

BUREAU DE RECHERCHES  
GEOLOGIQUES & MINIERES



PRECISION DE LA RECONNAISSANCE PAR SONDAGES  
A MAILLE RIGIDE D'UNE FORMATION STRATIFORIE  
- APPLICATION A UN GISEMENT DE BAUXITE -

par

G. MATHERON

Ingénieur des Mines

Paris, le 20 Octobre 1960.

Précision de la Reconnaissance par sondages à maille rigide  
d'une formation stratiforme - Application à un gisement de bauxite

par

G. MATHÉRON, Ingénieur des Mines (B.P.G.M.)

R E S U M E

L'exposé comporte deux parties : rappel des formules théoriques relatives aux sondages à maille rigide et vérification expérimentale dans le cas d'un gisement de bauxite.

1 - Formules théoriques

On considère la population statistique constituée par les teneurs  $x$  des sondages en nombre infini, qu'il est théoriquement possible d'implanter dans le gisement. Aucune hypothèse n'est faite sur la loi de distribution statistique de ces teneurs. Les teneurs  $x_1$  et  $x_2$  de deux sondages distants de  $d$  ne sont pas indépendantes. Plus petite sera la distance  $d$ , plus faible, en moyenne, la différence des teneurs  $x_1$  et  $x_2$ . Cette corrélation, au sens statistique, traduit l'existence d'un certain degré de continuité de la minéralisation et peut se représenter à l'aide d'un variogramme, c'est-à-dire la courbe  $\sigma_d^2 = f(d)$  donnant la variance de la différence  $(x_1 - x_2)^2$  en fonction de la distance  $d$ .

Le variogramme peut souvent être représenté par une formule logarithmique  $f(d) = A + 4\alpha \log d$ , où  $\alpha$  est la dispersion absolue. Pour les valeurs très faibles de la distance  $d$  cette expression n'est plus utilisable. Mais, si la maille des sondages est grande vis à vis de la puissance de la couche, le comportement de  $f(d)$  autour de  $d = 0$  n'a pas d'importance pratique.

La variance d'échantillonnage - variance fixent la précision avec laquelle est connue la teneur d'un gisement reconnu par  $n$  sondages à maille  $d$  se met sous la forme

$$\sigma^2 = \frac{1}{2n} f\left(\frac{d}{2,91}\right)$$

## 2 - Vérification expérimentale

Le gisement africain de bauxite de MEHENGUI a été reconnu par 300 sondages à maille 50 m. De plus, deux carrés de 50 x 50 ont été reconnus chacun par 121 sondages à maille 5 m.

Le variogramme a été construit à partir de 29 points expérimentaux compris entre les valeurs  $d = 5$  et  $d = 500$ . Le distinct on des couples orientés NS, EW, et NE-SW, NW-SE ne fait pas apparaître d'anisotropie. Les valeurs relatives à chacun des deux carrés 50 x 50 sont en remarquable parallélisme, ce qui indique que les petites irrégularités de la courbe correspondent à un phénomène réel. L'ensemble des points expérimentaux s'aligne approximativement en diagramme logarithmique. Les écarts ne sont pas dus à de simples fluctuations, mais leur amplitude est assez faible pour que les formules théoriques restent utilisables.

Parmi les 300 sondages à maille 50 m, il est possible, de 16 façons différentes, d'extraire un schéma de reconnaissance à maille 200. On étudie les distributions de la moyenne et de la variance de ces 16 schémas. La dispersion des moyennes, en particulier, permet une bonne estimation de la variance d'échantillonnage de la maille 200. La valeur expérimentale obtenue est en excellent accord avec la valeur déduite des formules théoriques :

	Prévision théorique	Valeur expérimentale
Variance d'échantillonnage $\sigma^2$	0.62	0.60
Erreur possible $\pm 26$	$\pm 1.57$	$\pm 1.55$

PRECISION DE LA RECONNAISSANCE PAR SONDAGES A MAILLE RIGIDE  
D'UNE FORMATION STRATIFORME - APPLICATION A UN GISEMENT DE BAUXITE -

I - GENERALITES SUR LA PRECISION D'UNE MAILLE DE RECONNAISSANCE

a) Notion de variance - Variance d'échantillonnage

Il est théoriquement possible d'implanter dans un gisement une infinité de sondages ou de puits. L'ensemble des teneurs  $x_i$  que donnerait la totalité de ces sondages ou de ces puits constitue une population statistique. Il est commode de caractériser une telle population par deux paramètres :

- une moyenne arithmétique

$$m = \frac{1}{n} \sum x_i \quad (n = \text{nombre de sondages})$$

qui n'est autre que la teneur moyenne du gisement.

- une variance ou écart quadratique moyen

$$\sigma^2 = \frac{1}{n} \sum (x_i - m)^2$$

qui mesure le degré de dispersion des valeurs individuelles autour de leur valeur moyenne.

La racine carrée de la variance est l'écart-type  $\sigma$ . Si les valeurs individuelles se répartissent selon la loi lognormale (courbe en cloche) 95 % d'entre elles sont comprises dans l'intervalle  $m \pm 2\sigma$ . Une valeur quelconque, choisie au hasard, a 95 chances sur 100 de tomber dans cet intervalle. L'intervalle  $\pm 2\sigma$  donne donc une mesure pratique de la marge de variation possible.

