



ETUDE GEOSTATISTIQUE DU GISEMENT

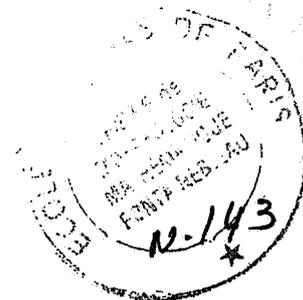
DE BAUXITE DE MEHENGUI

-----ooOoo-----

G. MATHERON

Mars 1960

ETUDE GEOSTATISTIQUE DU GISEMENT DE BAUXITE
DE MEHENGUI



-----ooOoo-----

LES DONNEES

Le gisement stratiforme de bauxite de Mehengui est analogue à celui de Minim Martap, avec des dimensions réduites. Il semble s'agir là aussi d'une couche de basalte bauxitisée. A la différence de Minim, où l'on n'a plus que des plateaux résiduels, la couche est restée ici d'un seul tenant. Le gisement a été reconnu par Péchiney 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.

Les données fournies par Péchiney, sur un schéma d'implantation, sont les teneurs du toit au mur en Al₂O₃ et les puissances de chaque sondage. En ce qui concerne les puissances, une annotation manuscrite nous prévient qu'elles ne sont pas trop représentatives, un certain nombre de sondages ayant dû, pour des raisons techniques, être arrêtés avant d'avoir traversé la couche.

BUT DE CETTE ETUDE

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

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

donnant la covariance σ_{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 500 m, soit un intervalle de variation de 1 à 100.

Les covariances seront celles des teneurs, et non de leurs logarithmes.

L'échelle logarithmique, pour des valeurs comprises en gros entre 35 % et 50 %, ne présenterait pas d'avantages particuliers. Pour éliminer l'effet de bordure, on calculera numériquement, non pas les covariances elles-mêmes, mais les variances σ_d^2 des différences $(x - y)$ des teneurs x et y des sondages distants de d . On a :

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

.../...

étant la variance des teneurs des sondages. L'expression σ^2 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 Al₂O₃ 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 $= 15.30$

Il est entendu que toutes les moyennes utilisées ici sont des moyennes non pondérées. Les variations 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 n 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 données disponibles. Un test χ^2 sur l'homogénéité de la distribution EW et NS, 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	28.6	30.3
7.06	161	129	290	31.4	33.6	32.4
8.49	137	100	237	34.9	31.8	33.7
9.90	119	73	192	38.2	33.8	37.0
11.30	103	52	155	28.3	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/le sont probablement pas, mais il est difficile de faire un test précis, ^{ne} 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	2.82	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 σ^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	17.8	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 isotropie 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 restant admissibles eu égard au petit nombre de données.

.../...

ZONE II

d Unité 5 m	Nombre de données			Variance $\sigma^2 d$		Ensemble
	N.S	E.W	NS + EW	NS	EW	
1	110	110	220	6.85	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	12.80	14.6	13.70
5	66	66	132	13.85	16.3	15.07
6	55	55	110	16.25	20.4	18.32
7	44	44	88	13.25	24.6	18.92
8	33	33	66	12.50	29.8	21.15
9	22	22	44	10.15	36.3	23.22
10	11	11	22	13.30	30.1	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			Variances $\sigma^2 d$		
	Zone I	Zone II	Ensemble	Zone I	Zone II	Ensemble
1.41	200	200	400	11.91	7.28	
2.83	162	162	324	15.60	10.20	
4.24	128	128	256	17.60	11.21	
5.65	98	98	196	20.90	13.34	
7.06	72	72	144	20.00	16.38	
8.49	50	50	100	24.04	21.62	
9.90	32	32	64	25.85	24.6	

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 identifiées aux points 7 et 10 de la première, et les données correspondantes groupées ensemble :

d Unité 5 m	Nombre de données			Variances $\sigma^2 d$		
	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	9.65	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	17.80	13.70	15.75
4.24	128	128	256	17.60	11.21	14.40
5	132	132	264	20.45	15.07	17.76
5.65	98	98	196	20.90	13.34	17.12
6	110	110	220	22.05	18.32	20.18
7	160	160	320	20.35	17.65	18.95
8	66	66	132	21.60	21.15	21.37
8.49	50	50	100	24.04	21.62	22.83
9	44	44	88	29.45	23.22	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 $\sigma^2 d$

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

σ_{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 $\sigma^2 d$ entre les deux zones proviennent uniquement de la différence des variances σ^2 des deux populations. Cette différence une fois imposée se répartit 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	308	15.75
4.24	256	14.4
5	264	17.76
5.65	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 somme 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 des 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

$$\sigma^2 d = 7,9 + 5,76 \log d \quad \left(\begin{array}{l} \text{log népérien} \\ d = \text{unité } 5 \text{ m} \end{array} \right)$$

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 :

$$\sigma^2 E = \frac{\sigma^2 - \sigma^2 z - \frac{\kappa}{2}}{n}$$

σ^2 - Variance des sondages dans le gisement

$\sigma^2 z$ - Variance dans le gisement du carré élémentaire de côté a

κ - Dispersion absolue.

