

ETUDE DE LA LOI DE REPARTITION DES
LONGUEURS DES FELDSPATHS D'UNE DALLE DE
GRANITE

—ooOoo—

G. MATHERON
avril 1960

—ooO—

Sur une dalle de granite porphyroïde de l'Adrar Renaissance (Hoggar), M. THEBAULT a fait choix d'un grand carré de 10 m sur 10 m. Ce grand carré a été divisé en 100 carrés élémentaires de 1 m sur 1 m dans chacun desquels il a procédé à la mesure des longueurs des cristaux de feldspath. Au total, 5.627 mesures ont été rassemblées. Le but de ce travail était :

- 1 - De vérifier sur un très grand nombre de données si la loi de distribution de ces longueurs était log normale.
- 2 - D'examiner comment variaient les paramètres de cette distribution en fonction de l'aire échantillonnée, et si, en particulier il était possible d'étendre la notion de dispersion absolue à la pétrographie.

Le premier objectif a été atteint. En ce qui concerne le deuxième, la variation des paramètres a bien été mise en évidence, mais n'a pu être ramenée à une loi simple. L'influence de facteurs subjectifs, particulièrement l'ordre dans lequel les mesures ont été effectuées carré par carré, perturbe profondément les lois de variations et interdit toute conclusion.

I.- LES DONNEES.

Les mesures ont été effectuées carré par carré en progressant selon des lignes. Seules ont été mesurées sur l'affleurement les longueurs des feldspaths présentant une section longitudinale, de sorte que les mesures représentent en principe les plus grandes longueurs des cristaux eux-mêmes. Cependant, s'il n'y avait guère de confusion possible pour les grands cristaux, les petites sections posaient un problème plus délicat. Certaines d'entre elles devaient indéniablement correspondre à des sections transverses de grands cristaux, donc introduire des données parasites. Mais d'autres aussi devaient représenter des sections longitudinales de cristaux réellement petits, sans qu'il soit toujours possible de faire la distinction. Il est probable en fait que l'on a introduit un certain nombre de petites valeurs parasites.

Au total 5.627 mesures ont été effectuées. Au tableau I, les résultats relatifs à chacun des 100 carrés sont résumés par trois chiffres : le nombre de mesures effectuées, la médiane de ces mesures exprimées en millimètre et la variance. Cette variance est la variance des log népériens; elle n'a subi ni correction de biais, ni correction de Sheppard. Le tableau II présente l'histogramme de l'ensemble des 5.627 mesures effectuées dans les 100 carrés.

TABLEAU I

Paramètre des carrés élémentaires

Premier chiffre : nombre de mesures
Deuxième " : médiane en mm
Troisième " : variance

69 23,2 0,0665	71 24,4 0,078	66 24,2 0,0855	77 25,8 0,066	65 25,7 0,0685	71 26,0 0,0746	99 22,4 0,097	77 24,7 0,0652	68 26,9 0,0719	84 25,3 0,0623
61 23,3 0,095	60 23,4 0,127	59 22,6 0,117	61 22,4 0,158	59 24,2 0,0755	57 23,5 0,119	63 22,0 0,109	63 23,6 0,083	63 23,2 0,084	56 26,0 0,101
55 25,6 0,0745	63 24,3 0,0775	57 23,9 0,078	62 26,0 0,0654	55 26,0 0,108	58 27,2 0,080	61 26,8 0,075	63 24,9 0,107	63 23,4 0,069	63 24,4 0,0935
56 26,6 0,0805	45 28,4 0,098	51 27,0 0,103	53 29,8 0,1075	57 27,5 0,0944	56 29,4 0,096	53 29,9 0,0965	61 27,4 0,0965	53 28,4 0,089	57 27,9 0,0695
45 25,5 0,106	52 27,7 0,0668	41 26,5 0,057	63 21,8 0,0985	63 21,6 0,0791	81 23,0 0,0696	63 23,0 0,102	62 21,7 0,0745	63 22,8 0,1051	63 23,6 0,0629
45 27,8 0,080	53 27,2 0,0768	52 24,1 0,071	53 27,6 0,0905	54 26,8 0,085	52 30,8 0,0565	55 30,2 0,0732	53 30,3 0,096	52 30,0 0,0857	52 28,7 0,128
53 24,6 0,0729	48 24,3 0,063	53 25,2 0,0653	51 25,8 0,0666	52 24,3 0,0835	46 26,6 0,0765	52 29,0 0,080	55 27,4 0,0725	52 29,7 0,1035	54 26,9 0,079
53 26,6 0,0458	54 26,0 0,0420	34 26,1 0,0865	44 26,5 0,0805	52 26,2 0,0582	53 26,6 0,0758	55 25,9 0,0585	52 27,0 0,0715	53 27,1 0,0613	54 27,9 0,0556
41 25,3 0,0752	51 28,4 0,0624	49 26,5 0,0731	39 29,4 0,0759	53 25,2 0,0653	53 27,7 0,0857	52 29,4 0,063	53 26,6 0,0958	53 25,0 0,0572	53 26,8 0,0798
54 27,8 0,0576	45 25,4 0,0578	52 26,6 0,0794	53 26,5 0,0803	46 26,8 0,0615	51 27,4 0,072	53 26,8 0,0593	47 26,3 0,0695	49 27,2 0,0562	53 27,3 0,040