Lorsqu'un gisement a été reconnu par un nombre limité  $n$  de sondages, la moyenne  $m'$  des teneurs de ces sondages n'est généralement pas rigoureusement égale à la teneur moyenne réelle  $m$  du gisement. On conçoit qu'il serait possible d'implanter d'une infinité de façons le même nombre  $n$  de sondages à la même maille régulière, les différents schémas possibles se déduisent les uns des autres par translation. On conçoit aussi que chacun de ces schémas donnerait une moyenne arithmétique  $m'$  différente. A ce titre,  $m'$  apparaît comme une variable aléatoire ; l'ensemble des valeurs moyennes  $m'$  correspondant à l'ensemble des implantations possibles constitue une population statistique. Et cette population nous pouvons la caractériser à son tour par deux paramètres :

- une valeur moyenne qui n'est autre que la teneur réelle du gisement,

- une variance, que nous appellerons variance d'échantillonnage  $\sigma^2_E$

Cette variance d'échantillonnage caractérise la dispersion des estimations possibles de la teneur moyenne pour un nombre de sondages et une maille donnés : autrement dit elle mesure la précision d'un schéma de reconnaissance.

L'écart-type d'échantillonnage,  $\sigma E$ , est intéressant à considérer. En effet, même si les teneurs des sondages ne sont pas rigoureusement distribuées selon une loi normale, les moyennes arithmétiques  $m'$  ont une distribution qui converge vers le type normal lorsque le nombre  $n$  augmente. Dans le cas de gisements de bauxite, on est pratiquement assuré de la normalité. Il en résulte que la teneur moyenne fournie par un schéma d'implantation quelconque comprenant  $n$  sondages à la maille imposée a 95 chances sur 100 de tomber dans l'intervalle  $m \pm 2 \sigma E$ .

En réalité,  $m$  n'est pas connue. Mais on conçoit qu'il y a aussi à peu près 95 chances sur 100 pour que la teneur moyennée <sup>réelle</sup>  $m$  ne diffère de la valeur  $m'$  fournie par les sondages que d'une quantité inférieure à  $2\sigma$  en valeur absolue.

Pratiquement, cela signifie que l'on peut prendre comme limite de l'erreur sur la teneur moyenne le double,  $+ 2 \sigma E$ , de l'écart-type d'échantillonnage.

b) Expression théorique de la variance d'échantillonnage - Le variogramme

Mathématiquement parlant, les teneurs  $x_1$  et  $x_2$  que donneraient deux sondages implantés à une distance  $d$  l'un de l'autre ne peuvent pas être considérées comme des données indépendantes. On conçoit que, si la distance  $d$  est faible, et si le sondage 1 a donné une teneur  $x_1$  élevée, le sondage 2 implanté à la distance  $d$  a plus de chances de donner une teneur élevée qu'un sondage implanté au hasard. Deux sondages voisins donneront, en moyenne, des teneurs moins différentes l'une de l'autre que deux sondages quelconques. Cette corrélation, au sens statistique, entre les teneurs de sondages voisins ne signifie rien d'autre que l'existence d'un certain degré de continuité de la minéralisation. Il est intéressant d'essayer de chiffrer ce degré de continuité, qui doit être évidemment différent selon les types de gisement. Indiquons simplement ici qu'il est commode de représenter cette relation de continuité au moyen d'une courbe que nous appellerons variogramme.

Si nous considérons l'ensemble des couples  $x_1$  et  $x_2$  des teneurs de sondages possibles implantés dans le gisement à la distance  $d$  l'un de l'autre, la variable  $x_1 - x_2$  se caractérise par une moyenne, qui est nulle et une variance :

$$\sigma^2 d = \frac{1}{n} \sum (x_1 - x_2)^2$$

qui donne une mesure de la dépendance l'une envers l'autre de ces deux teneurs. Le variogramme est la courbe représentant cette variance en fonction de la distance  $d$

$$(1) \sigma^2 d = f(d)$$

On s'attend à ce que  $f(d)$  soit une fonction croissante de la distance  $d$ , puisqu'eussi bien plus les sondages sont voisins moins leurs teneurs doivent différer en moyenne. A la limite, pour  $d = 0$   $f(d)$  doit s'annuler puisque les deux sondages coïncident. Toutefois, le comportement de la fonction au voisinage de l'origine n'a pas grande importance du point de vue pratique. Dans le domaine d'utilisation usuelle, le variogramme peut souvent se représenter de façon très acceptable par une équation de la forme :

$$(2) \quad f(d) = A + 4\alpha \log d$$

Le paramètre  $\alpha$ , ou dispersion absolue, caractérise le degré de discontinuité de la minéralisation. C'est un paramètre intrinsèque dont la signification théorique est étroitement liée au principe de similitude et présente une grande importance pour la métallogénie.

Sans insister ici sur ces implications théoriques, nous considérons cette relation comme représentant une bonne interpolation des données expérimentales.