Compte tenu de la formule logarithmique :

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

Il faut diviser par 2n 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

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

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

.../...

$$D'o\grave{u} \begin{cases} \text{Ecart type } \sigma_E = 0.785 \\ \text{Erreur possible (au niveau 95 \%) } \pm 2 \sigma_E = \pm 1.57 \end{cases}$$

Telles sont les pr\u00e9visions de la th\u00e9orie que nous allons maintenant comparer aux donn\u00e9es exp\u00e9rimentales.

2 - V\u00e9rification exp\u00e9rimentale.-

A partir des 300 sondages implant\u00e9s \u00e0 la maille de 50 m, nous pouvons choisir 16 sch\u00e9mas d'implantation possibles \u00e0 la maille 200. Pour chacun de ces sch\u00e9mas, on a calcul\u00e9 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 \u00e0 la maille 200, donc une bonne estimation de σ^2_E

Les 16 valeurs obtenues pour la moyenne s'\u00e9chelonnent entre 40.90 et 43.94 - soit dans un intervalle de 3 points. Il est int\u00e9ressant de voir que le nombre de sondages utiles varie entre 15 et 22, ce qui peut donner une id\u00e9e de la pr\u00e9cision avec laquelle le tonnage serait connu \u00e0 partir d'une maille 200. Les variances sont-elles aussi fortement dispers\u00e9es. Mais l'int\u00e9r\u00eat principal se porte sur la distribution des moyennes.

N\u00b0 du sch\u00e9ma	Nombre de donn\u00e9es	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 σ^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 avons 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. 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 théorique. Les 16 variances forment une population

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

$$\text{de variance } D^2(\sigma^2) = 51.4$$

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 négative. 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 les 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

.../...

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

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

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 $\frac{26^4}{15} = 23,1$

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 variances 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 10 m., 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 $\sigma^2 E \frac{1}{2n} f\left(\frac{d}{2.91}\right)$. 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

