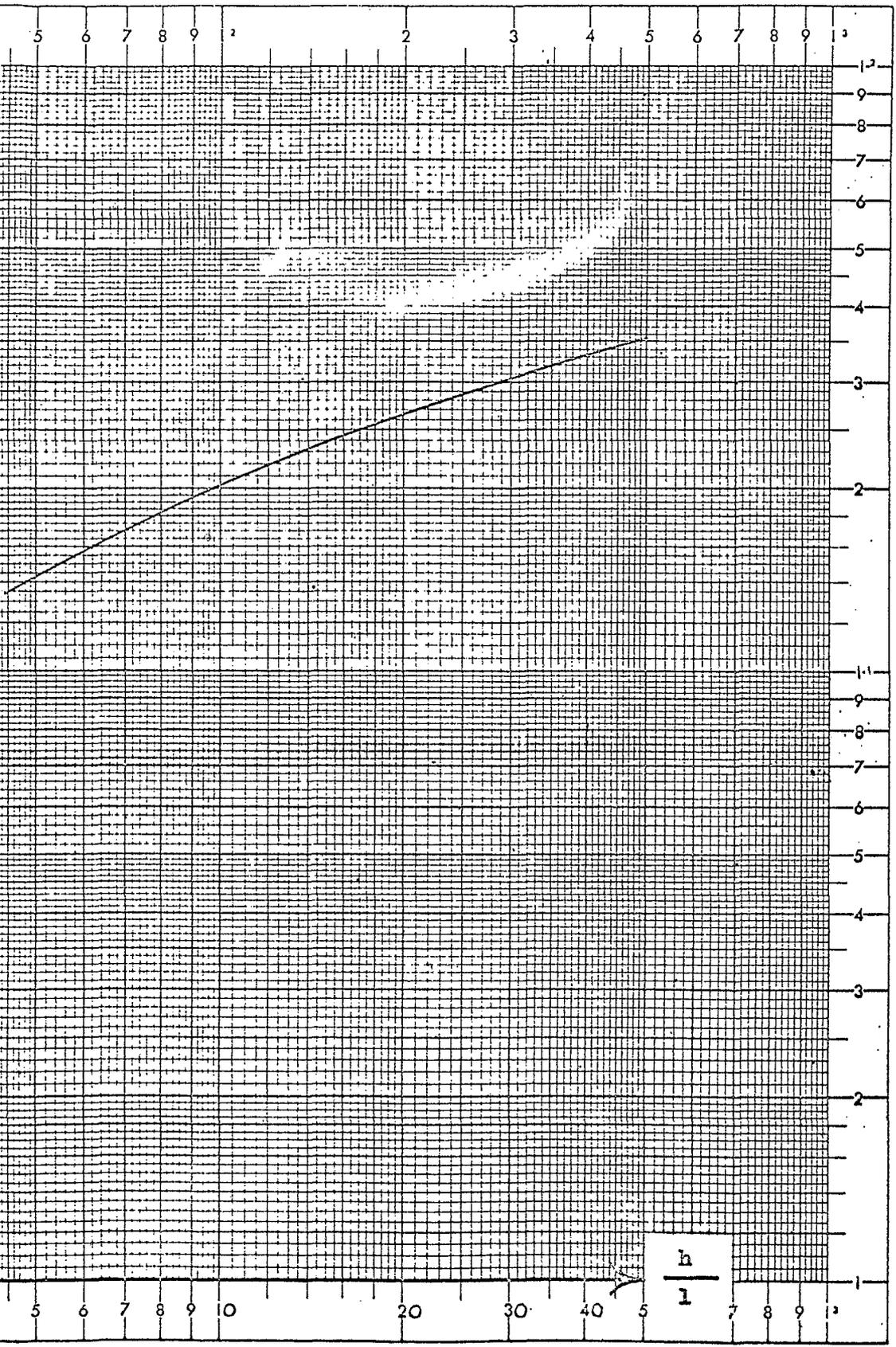




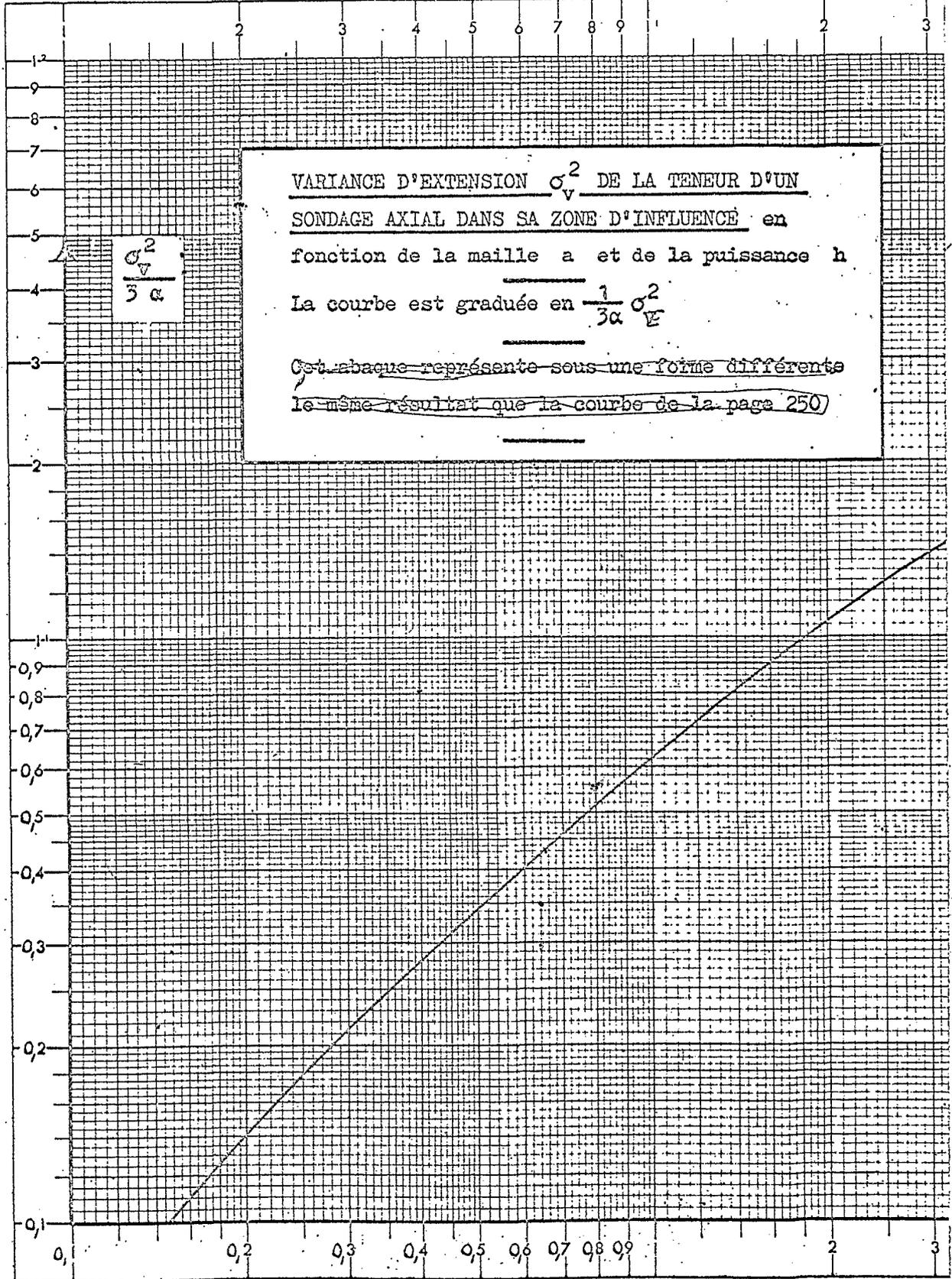
XI-5