On démontre que, si cette relation est vérifiée, la variance d'échantillonnage fixant la précision avec laquelle est connue la teneur d'un gisement reconnu par  $n$  sondages implantés selon une maille carrée de côté  $d$ , est donnée par la formule :

$$(3) \quad \sigma^2 = \frac{\sigma^2 d}{2n} - \frac{2,14\alpha}{n} = \frac{1}{2n} f\left(\frac{d}{2,91}\right)$$

Il n'est pas étonnant que la variance d'échantillonnage prenne une forme en  $1/n$  et soit liée étroitement au variogramme. En effet, l'erreur possible provient du fait que la teneur  $x_i$  de chaque sondage diffère de la teneur réelle  $m_i$  de son "polygone d'influence". L'erreur totale est de la forme :

$$\frac{1}{n} \sum (x_i - m_i)$$

Si les différences  $(x_i - m_i)$  peuvent, en moyenne, être considérées comme indépendantes les unes des autres, et si  $h^2$  est la variance de  $(x_i - m_i)$  la variance d'échantillonnage doit être égale à :  $\frac{h^2}{n}$ . Donc :

- La variance est bien en  $\frac{1}{n}$ , ce qui signifie qu'à maille égale un gisement 4 fois plus grand  $n$  (reconnu par quatre fois plus de sondage) est connu avec une précision double.

- Elle est proportionnelle à  $h^2$  qui représente une sorte de demi-moyenne des valeurs du variogramme pour les valeurs de la distance inférieure à  $d$ . Il n'est donc pas étonnant de trouver une formule type (3). Seul, en fait, la valeur numérique 2,91 du facteur de  $d$  dépend de la nature logarithmique du variogramme.

L'intérêt de la formule (3), c'est que le variogramme  $f(d)$  est directement accessible à l'expérience.

## II - APPLICATION AU GISEMENT DE BAUXITE DE MEHENGUI

### LES DONNEES

Le gisement stratiforme de bauxite de Mehengui forme une couche continue peu puissante. Elle a été reconnue au moyen de 300 sondages utiles implantés à la maille carrée de 50 m. De plus, deux carrés de 50 x 50 désignés par la suite sous le nom de Zone I et Zone II, ont été reconnus chacun par 121 sondages à maille de 5 m dans le but de tester l'homogénéité des teneurs. Chaque sondage est caractérisé par sa teneur en  $Al_2O_3$ , prise entre toit et mur.

### BUT DE CETTE ETUDE

L'objet de cette étude est avant tout de vérifier, sur des données expérimentales, la validité de la loi.

$$\sigma_{xy} = A - 2\alpha \log d$$

donnent la covariance  $xy$  des teneurs des sondages distants de  $d$ . Il est possible pour Mehengui, de calculer numériquement de telles covariances pour des valeurs de la distance  $d$  comprise entre 5 et 500m soit un intervalle de variation de 1 à 100.

Pour éliminer l'effet de bordure, on calculera numériquement non pas les covariances elles-mêmes, mais les variances  $\sigma^2 d$  des différences  $(x - y)$  des teneurs  $x$  et  $y$  des sondages distants de  $d$ . On a :

$$\sigma^2 d = 2(\sigma^2 - \sigma_{xy}) = B + 4\alpha \log d$$

$\sigma^2$  étant la variance des teneurs des sondages. L'expression  $\sigma^2 d$  présente de plus l'avantage d'être indépendante de la dimension du gisement ou de la portion de gisement d'où proviennent les données utilisées pour le calcul. C'est une fonction intrinsèque de la distance  $d$ . Il nous sera possible, par conséquent de porter sur un même graphique les données tirées de la maille de 50 m et des deux carrés reconnus à la maille de 5 m.

### I - Etude de la maille de 50 m

L'histogramme des teneurs en  $Al_2O_3$  est reproduit ci-contre :

.../...

Teneurs	Nombre de sondages	:	Teneurs	Nombre de sondages
27	1	:	43	37
32	1	:	44	26
33	3	:	45	36
34	3	:	46	24
35	6	:	47	15
36	7	:	48	13
37	11	:	49	6
38	14	:	50	5
39	12	:	51	1
40	22	:	52	1
41	19	:	55	1
42	36	:		

On en déduit :

- Nombre de sondages  $N = 300$
- Moyenne des teneurs  $m = 42,63$
- Variance des teneurs  $= 15,30$

Les variances des différences  $(x - y)$  des sondages distants de  $d$  ont été calculées pour 10 valeurs de  $d$ , de 50 à 500 m, pour les alignements EW et NS séparément :

d (unité 50 m)	Nombre de données		Variance $\sigma^2 d$		
	EW	NS	EW	NS	EW + NS
1	267	266	23.7	26.0	25.1
2	245	243	26.4	24.6	25.5
3	223	233	26.6	28.4	27.6
4	199	222	30.2	26.8	28.5
5	180	204	28.8	31.4	30.1
6	159	185	31.0	31.0	31.0
7	138	173	35.3	29.4	31.9
8	118	161	30.5	25.7	27.8
9	95	147	30.5	30.5	30.5
10	76	130	31.6	27.8	29.2

Les variances EW et NS sont compatibles entre elles, compte tenu du nombre de données disponibles. Un test  $\chi^2$  sur l'homogénéité de la distribution EW et NS des différences d'ordre 1, donne  $\chi^2 = 11.09$  pour 10 degrés de liberté, soit  $P = 0.35$ , ce qui est tout à fait admissible.

Il n'y a pas d'anisotropie.