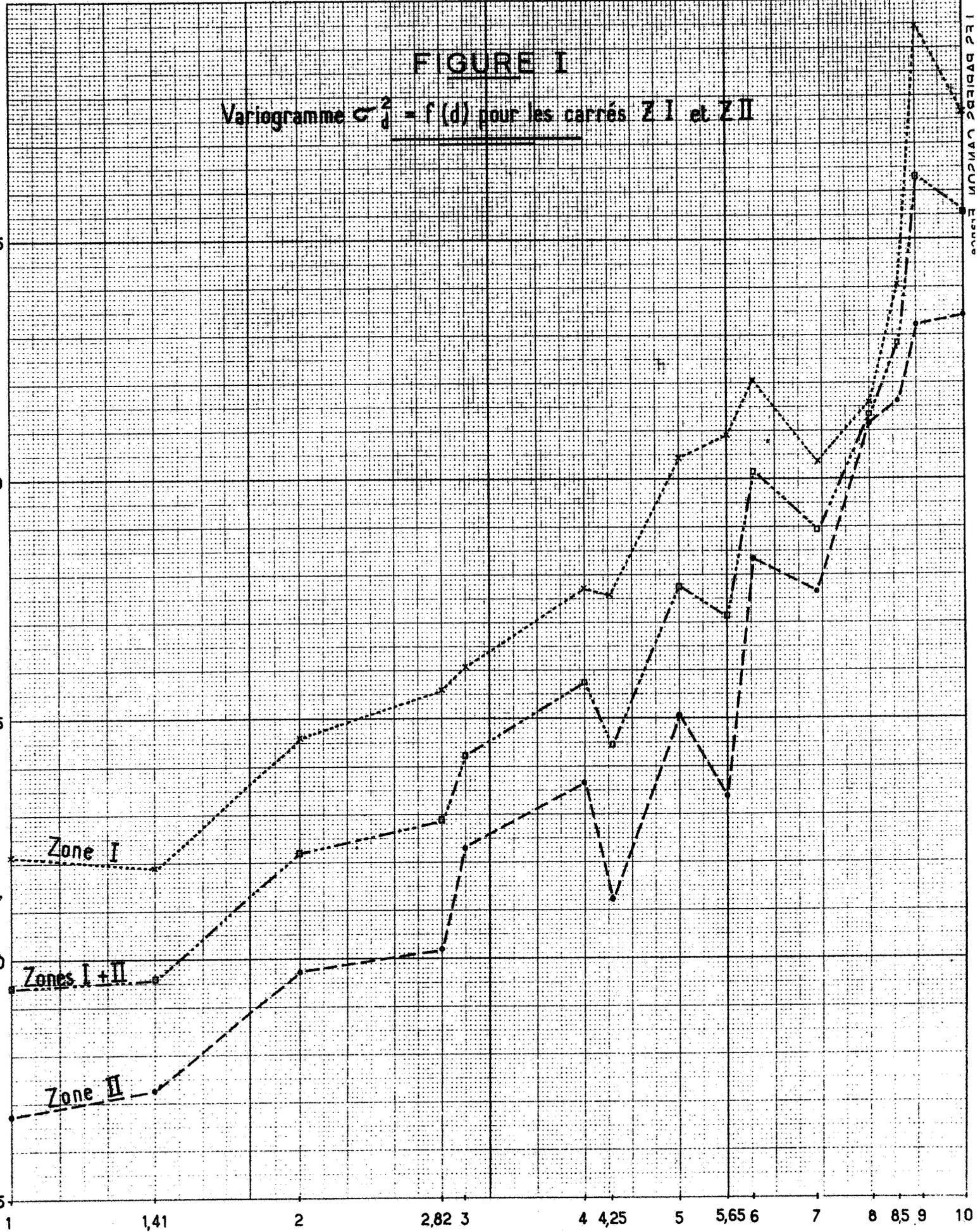


FIGURE II
 Covariances σ_{xy} en fonction de la distance d des sondages (unité = 5m) pour chacun des carrés ZI et ZII

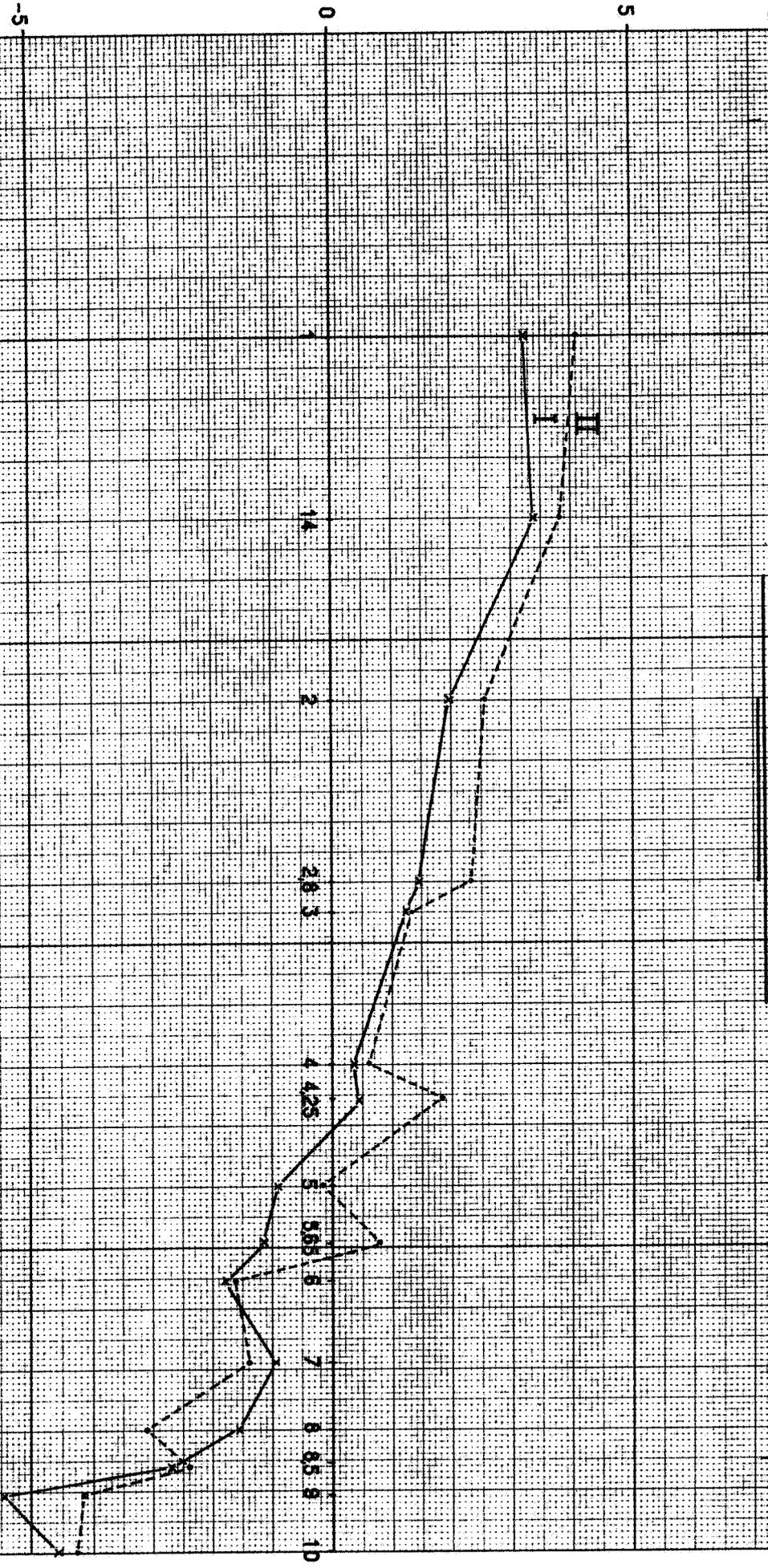


FIGURE III - MEHENGUI

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

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

En ordonnées : Variance σ^2
Interpolation par la droite

$$\sigma^2 = 7,9 + 1,53 \log_{10}(d) \quad (\log \text{ décimaux})$$

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