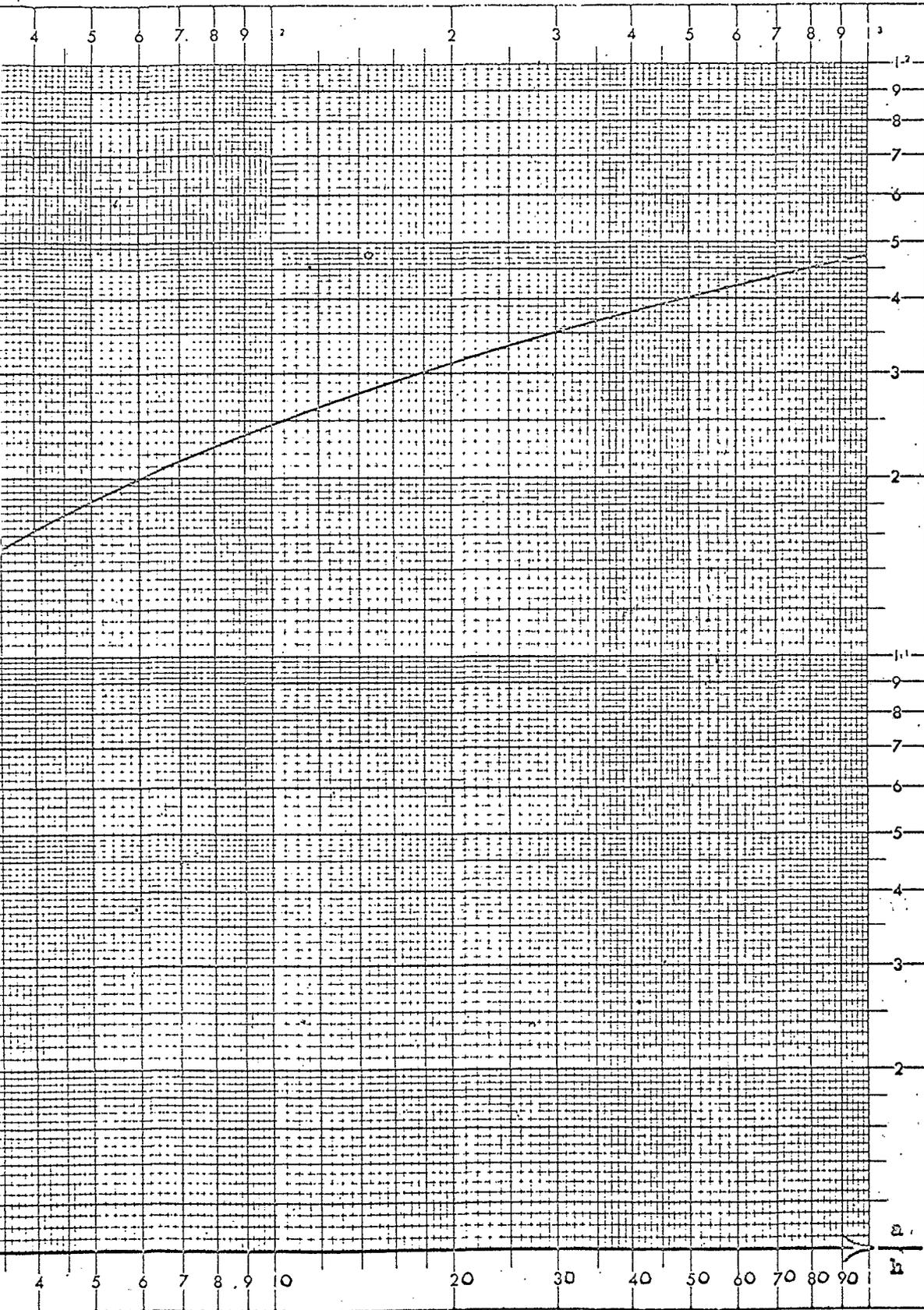
(60)

295



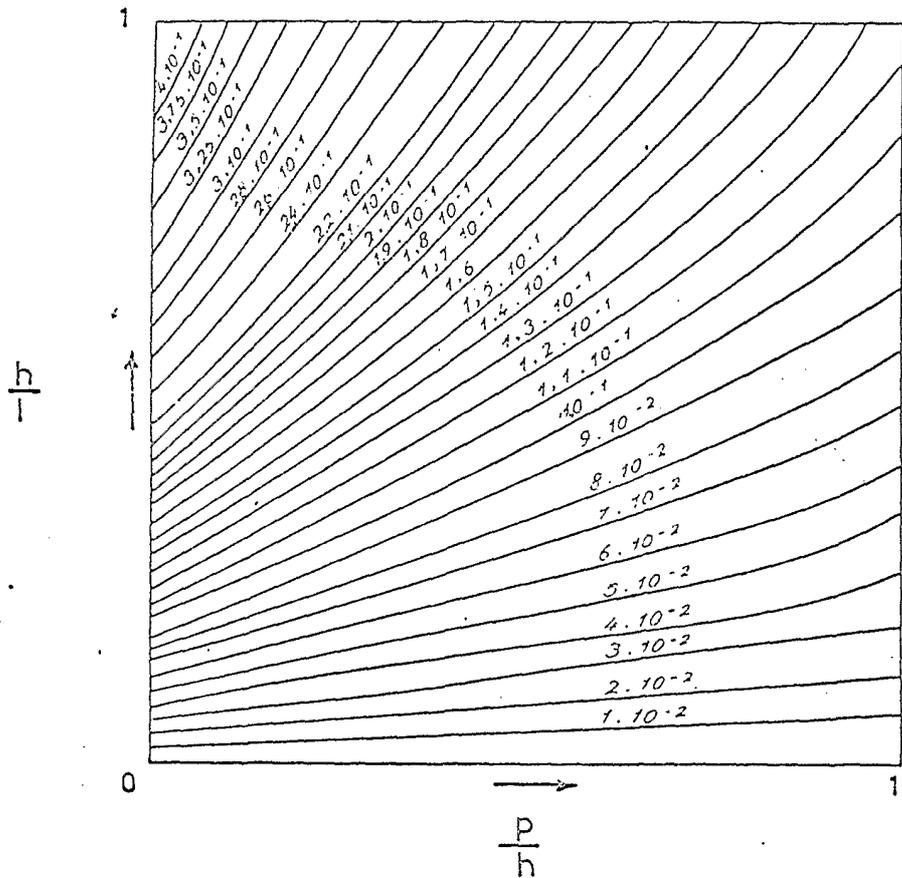
ordonnées et axe des abscisses à division logarithmique - graduation de l'un - 1 à 100 - de l'autre - 1 à 1000 - Unité 100 %