II.- AJUSTEMENT LOG NORMAL DES 5.627 MESURES.-

Disposant d'un grand nombre de données, nous pouvons tester de manière très précise la lognormalité de la répartition du tableau II du moyen d'un test χ^2 . Toutefois l'examen direct de l'histogramme montre que l'attraction psychologique des chiffres "ronds" a exercé un effet très important. Les valeurs 20, 25, 30 etc ... de l'histogramme correspondent à des pointes très fortes, et les valeurs immédiatement inférieures : 19, 24, 29, 34 etc ... à des creux également très prononcés. De même, quoique dans une moindre mesure, les valeurs 16, 22, 32, etc ... donnent des pointes, les valeurs 23, 33, etc ... ne correspondant cependant pas à des creux visibles par suite de la proximité des valeurs encore plus deshéritées : 24, 34 ...

Pour éliminer autant que possible cet effet, les données ont été groupées par classes de longueur 5 centrées sur les multiples de 5 :

Classe	Nombre de mesures
8-12	64
13-17	550
18-22	1 248
23-27	1 566
28-32	1 130
33-37	560
38-42	279
43-47	136
48-52	53
53-57	21
58-62	12
63-67	4
78-82	1

Ce groupement élimine l'influence de l'attraction des multiples de 5, mais laisse subsister l'influence, moindre il est vrai, de l'attraction des valeurs : 22, 32 ...

Puisque l'emploi de classes arithmétiques nous est imposé, il est naturel de calculer les paramètres arithmétiques de cette distribution (moyenne et variance arithmétiques) et d'en déduire les paramètres logarithmiques correspondants par les formules théoriques de la loi log normale. L'étape suivante consistera à tester l'ajustement de la distribution empirique avec la distribution log normale possédant les paramètres ainsi calculés.

On trouve pour les paramètres arithmétiques :

$$\begin{aligned} \text{moyenne } m &= 26.441 \\ \text{variance } \sum^2 &= 65.536 \end{aligned}$$

.../

A la variance ainsi calculée, il convient de retirer le terme correctif de Sheppard (correction pour groupement), égal à :

$$\frac{h^2}{12}, h \text{ étant l'intervalle de classe : } \frac{h^2}{12} = \frac{25}{12} = 2.083. \text{ Il vient :}$$

$$\begin{cases} \bar{m} = 26.441 \\ \sum^2 = 63.453 \end{cases}$$

On remarque que cette correction de Sheppard revient implicitement, en ce qui concerne le calcul de la variance, à éliminer l'effet de la concentration préférentielle des données sur les multiples de 5. De ces paramètres arithmétiques, on déduit ce que devraient être les paramètres logarithmiques, si la loi était log normale :

$$\begin{cases} \text{Variance} & : \sigma^2 = L(1 + \frac{\sum^2}{\bar{m}^2}) = 0.086876 \\ \text{Ecart-type} & : \sigma = 0.29475 \\ \text{Médiane} & : \gamma = \bar{m} e^{-\frac{\sigma^2}{2}} = 25.318 \end{cases}$$