Le même calcul a été exécuté pour les différences prises sur des diagonales NE-SW et NW-SE séparément :

.../...

d Unité 50 m :	Nombre n de données :			Variance $\sigma^2_d$ :		Ensemble
	NE-SW	NW-SE	Total	NE-SW	NW-SE	
1.41	249	248	497	27.10	21.75	24.45
2.83	224	216	440	33.3	29.8	31.5
4.24	204	189	393	29.0	30.3	29.7
5.65	182	160	342	31.8	29.6	30.3
7.06	161	129	290	31.4	23.6	32.4
8.49	137	100	237	34.2	31.2	33.7
9.90	119	73	192	39.6	33.2	37.0
11.30	103	52	155	28.2	27.6	28.3

II - Etude de la maille de 5 m (Zones I et II)

	Zone I	Zone II
Nombre de données :	121	121
Moyenne :	44.6	43.3
Variance :	9.30	7.45

Les deux moyennes sont significativement différentes. Les deux variances ne le sont probablement pas, mais il est difficile de faire un test précis, les 121 données de chaque carré n'étant pas indépendantes.

L'étude des  $\sigma^2_d$  ne met en évidence de différences significatives entre données NS et EW, ou NE-SW et NW-SE ni dans la zone I ni dans la Zone II. Par contre, les valeurs globales sont significativement différentes entre Zone I et Zone II. La courbe I montre le remarquable parallélisme des courbes  $\sigma^2_d = f(d)$  pour chacune des deux zones. Il est frappant de constater que chaque petite irrégularité d'une courbe, que l'on serait tenté d'attribuer à une fluctuation non significative, a son exact correspondant sur l'autre courbe. Très caractéristique est la variation de la pente (dérivée seconde) :

	1.41	2	2.80	3	4	4.25	5	5.65	6	7	8	8.5	9
Zone I	+	-	+	-	-	+	-	+	-	+	+	+	-
Zone II	+	-	+	-	-	+	-	+	-	+	-	+	-

Le signe de ces variations est le même en 12 points sur 13. Cela signifie que ces petites irrégularités d'apparence fortuite correspondent à un phénomène réel, d'ampleur assez modeste il est vrai pour ne pas interdire une interpolation par une courbe continue.

Zone I

d Unité 5 m	Nombre de données			Variance $\sigma^2 d$		Ensemble
	N.S	E.W	NS + EW	N.S	E.W	
1	110	110	220	12.65	11.6	12.12
2	99	99	198	14.7	11.55	14.62
3	88	88	176	17.3	14.3	16.05
4	77	77	154	20.0	15.6	17.80
5	66	66	132	22.5	18.4	20.45
6	55	55	110	21.2	22.9	22.05
7	44	44	88	24.0	17.4	20.70
8	33	33	66	24.3	18.9	21.60
9	22	22	44	33.8	25.1	29.45
10	11	11	22	32.1	27.8	29.95

Il n'est guère possible de conclure à une hétérogénéité des valeurs relatives aux séries EW et NS. Il y a isotrope horizontale dans la Zone I pour la maille de 5 m. Il en est de même pour la Zone II, où l'on ne doit pas attacher trop d'importance aux divergences qui se manifestent à partir du rang 7, celles-ci restent admissibles en regard au petit nombre de données.

**Zone II**

d Unité 5 m	Nombre de données			Variance $\sigma^2 d$		Ensemble
	N	S	N + S	N	S	
1	110	110	220	6.35	6.61	6.73
2	99	99	198	9.30	10.2	9.75
3	88	88	176	11.72	12.9	12.31
4	77	77	154	13.70	14.6	13.70
5	66	66	132	15.35	16.2	15.07
6	55	55	110	16.75	18.4	17.59
7	44	44	88	18.75	21.6	19.92
8	33	33	66	21.50	25.2	21.15
9	22	22	44	25.25	30.2	23.42
10	11	11	22	30.30	36.1	21.70

Enfin, les  $\sigma^2 d$  ont été calculés, pour chaque zone, le long des diagonales à 45°, sans distinguer les alignements NE-SW et NW-SE. (Les histogrammes établis séparément pour chacune des directions sont pratiquement identiques, comme il fallait s'y attendre, en l'absence d'anisotropie).

d Unité 5 m	Nombre de données			Variance $\sigma^2 d$		
	Zone I	Zone II	Ensemble	Zone I	Zone II	Ensemble
1.41	200	200	400	11.91	7.28	9.59
2.83	162	162	324	15.60	10.20	12.90
4.24	128	128	256	17.60	11.21	14.40
5.65	98	98	196	20.90	13.34	17.12
7.06	72	72	144	20.00	16.38	18.19
8.49	50	50	100	24.04	21.62	22.82
9.90	32	32	64	25.35	24.6	25.22

Ayant admis l'absence d'anisotropie, nous pouvons porter sur un même diagramme les valeurs du  $\sigma^2 d$  de 5 en 5 m et de  $5\sqrt{2}$  en  $5\sqrt{2}$ . Les points 7.06 et 9.90 de cette dernière maille seront identifiés aux points 7 et 10 de la première, et les données correspondantes groupées ensemble :

d	Nombre de données			Variances $\sigma^2 d$			
	Unité 5 m	Zone I	Zone II	Ensemble	Zone I	Zone II	Ensemble
1		220	220	440	12.12	6.73	9.42
1.41		200	200	400	11.91	7.28	9.60
2		198	198	396	14.62	8.65	12.18
2.83		162	162	324	15.60	10.20	12.90
3		176	176	352	16.05	12.31	14.18
4		154	154	308	17.22	13.71	15.75
4.24		128	128	256	17.60	11.21	14.40
5		132	132	264	20.15	15.07	17.76
5.65		98	98	196	20.92	12.34	17.12
6		110	110	220	20.05	12.32	20.18
7		160	160	320	20.35	17.65	18.95
8		66	66	132	21.69	21.15	21.37
8.49		50	50	100	21.04	21.62	22.33
9		44	44	88	22.45	21.22	26.33
10		54	54	108	27.70	23.40	25.55