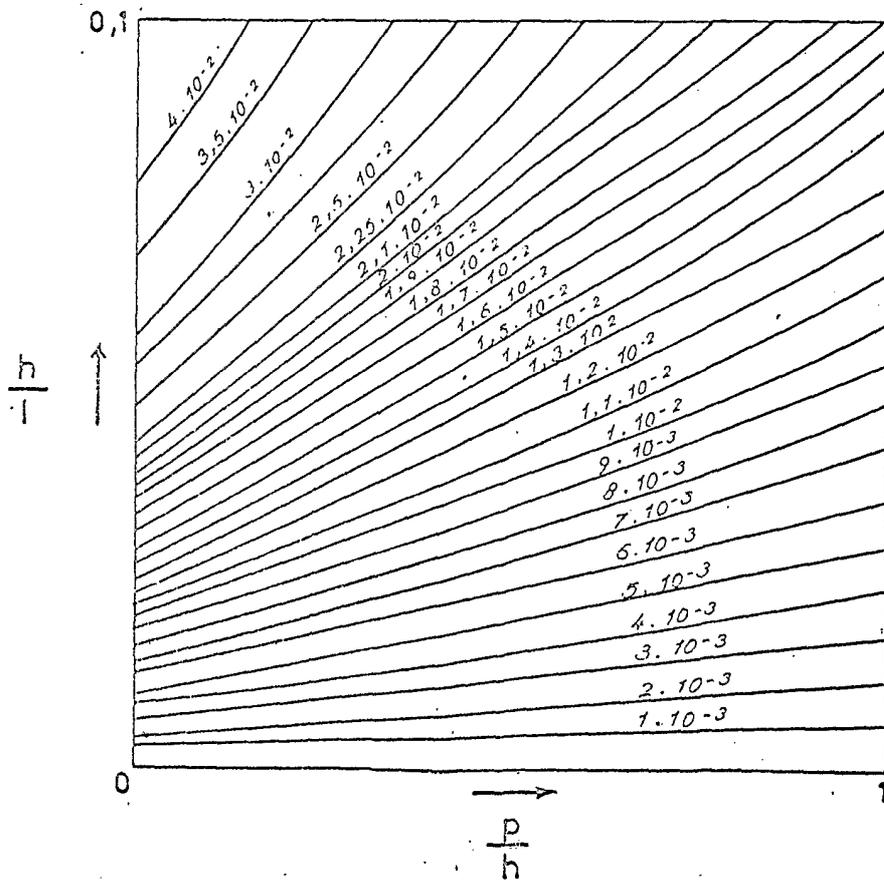
Axe des ordonnées et axe des abscisses à division logarithmique graduation de l'un 1 à 100 de l'autre 1 à 1000 Unité 100 %