Appliquons le test χ^2 à l'ajustement de la distribution observée à la distribution log normale ayant les paramètres précédents

.../

TABLEAU III

Test χ^2

Classe	Fréquence théorique n_t	Fréquence observée n_o	$\frac{(n_o - n_t)^2}{n_t}$
0 - 12,5	46,83	64	6,28
12,5 - 17,5	544,60	550	0,05
17,5 - 22,5	1 347,30	1 248	7,43
22,5 - 27,5	1 495,60	1 566	3,34
27,5 - 32,5	1 075,80	1 130	2,73
32,5 - 37,5	602,77	562	2,76
37,5 - 42,5	292,03	279	0,58
42,5 - 47,5	129,70	136	0,31
47,5 - 52,5	54,72	53	0,05
52,5 - 57,5	22,415	21	0,09
57,5 -	15,235	18	0,50
TOTAL	5 627,000	5 627	24,12

.../

Les trois dernières classes ont été groupées en une seule pour avoir dans chaque classe le minimum de 10 données requis en pratique pour l'application du test.

La valeur observée $\chi^2 = 24,12$ correspond à une probabilité de $P = 0,00229$, soit 2,3 chances sur mille seulement, avec $11 - 3 = 8$ degrés de liberté, pour qu'une distribution log normale ayant les paramètres calculés se traduise, par tirage au sort de 5.627 échantillons, par une répartition numérique possédant un tel χ^2 . Ceci équivaut à un rejet de la lognormalité.

Cependant, dans la colonne de droite du tableau II, nous observons que les données apportant au χ^2 la plus forte contribution sont celles des classes :

0 - 12,5	=	6,28
17,5 - 22,5	=	7,43
22,5 - 27,5	=	3,34
		17,05

avec une contribution de 17,05 sur un χ^2 de 24,12. Or :

- 1 - Nous savons que, sur le terrain, des sections transverses de feldspaths grands ou moyens ont été prises pour des sections longitudinales de petits feldspaths. Nous avons donc introduit, dans les petites classes, un nombre plus grand de données qu'il n'y en avait réellement. Or, justement, la classe 0 - 12,5 contient 64 observations contre une prévision théorique de 46,83, soit un excès de 17 environ responsable de 6,28 dans la somme du χ^2 . Il est fort possible, en effet, qu'on ait introduit 17 mesures parasites de sections transverses dans cette classe. Pour éliminer cette cause très réelle d'erreur, il convient d'ajuster seulement les observations supérieures à 12,5 à la loi lognormale tronquée à 12,5.
- 2 - Le groupement des données par 5 a éliminé l'influence de l'attraction psychologique des multiples de 5, mais non pas celle des valeurs 22, 32, etc ... avec leur contrepartie négative sur les classes 23, 33, etc ... Les deux classes apportant une importante contribution au χ^2 , en dehors de la classe $< 12,5$, ont justement l'une comme limite supérieure la valeur 22, l'autre la valeur 23. Sur l'histogramme du tableau II nous lisons :

322 valeur 22
277 valeur 23

Comme on est dans une partie croissante de l'histogramme, cela indique qu'au moins une vingtaine de mesures légèrement supérieures à 23 ont été classées à la valeur 22. Si l'on admettait, par exemple :

299 valeur 22
300 valeur 23

.../

On aurait :

<u>Classe</u>	<u>Fréquence théorique</u>	<u>Fréquence observée</u>	<u>χ^2</u>
17,5 - 22,5	1 314,30	12,48	3,34
22,5 - 27,5	1 518,60	15,66	1,82

soit une contribution au χ^2 de 3,34 et 1,82 au lieu de 7,43 et 3,34. Il va de soi qu'on n'a pas le droit de se livrer à une opération aussi arbitraire. Par contre, puisque nous avons des raisons solides de penser qu'un certain nombre de valeurs de la classe supérieure ont été rangées par erreur dans la classe inférieure, nous avons le droit, pour éliminer cette cause d'erreur, de grouper ensemble les données de ces deux classes, en supprimant naturellement un degré de liberté pour le χ^2 .