Ces valeurs sont portées sur la figure I ; le diagramme met en évidence le parallélisme étroit des deux courbes relatives aux zones I et II, parallélisme qui va jusqu'à reproduire dans le détail les fluctuations de la variation de la pente. Nous en avons conclu plus haut que les petites irrégularités de la courbe, malgré leur apparence fortuite, devaient correspondre à un phénomène réel, quoique d'amplitude faible. Il apparaît également que les deux séries de valeurs sont significativement différentes.

La Zone I, dans laquelle la variance des 121 sondages est 9.30, donne des valeurs plus fortes que la Zone II, où cette variance est 7.45. La différence entre les deux séries de valeurs est à peu près constante, et égale au double  $(9.30 - 7.45) = 3.70$  de la différence des variances. Or, on a, pour chaque  $\sigma^2 d$

$$\sigma^2 d = 2 (\sigma^2 - \sigma_{xy})$$

$\sigma_{xy}$  étant la covariance (calculée dans le carré, c'est-à-dire dans la population des 121 sondages) des sondages distants de d. Il apparaît ainsi que la courbe donnant cette covariance en fonction de d est à peu près la même pour chacun des deux carrés. Ceci est représenté en figure II.

Ainsi, les différences des  $\sigma^2$  entre les deux zones proviennent uniquement de la différence des variances  $\sigma^2$  des deux populations. Cette différence une fois imposée se répercute numériquement sur tous les  $\sigma^2 d$  malgré l'identité des covariances. La différence des  $\sigma^2 d$  n'est donc pas plus significative que celle de ces deux variances (et nous avons vu qu'on ne pouvait affirmer qu'elle l'était).

### III - Représentation de l'ensemble du phénomène

La figure III représente l'ensemble des données regroupées sur une courbe unique.

d Unité 5 m	Nombre de données	Variance $\sigma^2 d$
1	440	9.42
1.41	400	9.60
2	396	12.18
2.83	324	12.90
3	352	14.18
4.24	308	15.75
4	256	14.4
5.65	264	17.76
5	196	17.12
6	220	20.18
7	320	18.95
8	132	21.37
8.49	100	22.83
9	88	26.33
10	641	25.20
14.1	497	24.45
20	488	25.50
28.2	440	31.50
30	456	27.60
40	421	28.50
42.5	393	29.70
50	384	30.1
56.5	342	30.3
60	344	31.0
70	601	32.1
80	279	27.8
84.9	107	33.7
90	242	30.5
100	398	33.2

L'alignement, en échelle logarithmique, est bonne toute satisfaisant. Les trois anomalies pour 9, 28 et 80 sont presque significatives.

Bien qu'on ne puisse l'affirmer, il est probable qu'elles correspondent à de petites déviations réelles. Il n'en reste pas moins qu'en gros le phénomène obéit à une loi logarithmique : les écarts qui paraissent graves disparaîtraient presque si l'on rétablissait l'échelle arithmétique de 1 à 100. Nous pensons que la loi logarithmique

$$\sigma^2 d = 7,9 + 5,76 \log d \quad (\log \text{ népérien} \\ d = \text{unité } 5 \text{ m})$$

peut être utilisée pour établir des prévisions.

#### IV - Discussion d'une maille

A partir de la formule logarithmique précédente, nous allons établir la loi donnant la précision d'une maille déterminée. Nous comparerons ensuite, à la précision théorique trouvée, la précision réellement obtenue pour une maille de 200 m.

##### 1 - Formule théorique

Pour un gisement reconnu par n sondage à la maille a, la variance d'échantillonnage est :

$$\sigma^2 E = \sigma^2 \frac{a}{2n} - \frac{2.14\alpha}{n} = \frac{1}{2n} f \left( \frac{a}{2.91} \right)$$

Il faut diviser par 2 n la valeur du  $\sigma^2 d$  pour  $\frac{a}{2.91}$

Par exemple, pour une maille de 200 m ( a = 40 en unité 5 m), nous prendrons le  $\sigma^2 d$  correspondant à  $\frac{40}{2.91} = 13.75$  avec  $\sigma^2 d = 23.2$  D'où

$$\sigma^2 E = \frac{23.2}{2n}$$

Pour le gisement de Mehengui, reconnu par 300 sondages à maille de 50 m, on aurait à la maille 200

$$n = \frac{300}{16} = 18.7 \text{ et}$$

$$\sigma^2 E = \frac{23.2}{2 \times 18.7} = 0.62$$

$$\text{D'où } \left\{ \begin{array}{l} \text{Ecart type } \sigma_E = 0.785 \\ \text{Erreur possible ( au niveau } 95 \% ) \pm 2\sigma_E = \pm 1.57 \end{array} \right.$$

Telles sont les prévisions de la théorie que nous allons maintenant comparer aux données expérimentales.