Abaques XI-8

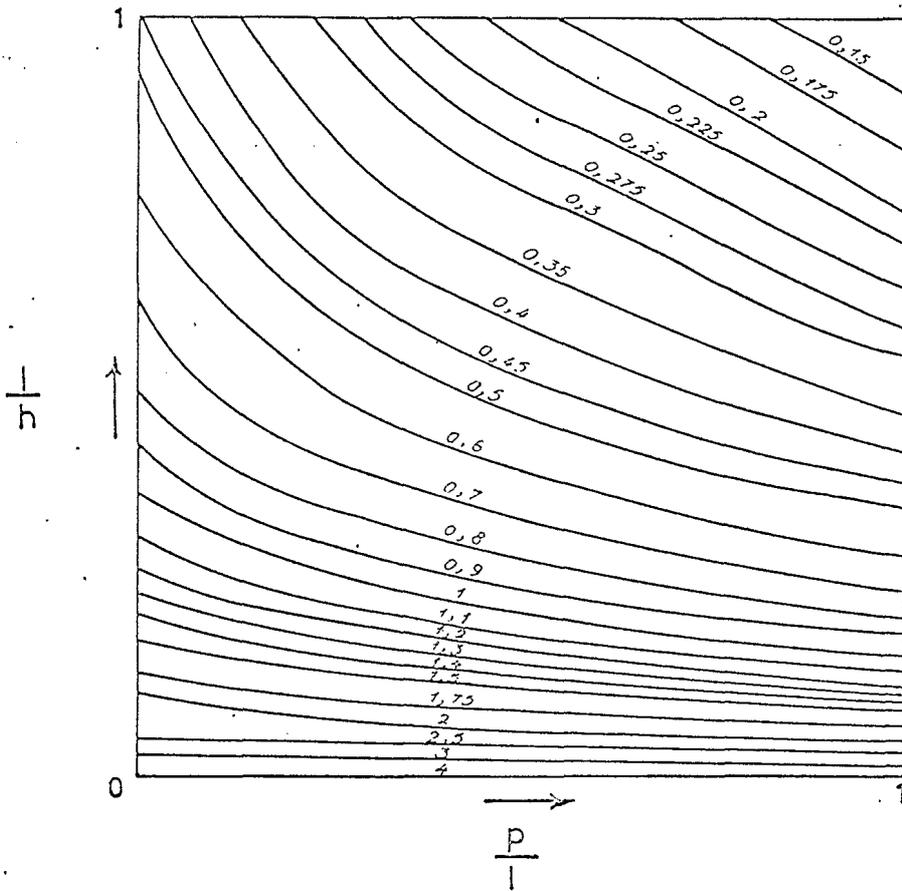


Abaque III. — Variance d'extension d'une section intermédiaire $l \times p$ dans sa zone d'influence de relevée $h = l \geq h \geq p$.

Les courbes sont graduées en $\frac{I}{3\alpha} \sigma_B^2$.



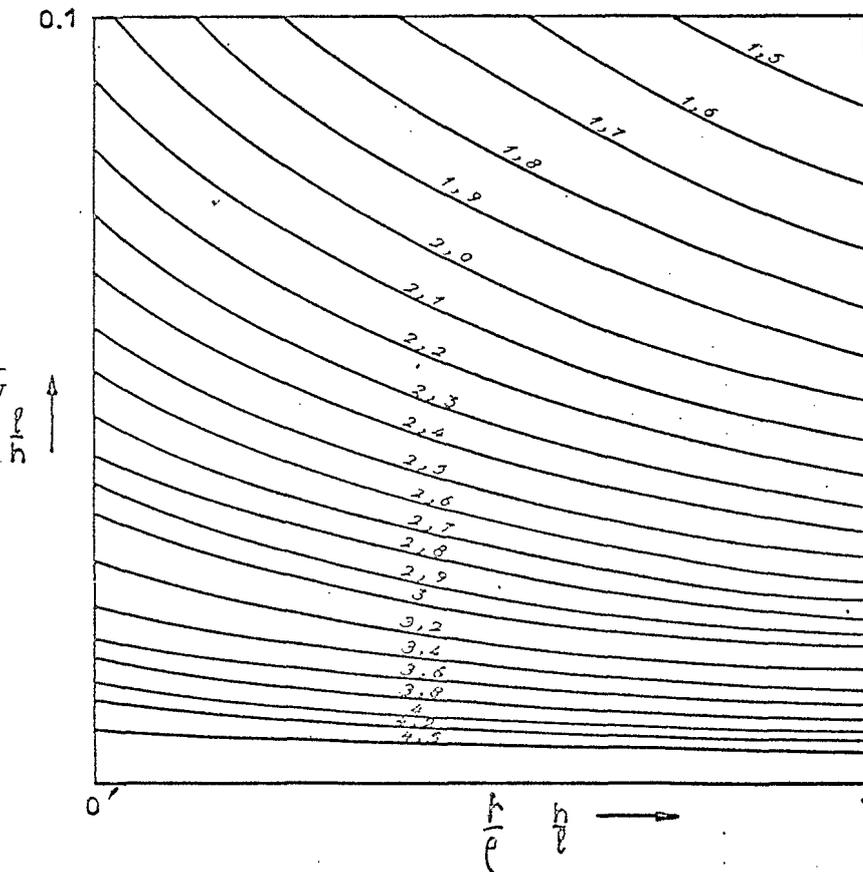
Abaque III bis. — Agrandissement de l'abaque III pour $\frac{h}{l}$ petit.