On pourrait remarquer que les deux classes suivantes 27,5-32,5 et 32,5-37,5, contiennent, l'une la valeur favorisée 32 et un excès de valeurs sur la prévision théorique, l'autre la valeur défavorisée 33 et un défaut sur la prévision théorique, avec pour ces deux classes une contribution totale au χ^2 de 5,49. Cependant, nous ne les grouperons pas.

.../

TABLEAU IV

Test χ^2 après tronquage à 12.5 et fusion des classes 17.5 - 22.5
et 22.5 - 27.5

Classe	Fréquence théorique n_t	Fréquence observée n_o	$\frac{(\mathbf{n_t} - \mathbf{n_o})^2}{\mathbf{n_t}}$
12.5 - 17.5	544.60	550	0.05
17.5 - 27.5	2 842.90	2 814	0.29
27.5 - 32.5	1 075.80	1 130	2.73
32.5 - 37.5	602.77	562	2.76
37.5 - 42.5	292.03	279	0.58
42.5 - 47.5	129.70	136	0.31
47.5 - 52.5	54.72	53	0.05
52.5 - 57.5	22.415	21	0.09
57.5 -	15.235	18	0.50
			$\chi^2=7.36$

Le χ^2 atteint une valeur de 7.36 pour $9-3 = 6$ degrés de liberté, correspondant à une valeur $P = 0.285$. Si la loi était réellement lognormale avec les paramètres admis, il y aurait 28,5 chances sur cent pour que la valeur observée du χ^2 dépasse 7.36. Le test conclut en faveur de la lognormalité.

On remarquera bien que les deux opérations auxquelles nous nous sommes livrés

- tronquage à 12.5 pour mettre hors jeu les petites valeurs
- fusion des classes 17.5 - 22.5 et 22.5 - 27.5

ne sont pas arbitraires. Leur but était d'éliminer des causes d'erreur d'origine subjective (petites sections et répulsion psychologique pour la valeur 23) dont la réalité est indéniable et qui déformaient réellement la représentation du phénomène naturel.

III - Les lois de variations des paramètres

Nous étudierons successivement :

- 1 - Le variogramme des médianes dans les lignes et les colonnes
- 2 - La loi d'évolution de la variance des longueurs des feldspaths distribués dans les carrés 1, 2, 4 et 10 m.

.../...

- 3 - Les lois de distribution, dans le carré 10 x 10, des médianes des longueurs des feldspaths distribuées dans les carrés de côté 1, 2 et 4.

Ce sont là trois façons différentes d'essayer de dégager une dispersion absolue ou, si l'on veut, d'examiner si la loi de Wijs et le principe de similitude s'appliquent à la distribution des longueurs des feldspaths.

1 - Le variogramme des médianes -

Nous cherchons en premier lieu si la covariance (ou la corrélation) des longueurs des feldspaths distants de d sur la dalle est une fonction régulière de la distance d . En pratique, il était exclu de relever sur un plan la longueur et la position de chaque feldspath. Nous sommes réduits à examiner la corrélation des médianes des feldspaths dans les carrés élémentaires.

En fait, cette limitation n'est pas gênante. Nous cherchons en fait si la covariance σ_{xy} des longueurs x et y des feldspaths distants de d est de la forme

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

La covariance des médianes de deux carrés distants de d est en fait la valeur moyenne de la covariance de chaque feldspath du carré 1 avec chaque feldspath du carré 2. Si ces covariances sont fonction logarithmique de la distance, on démontre que leur valeur moyenne entre les deux carrés est égale à la covariance des longueurs des feldspaths placés au centre de chacun des carrés, c'est-à-dire distants de d : les covariances des médianes coïncident avec celles des feldspaths eux-mêmes.

Au lieu des covariances elles-mêmes, nous calculerons le variogramme, c'est-à-dire la courbe

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

donnant la variance du rapport γ_1/γ_2 des médianes des carrés distants de d . On sait que le $\sigma^2 d$ est lié à la covariance σ_{xy} correspondante par la relation

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

σ^2 étant la variance des médianes elles-mêmes.