2 - Vérification expérimentale

A partir des 300 sondages implantés à la maille de 50 m, nous pouvons choisir 16 schémas d'implantation possibles à la maille 200. Pour chacun de ces schémas, on a calculé moyenne et variance des teneurs. Du fait qu'ils sont au nombre de 16, on obtient ainsi une bonne approximation de la loi de distribution de l'estimation de la moyenne à la maille 200, donc une bonne estimation de  $\sigma^2 E$ .

Les 16 valeurs obtenues pour la moyenne s'échelonnent entre 40.90 et 43.94 - soit dans un intervalle de 3 points. Il est intéressant de voir que le nombre de sondages utiles varie entre 15 et 22, ce qui peut donner une idée de la précision avec laquelle le tonnage serait connu à partir d'une maille 200. Les variances sont elles aussi, fortement dispersées. Mais l'intérêt principal se porte sur la distribution des moyennes.

N° du schéma	Nombre de données	Moyenne	Variance
1	19	43.58	13.05
2	22	43.54	10.60
3	19	41.79	16.05
4	19	43.00	7.90
5	22	42.73	13.80
6	22	42.86	16.40
7	20	40.90	9.24
8	19	42.73	13.10
9	19	43.94	17.60
10	20	43.15	9.84
11	18	42.11	15.75
12	19	41.79	12.90
13	15	42.93	30.60
14	16	42.75	23.40
15	15	41.60	33.60
16	16	42.31	11.40

Distribution des moyennes

Cette distribution est bien caractérisée par les deux paramètres habituels :

$$\text{moyenne } m = 42.59$$

$$\text{variance } \sigma^2_E = 0.596$$

La moyenne  $m = 42.59$  est différente de la moyenne des 300 sondages (42,63) parce-que les 16 schémas comportent un nombre variable de sondages. La variance  $\sigma^2_E$  est remarquablement voisine de la valeur théorique calculée plus haut (0.62 contre 0.596). Ce n'est pas là coïncidence fortuite. Si nous avions estimé cette variance par le procédé vulgaire qui consiste à diviser par  $n = 18.7$  la variance des sondages, soit  $\frac{15.30}{18.7} = 0.82$ , la valeur obtenue serait considérablement plus forte. N

Nous avons donc là une bonne confirmation des théories de l'estimation fondées sur la formule de la covariance logarithmique.

Répartition des variances

Cette répartition est intéressante à considérer sur le plan théoriques. Les 16 variances forment une population

$$\text{de moyenne } E(\sigma^2) = 15.95$$

$$\text{de variance } D^2(\sigma^2) = 54.4$$

Si la moyenne ne diffère pas significativement de la variance générale (15.30) des 300 sondages, par contre la variance  $D^2(\sigma^2)$  est notablement plus forte que la variance  $\frac{2}{n} \sigma^4 = 32$  que l'on aurait observé si nos 16 sous populations avaient été tirées au hasard dans le gisement.

Cette augmentation est la contrepartie logique du fait que la variance  $D^2(m)$  de l'estimation de la moyenne est plus petite que si les sous populations étaient aléatoires. La raison profonde en est que les 18 éléments de chaque sous population ont entre eux une corrélation négatives. Les moyennes devant être moins dispersées que celles de sous populations aléatoires, une sous population contenant une valeur anormalement forte aura plus de chance, aussi, de contenir une valeur anormalement faible ; et réciproquement une population dénuée de fortes valeurs aura aussi tendance à être dénuée de valeurs très faibles : d'où augmentation de la dispersion des variances par rapport aux sous populations aléatoires.

Si donc la maille rigide améliore l'estimation de la moyenne, elle détériore l'estimation de la variance. La théorie correspondante est à faire. Elle présente une grande importance dans le cas lognormal pour l'estimateur de Sichel où figurent à la fois médiane et variance. Cet estimateur devra être profondément modifié.

3) Etude des carrés 200 x 200

Profitant du groupement des données effectuées pour les vérifications expérimentales du paragraphe précédent, nous avons voulu étudier aussi la répartition dans le gisement des teneurs des carrés de 200 x 200. Nous avons retenu les 16 carrés contenant au moins 13 sondages, et étudié les 16 sous populations correspondantes : voir moyennes et variances de ces sous populations.

<u>Carré</u>	<u>Nombre de sondages</u>	<u>Moyenne</u>	<u>Variance</u>
a	16	42.31	33.30
b	15	44.00	11.13
c	16	43.12	8.66
d	14	40.36	8.74
e	16	40.56	11.60
f	16	44.13	15.80
g	14	40.07	16.25
h	16	44.00	24.80
i	16	44.00	8.00
j	13	43.31	5.90
k	14	41.88	15.55
l	16	43.12	12.90
m	16	44.69	7.54
n	13	41.07	8.90
o	16	42.31	3.71
p	15	41.73	8.06

L'implantation des carrés sur le terrain est schématisé comme suit :

```

      a   b   c
      d   e   f
      g   h   i
    j k   l   m
      n   o   p
  
```

On obtient les paramètres suivants, pour les distributions des moyennes et des variances

$$\left\{ \begin{array}{l} \text{moyenne des moyennes } E(m) = 42.54 \\ \text{variance des moyennes } D^2(m) = 2.18 \end{array} \right.$$

$$\left\{ \begin{array}{l} \text{moyenne des variances } E(\sigma^2) = 13.19 \\ \text{variance des variances } D2(\sigma^2) = 56.90 \end{array} \right.$$

La valeur de la variance des moyennes, comme on pouvait s'y attendre, est significativement plus forte que pour des sous populations aléatoires : les carrés ont, en effet, des teneurs réellement différentes.