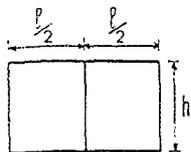
Abaque XI-9

Abaque IV. — Variance d'extension d'une petite section $l \times p$ dans sa zone d'influence de relevée $h = h \geq l \geq p$. Les courbes sont graduées en $\frac{\pi}{3\alpha} \sigma_0^2$.

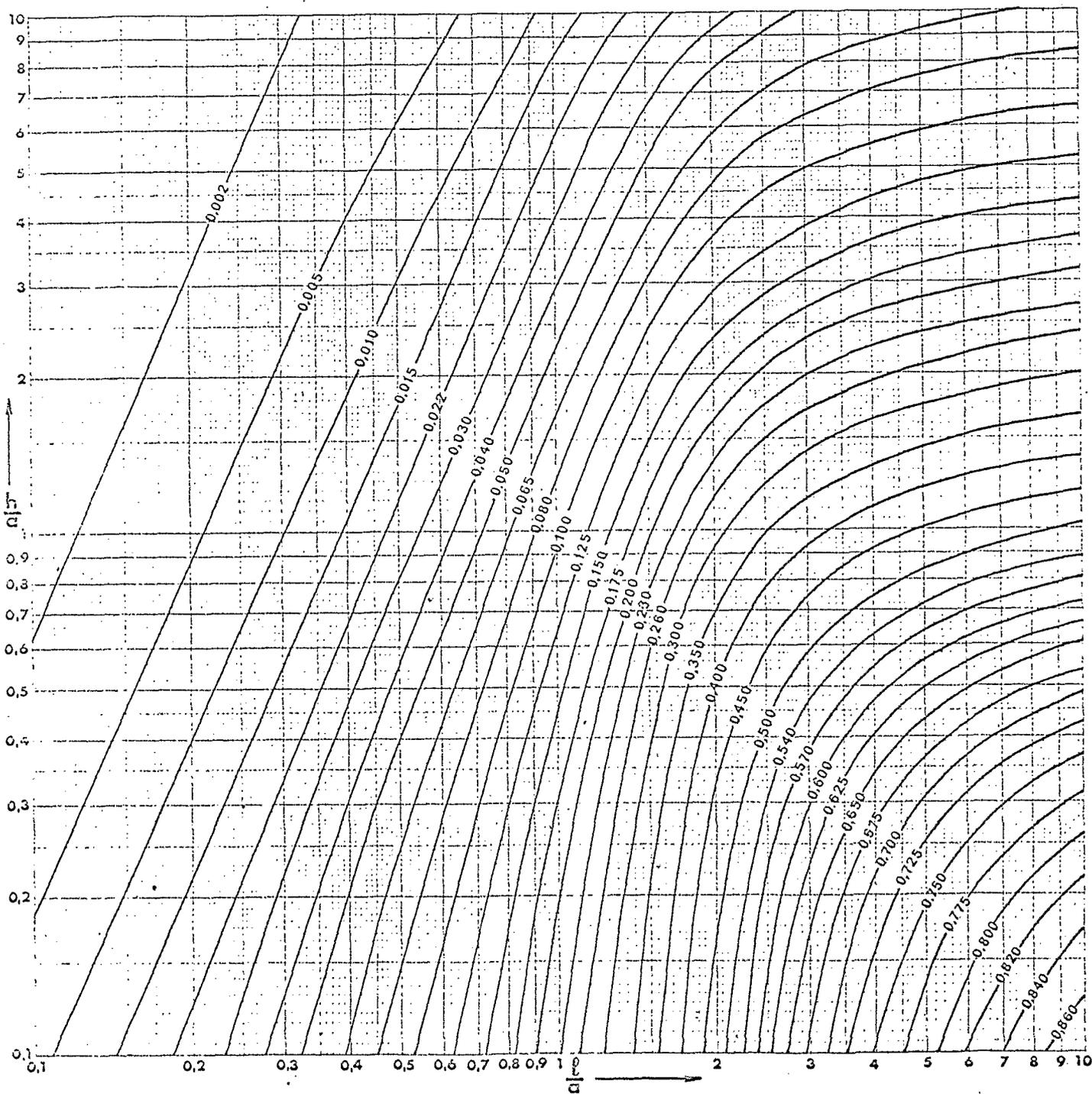
Abaque IV bis. — Agrandissement de l'abaque IV pour $\frac{l}{h}$ petit.