Le variogramme est établi de mètre en mètre (jusqu'à 9 m.) séparément dans les lignes et les colonnes.

Les tableaux V et VI fournissent les résultats.

.../...

TABLEAU V
Variogramme dans les lignes

Distance d	Nombre de données	Variance $\sigma^2 d$
1	90	0.469 10^{-2}
2	80	0.589 10^{-2}
3	70	0.729 10^{-2}
4	60	0.666 10^{-2}
5	50	0.754 10^{-2}
6	40	0.884 10^{-2}
7	30	0.855 10^{-2}
8	20	0.764 10^{-2}
9	10	0.409 10^{-2}

TABLEAU VI
Variogramme dans les colonnes

Distance d	Nombre de données	Variance $\sigma^2 d$
1	90	1.610 10^{-2}
2	80	1.393 10^{-2}
3	70	1.342 10^{-2}
4	60	1.761 10^{-2}
5	50	1.583 10^{-2}
6	40	1.070 10^{-2}
7	30	1.388 10^{-2}
8	20	1.748 10^{-2}
9	10	0.890 10^{-2}

La figure I montre graphiquement les variogrammes des lignes et des colonnes. L'échelle des distances d est logarithmique. Le contraste des deux courbes est frappant :

.../...

Le variogramme des lignes présente des valeurs faibles du $\sigma^2 d$ et une allure très régulière. Les points s'alignent de façon satisfaisante jusqu'à la valeur 7 de la distance. Il convient de ne pas attacher d'importance à la chute qui intervient aux valeurs 8 et 9. Les variances correspondantes, calculées à partir de 20 et 10 données seulement, n'ont qu'une valeur très relative, et la chute n'a rien de significatif.

Le variogramme des colonnes frappe au contraire par son caractère incohérent et la forte valeur moyenne du $\sigma^2 d$. Il n'y a pas, dans les colonnes, de réseau de corrélations bien structurées. En fait, la valeur moyenne des 9 valeurs du $\sigma^2 d$ des colonnes est 1.42×10^{-2} . Elle ne diffère pratiquement pas du double, 1.39×10^{-2} de la variance des médianes (0.695×10^{-2}). On a l'impression qu'il n'y a aucune corrélation dans les colonnes et que les $\sigma^2 d$ fluctuent au hasard autour de la valeur 1.39×10^{-2} correspondant à une covariance nulle.

Ce contraste entre une structure presque parfaite selon les lignes et une incohérence presque totale selon les colonnes mérite réflexion.

- On peut penser, en premier lieu, selon une suggestion de M. THEBAULT, à une influence possible de la zonalité du granite. La dalle est située à proximité du contact du granite porphyroïde, et ce contact est, en gros, parallèle aux lignes, d'où zonalité dans cette direction. Cependant, il semble que si une telle zonalité crée bien une anisotropie, celle-ci doit se traduire par une allure différente des deux variogrammes, la croissance du $\sigma^2 d$ étant plus rapide dans les colonnes que dans les lignes, mais chacun d'eux devrait rester structuré. On voit mal pourquoi les relations seraient détruites dans la direction perpendiculaire à la zonalité.

- On peut craindre, en réalité, que le contraste des variogrammes ne soit dû à des causes entièrement subjectives. Les mesures ont été faites carré par carré en progressant selon des lignes. Il est possible, pour des raisons subjectives, que la façon de travailler se modifie lentement d'heure en heure, au cours d'une journée de travail et surtout de jour en jour. Les carrés d'une ligne sont mesurés à quelques heures d'intervalle seulement, ceux d'une même colonne à plusieurs jours. Dans le premier cas, la méthode de travail s'est peu modifiée dans l'intervalle, dans le deuxième, on a affaire à des mesures peu comparables entre elles. D'où réseau régulier de corrélations et allure logarithmique du variogramme dans les lignes, et incohérence ou absence de corrélations dans les colonnes. Si cette explication était la bonne, il existerait réellement une dispersion absolue, et c'est le variogramme des lignes qui en donnerait la meilleure image. Mais alors nos paramètres globaux (médianes et variances) seraient profondément altérés par la lente modification des conditions de travail et ne représenteraient plus des propriétés de la distribution réelle.

entre
Il n'est en fait pas possible de trancher/les deux explications, zonalité ou facteurs subjectifs, sans recourir à de nouvelles mesures. L'attention en tout cas aura été attirée sur l'importance des facteurs subjectifs sur le résultat d'opérations en apparence aussi précises et objectives que des mesures de longueurs : attraction des chiffres ronds, tout d'abord, et maintenant, plus profondément, perturbation du réseau de corrélation par une évolution imperceptible des conditions subjectives du travail.