La somme  $D^2(m) + E(\sigma^2) = 15.37$  est, comme il se doit, très voisine de la variance 15.30 des sondages dans l'ensemble du gisement. La valeur moyenne de la variance, 13.19 est voisine de la valeur que l'on pouvait attendre : La valeur moyenne dans un carré 200 x 200 de la fonction  $\sigma^2 d = 7.9 + 5.76 \log d$  doit être égale au double de la variance des sondages dans ce carré. On trouve

$$\frac{1}{2} E ( 7.9 + 5.76 \log d ) = 14.3$$

valeur théorique en assez bon accord avec la valeur expérimentale 13.19 (moins de 10 % d'écart).

Enfin, la variance - 56.90 - est considérablement plus élevée que celle que l'on aurait pour des sous populations aléatoires  $\frac{26^4}{15} = 23$

La raison est la même qu'au paragraphe précédent : les sondages d'un carré, implantés à maille rigide dans ce carré sont entre eux en corrélation négative - d'où une majoration de la variance des variance du même ordre de grandeur qu'au paragraphe précédent.

#### CONCLUSION

La formule logarithmique de la covariance se vérifie bien pour Mehengui. Les petits écarts observés, peu significatifs au strict point de vue statistique, correspondent sans doute cependant à des phénomènes réels, mais d'amplitude assez faible pour que la formule puisse être utilisée quand même. Plutôt que la covariance, il convient d'étudier la variance  $\sigma^2 d$  des différences des teneurs des sondages distants de dm., et de construire la courbe  $\sigma^2 d = f(d)$ . La variance d'échantillonnage d'un gisement reconnu par n sondages à maille d est alors donnée par la formule théorique  $\sigma^2 = \frac{E 1}{2 n} f\left(\frac{d}{2.91}\right)$ . Pour une maille

de 200 m, à Mehengui, cette formule donne une variance d'échantillonnage de 0.62 en remarquable accord avec la valeur expérimentale qui est 0,596.

Mars 1960.

G. MATHERON.

En annexe : 3 figures.