N° 452
 (66) IV-29



Variance d'extension du tracé médian h dans le rectangle $l \times h$

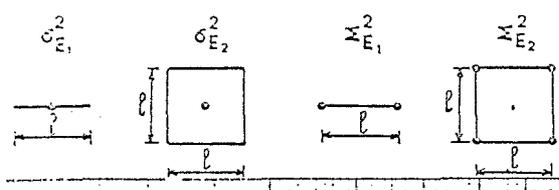


Abaque XIV-3
 Fig. 13 Schéma de Matheron

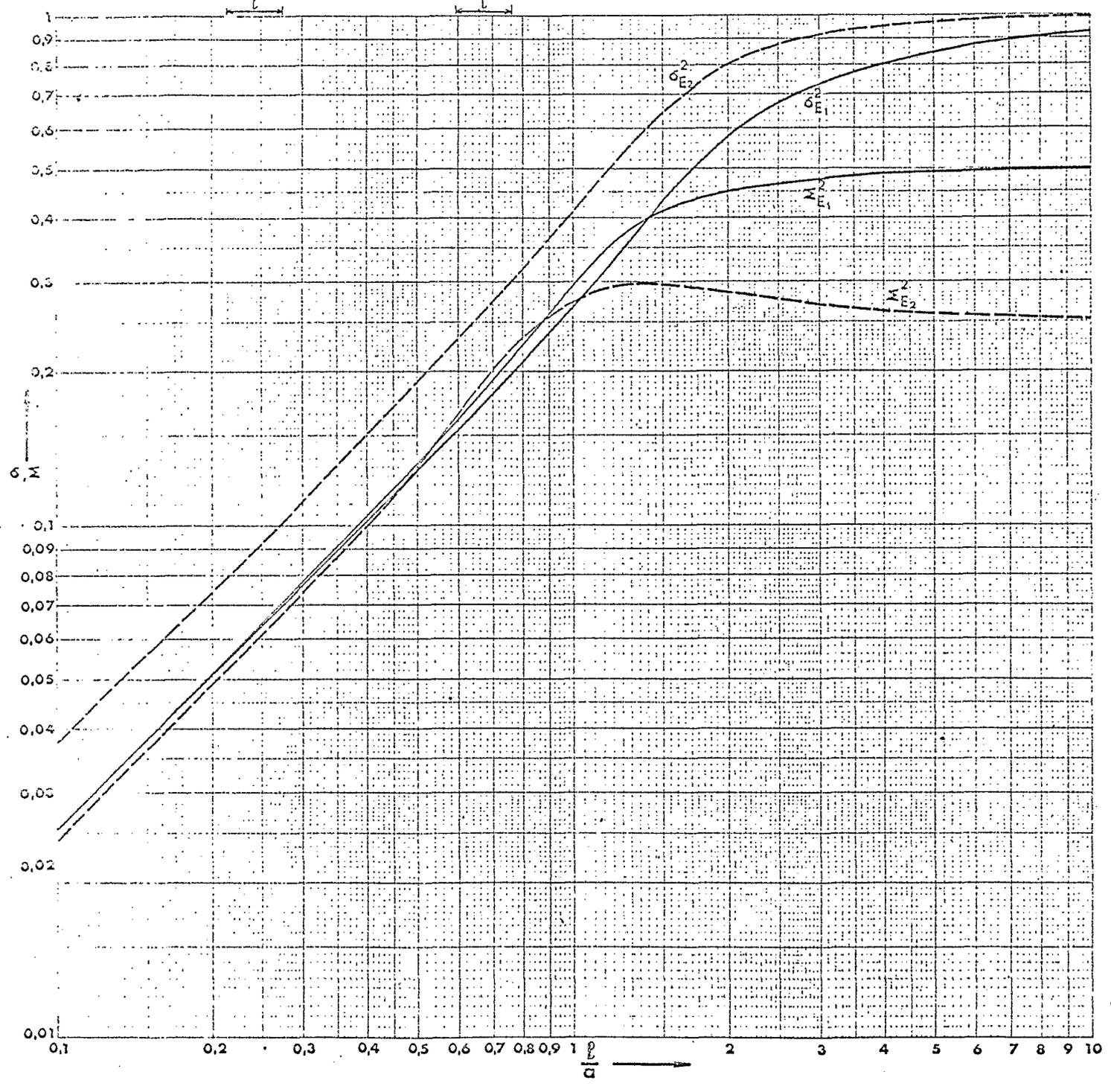
453
IV-27

67

AB XIV-4



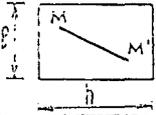
Variances d'extension diverses.