.../...

2 - Variance des longueurs des Feldspaths dans les carrés de côté 1, 2, 4 et 10

Carrés de côté 1 -

Les paramètres des distributions des longueurs des feldspaths dans les carrés élémentaires ont été réunis dans le tableau I. Il serait intéressant d'étudier les variances des variances et, en particulier, d'examiner s'il existe un réseau de corrélation pour les variances comme pour les médianes. Ce travail n'a pas été entrepris en raison du doute qui pèserait sur les corrélations découvertes du fait de l'influence de facteurs subjectifs.

Sur cent carrés, la moyenne arithmétique des variances est 0.0796. Il convient de faire subir à cette valeur une correction de biais et la correction de Sheppard. On trouve

$$\sigma^2_1 = 0.0766$$

Carrés de côté 2 -

On a calculé les variances des longueurs des feldspaths dans les 81 carrés de côté 2 que l'on peut former avec les 100 carrés élémentaires. Après les corrections de biais et de Sheppard, la moyenne de ces variances est :

$$\sigma^2_2 = 0.0826$$

Carrés de côté 4 -

On a procédé de même pour les 49 carrés de côté 4. On trouve toutes corrections faites :

$$\sigma^2_4 = 0.0840$$

Carré de côté 10 -

La répartition des longueurs des feldspaths dans le grand carré a déjà été étudiée. Elle a pour variance :

$$\sigma^2_{10} = 0.0869$$

La figure II présente ces variances en fonction du côté a du carré ($a = 1, 2, 4$ et 10). L'échelle des a étant logarithmique, ces 4 points s'aligneraient si le réseau de corrélations logarithmique régnait sur les distributions. Les écarts sur cet alignement sont dus aux mêmes causes (subjectives ou non) que les perturbations du variogramme des colonnes. Mais ils sont moins graves que sur le variogramme, et il est peut-être possible de tracer ici une droite moyenne, donc de définir une dispersion absolue. On hésite cependant entre une droite vraiment moyenne (D 1 sur la figure), et une droite passant par les trois derniers points (D 2 sur la figure). Les dispersions absolues correspondantes seraient $\frac{1.8}{1000}$ et $\frac{1.3}{1000}$.

.../...

3 - Distribution des médianes des carrés de côtés 1, 2 et 4

Les variances des médianes des carrés de côtés 1, 2 et 4 ont été calculées :

$$\sigma^2_1 = 7.01 \cdot 10^{-3}$$

$$\sigma^2_2 = 2.94 \cdot 10^{-3}$$

$$\sigma^2_4 = 0.97 \cdot 10^{-3}$$

La figure III représente la courbe correspondante en échelle logarithmique. Compte tenu du fait que la variance de la médiane du carré de côté 10 doit s'annuler, on obtient une contre épreuve de la figure II, avec les mêmes difficultés pour l'interpolation linéaire et la même interprétation de ces difficultés.

CONCLUSION -

Ce travail a permis d'établir la lognormalité de la distribution des longueurs des feldspaths. Il a mis en évidence l'influence de la dimension de l'aire échantillonnée sur les paramètres de dispersion, sans cependant permettre de dégager, de façon sûre, la notion d'une dispersion absolue. Cependant, ce demi-échec pourrait être imputable à l'influence de facteurs subjectifs, ou plus précisément à une lente et imperceptible évolution dans le temps des conditions subjectives des opérations de mesure exécutées sur le terrain, qui perturberait le réseau des corrélations réelles et en voilerait l'image. Nous proposons, en vue de l'application ultérieure de la géostatistique à la pétrographie, de recourir à des données dont la mesure serait moins soumise à de tels facteurs.