# FIGURE I

Variogramme  $\sigma_d^2 = f(d)$  pour les carrés Z I et Z II

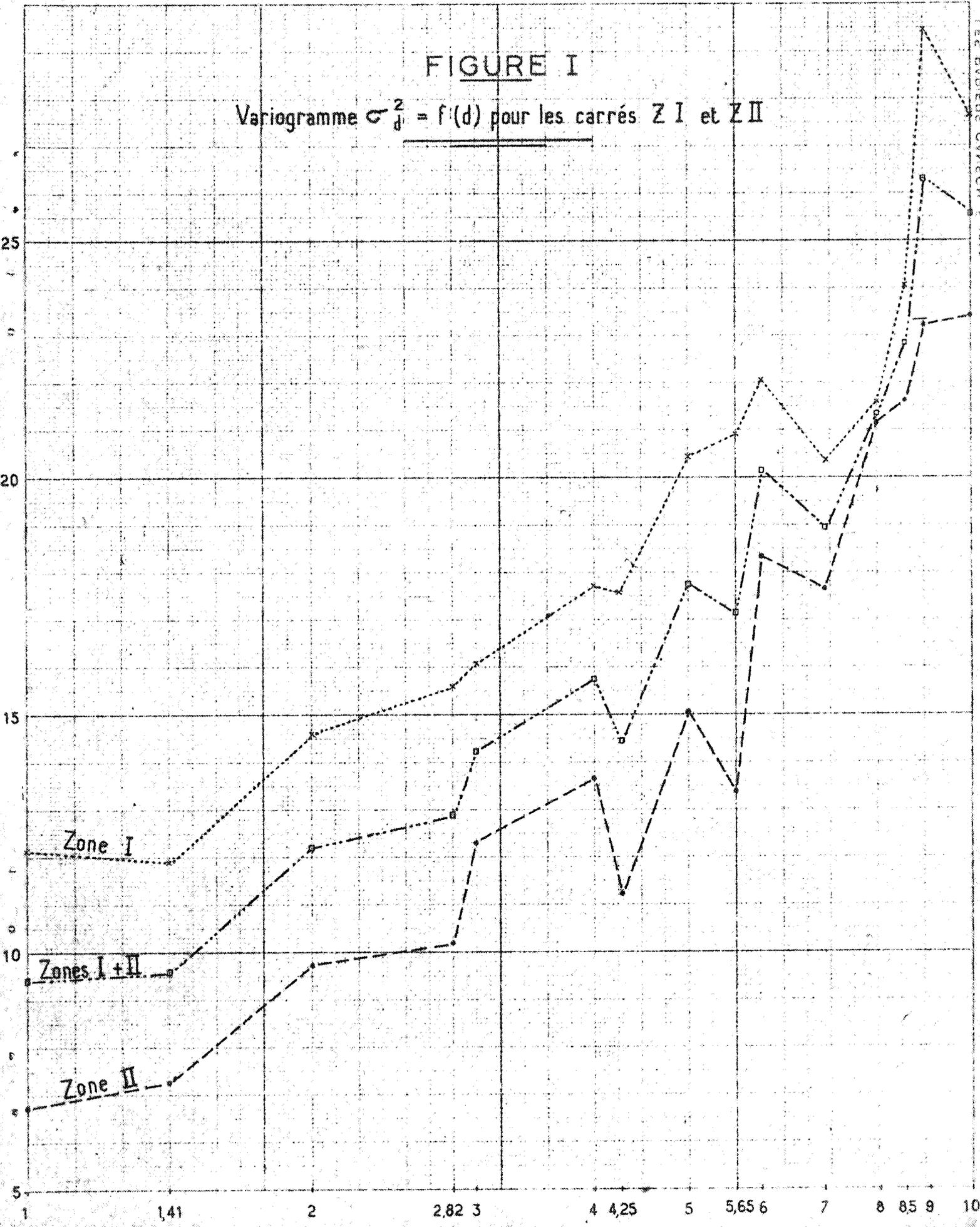
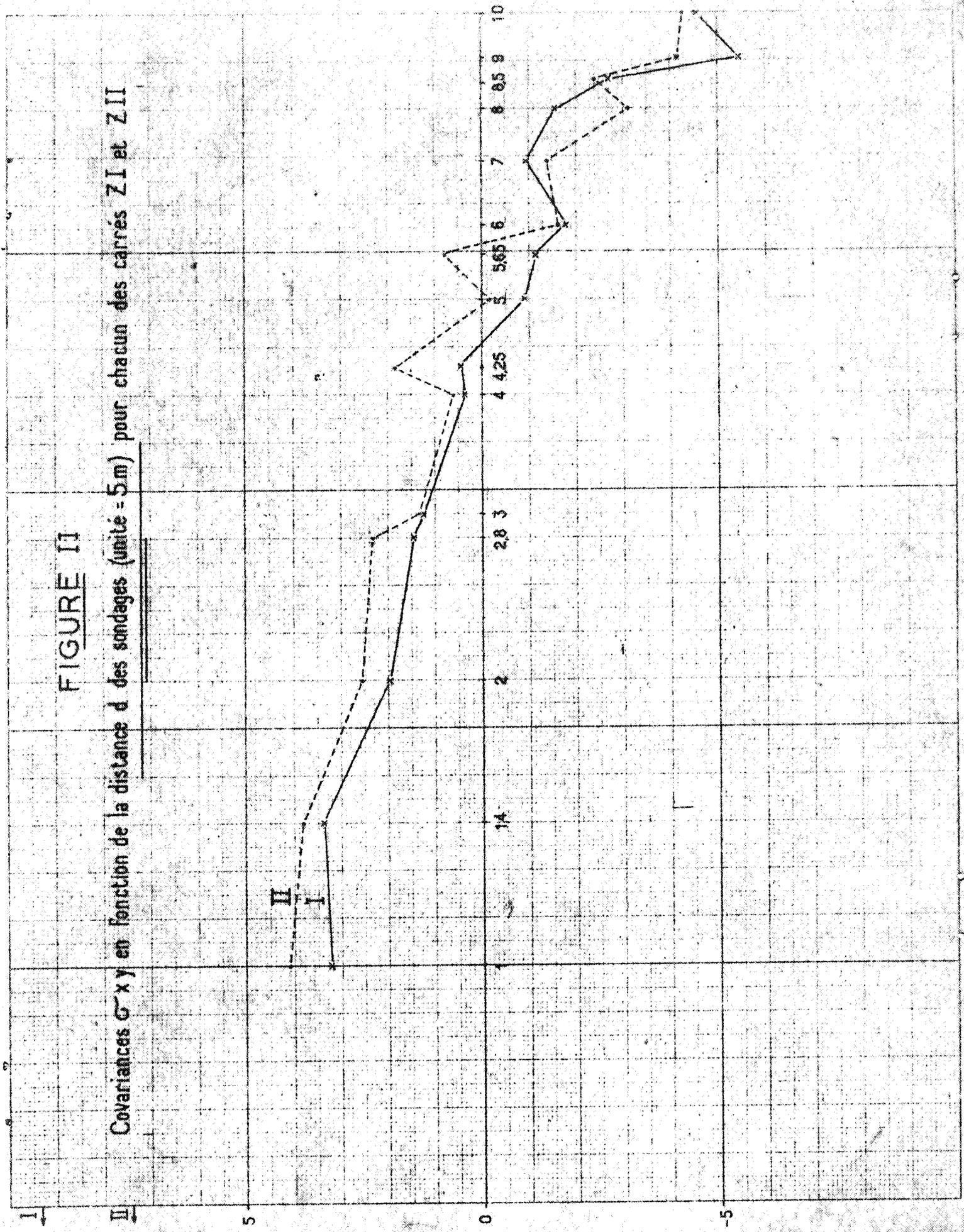


FIGURE II

Covariances  $\sigma_{xy}$  en fonction de la distance d des sondages (unité = 5m) pour chacun des carrés Z I et Z II



# FIGURE III - MEHENGUI

Variogramme  $\sigma_d^2 = f(d)$

En abscisse : distance d en échelle logarithmique  
Unité = 5 m.

En ordonnées : Variance  $\sigma_d^2$   
Interpolation par la droite

$$\sigma_d^2 = 7,9 + 1,33 \log_{10}(d) \quad (\log \text{ décimaux})$$

$$= 7,9 + 5,76 \log_e d \quad (\log \text{ népérien})$$