~~Fig 11 - Schéma de Matheron~~

Abaque XIV-4

10-25 AB. XIV-2
 (68)



Fonction $\frac{1}{c} F\left(\frac{p}{a}, \frac{h}{a}\right)$ = Variance d'un point dans le rectangle $E \times h$.

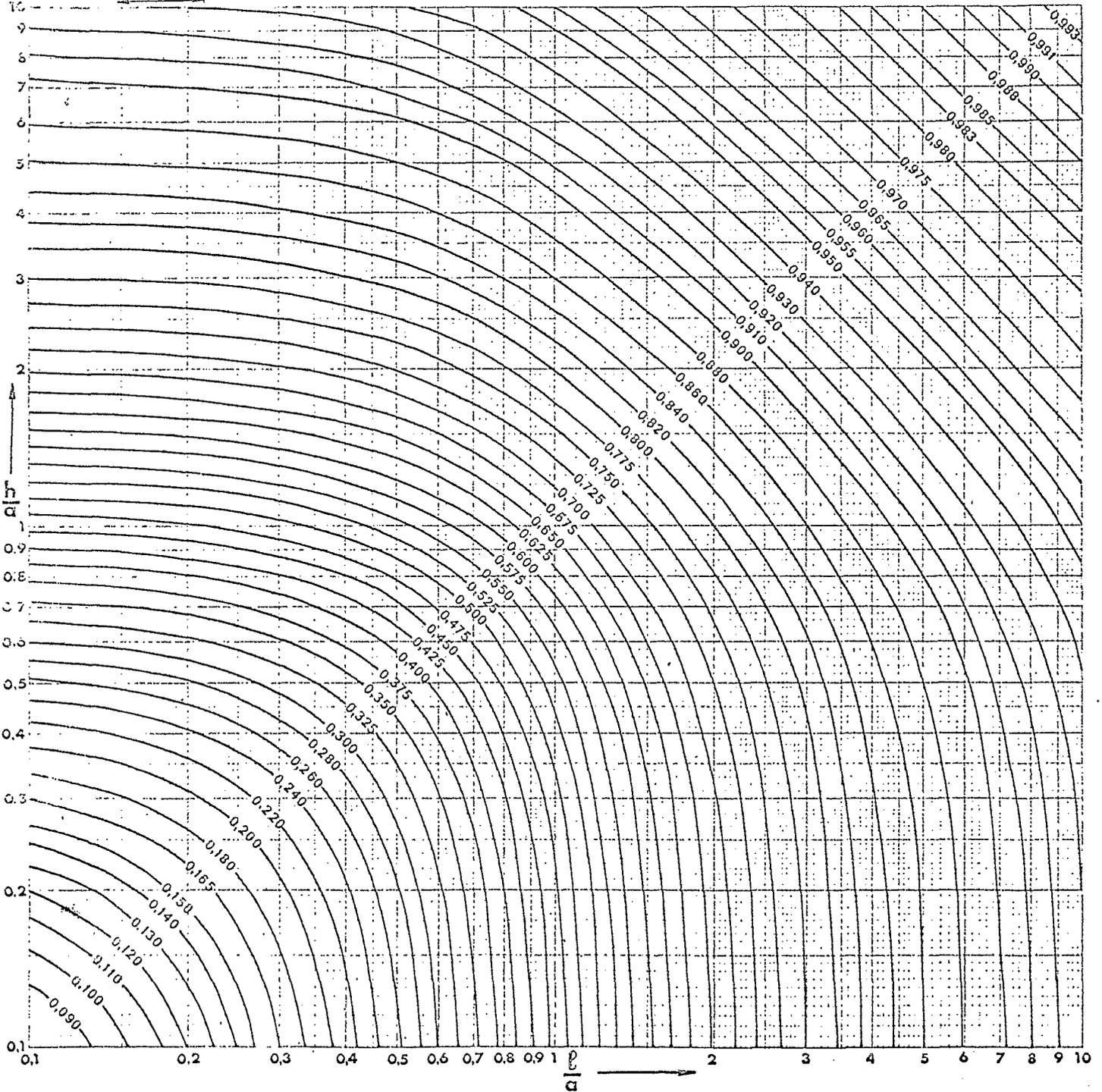


Fig. 9 Schéma de Matheron

A Gaqu XIV-2