G. MATHERON

FIGURE 1

Variogramme des longueurs des fieldspaths

FIGURE I

Variogramme des longueurs des feldspaths

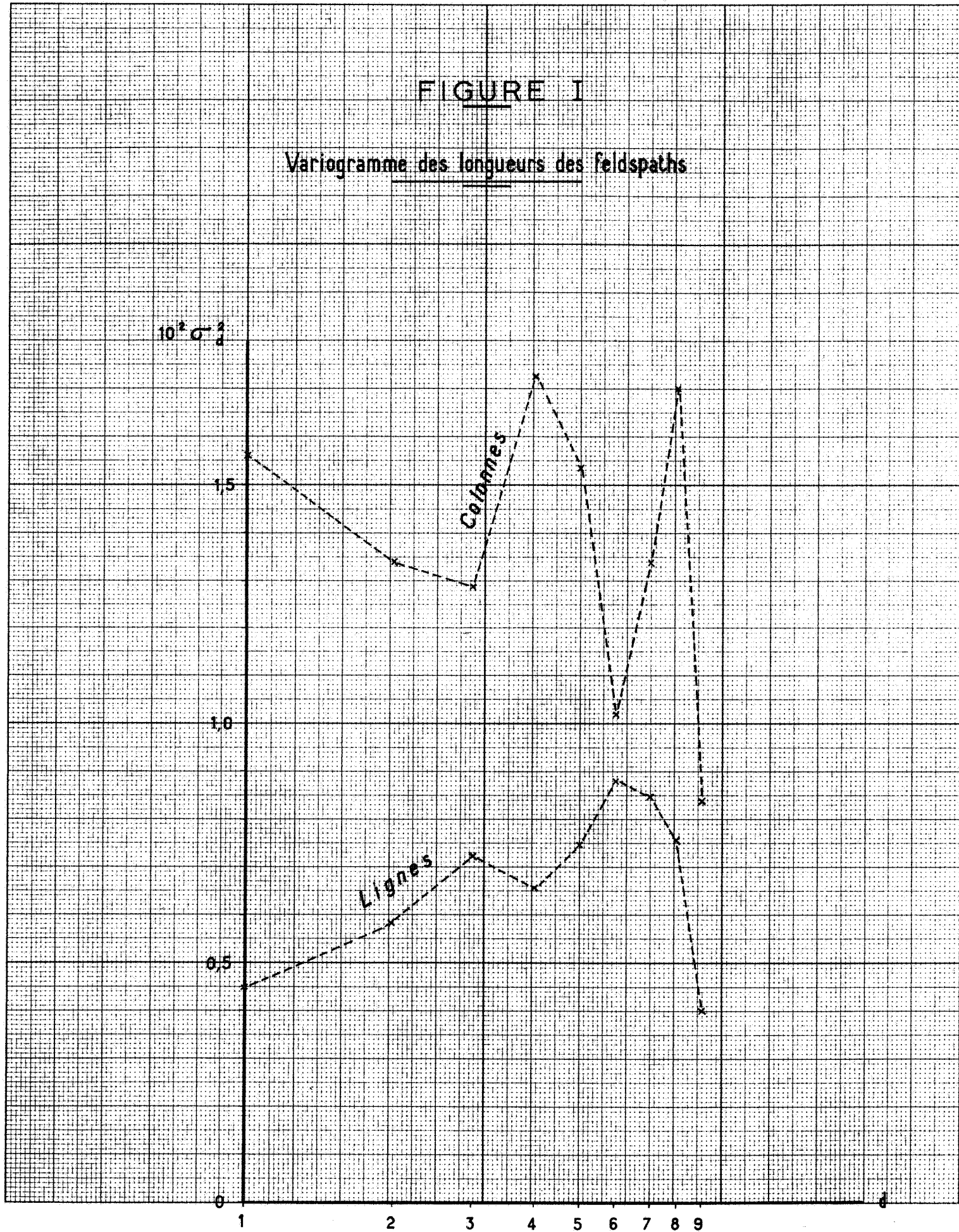


FIGURE II

Variance des longueurs des feldspaths dans les carrés de côté a

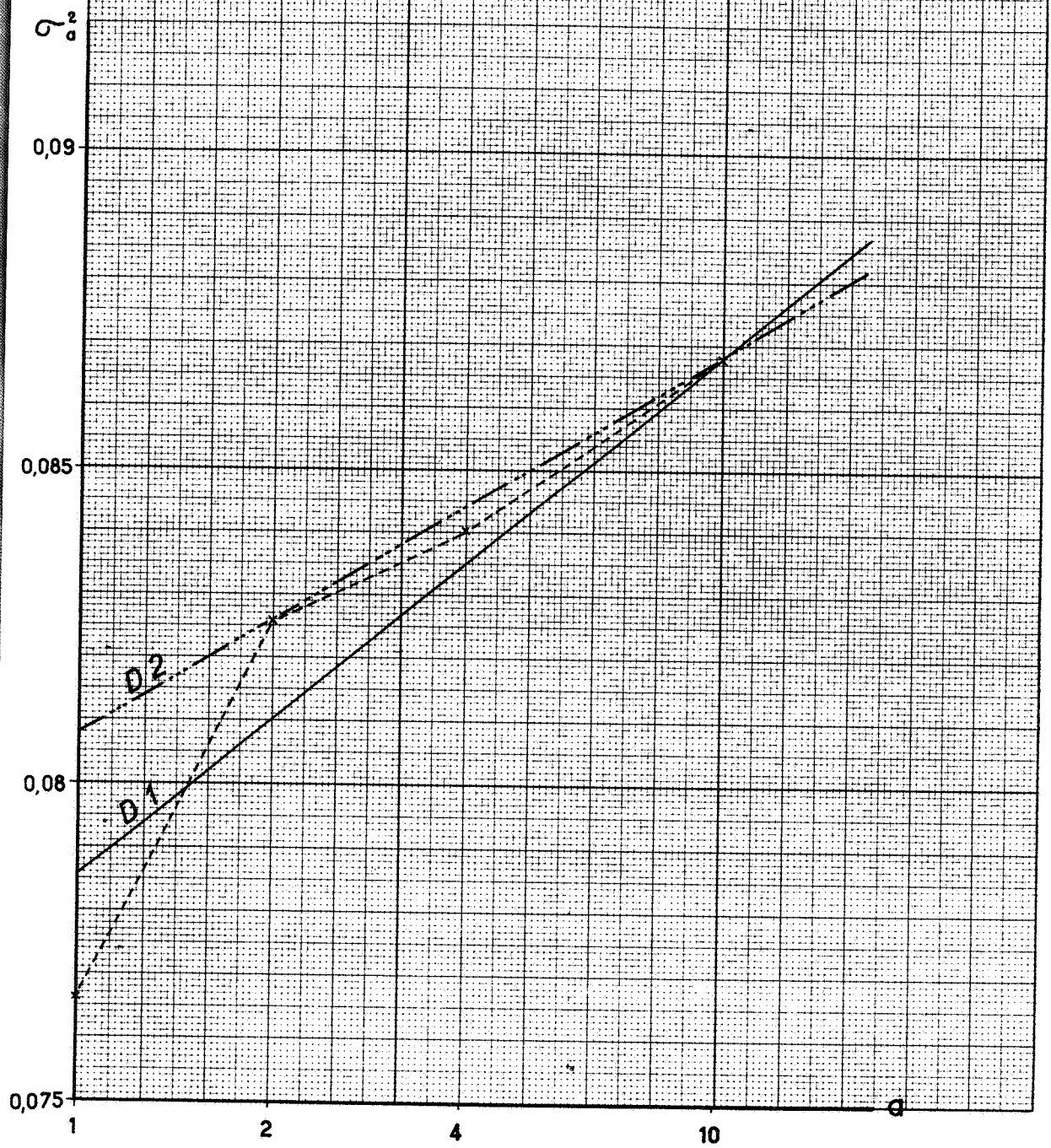


FIGURE III

Variance des medianes des carrés de côté 1, 2 et 4.

