

Note Géostatistique n° 25

---

Etude Expérimentale de l'effet de bordure dans l'estimation d'une  
aire minéralisée.

--:--:--:--:--:--:--:--:--:--

Le problème de l'effet de bordure est celui de la représentation de l'erreur commise lorsque l'on estime l'aire d'une surface minéralisée  $S$  à partir d'un nombre  $n$  de sondages positifs ponctuels implantés dans  $S$  selon une maille régulière, par exemple une maille carrée de côté  $a$ . Conformément au caractère géométrique pur d'un tel problème, nous admettons qu'il existe un critère simple, permettant de juger sans aucune ambiguïté si un sondage donné est positif ou négatif. Par exemple un sondage sera dit positif s'il rencontre une certaine formation géologique, et négatif s'il ne la rencontre pas.<sup>1</sup> Le problème est alors d'interpoler entre les derniers sondages positifs et les premiers sondages négatifs une ligne continue, représentant au mieux la frontière réelle inconnue de l'aire minéralisée. La marge d'incertitude est parfois considérable. D'autre part, la position du point frontière entre chaque couple positif/négatif n'est pas indépendante de la position adoptée pour les couples voisins : il y a une cohérence géométrique, qui rend solidaires les tracés des différentes portions du contour, et qui ne permet pas de représenter de façon simple l'erreur commise dans l'estimation de la surface, par la théorie

1. Il arrivera souvent que le critère employé en pratique ne présente pas ce caractère simple, par exemple s'il s'agit d'un critère d'exploitabilité. On voit facilement qu'en pareil cas le problème n'est plus de nature géométrique. Il ne s'agit plus, alors, d'interpoler entre des sondages positifs et négatifs la frontière d'un phénomène géométrique ou géologique réel, mais de choisir sur plan, au vu des résultats quantitatifs de sondages infra ou supra marginaux, les zones que l'on décide d'exploiter effectivement. Il n'y a plus d'effet de bordure, puisque l'on connaît parfaitement l'aire que l'on a soi-même décidé d'exploiter. L'erreur se reporte alors toute entière sur l'estimation de la teneur de la surface retenue (cette surface peut très bien, en effet, se trouver englober des enclaves inframarginales). Le lecteur pourra se référer à mon "Traité de Géostatistique Appliquée". Le critère géométrique simple postulé ici sera utilisable en pratique chaque fois que le gisement sera du type tout ou rien : charbon, potasse, gisement métallique riche dans un encaissant pratiquement stérile etc....

classique du calcul des probabilités. Il n'est du reste pas possible d'assimiler l'aire  $S$  à une variable aléatoire, puisque  $S$ , bien qu'inconnue, possède une valeur physiquement déterminée et nullement aléatoire.

D'un point de vue simplement qualitatif, on sait qu'il est possible de déterminer géométriquement deux surfaces  $S_0$  et  $S_p$  entre lesquelles la surface réelle  $S$  est presque nécessairement comprise. Les contours de  $S_0$  - surface limite optimiste - et  $S_p$  - surface limite pessimiste se construisent en joignant les premiers sondages négatifs ou les derniers sondages positifs par des segments de droite parallèles aux deux directions principales de la maille. Sur la figure 1, la surface doublement hachurée four-

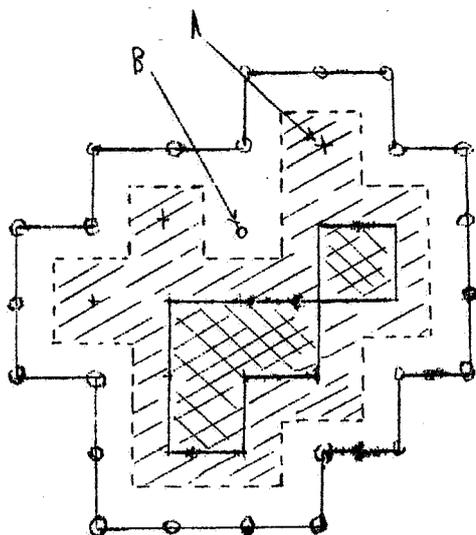


Figure 1.

nit un exemple de surface  $S_p$ , tandis que le contour extérieur limite la surface  $S_0$ .

On remarque qu'un sondage positif en flèche tel que A reste extérieur à  $S_p$ , tandis qu'un sondage négatif isolé tel que B est sans influence sur  $S_0$ .

Entre ces deux limites extrêmes  $S_0$  et  $S_p$ , on obtient une estimation moyenne  $S_m$  en attribuant à chaque sondage positif sa zone d'influence, qui est un carré de côté  $a$  égal à la maille. (hachures simples sur la figure). On a  $S_m = n a^2$ ,  $n$  étant le nombre de sondages positifs. Si l'on désigne par  $2p$  le périmètre de la surface  $S_m$ , on a les égalités très simples suivantes :

$$(I) \quad \begin{cases} S_p = (n + 1)a^2 - p a \\ S_m = n a^2 \\ S_0 = (n + 1) a^2 + p a \end{cases}$$

Sur l'exemple de la figure 1, avec  $n = 14$  et  $2p = 22$ , on trouve  $S_p = 4a^2$ ,  $S_m = 14 a^2$  et  $S_0 = 26 a^2$ . On peut dire que la surface réelle  $S$  est nécessairement comprise entre 4 et  $26 a^2$ , avec une valeur probable égale à  $14 a^2$ . La fourchette ainsi obtenue est certainement beaucoup trop large pour les besoins de la pratique, et il convient d'essayer de la resserrer. Mais les formules I mettent en évidence deux conclusions qualitatives importantes :

- l'ordre de grandeur de la valeur probable de  $S$ , soit

$S_n = n a^2$ , est lié, comme il est naturel, au nombre n des sondages positifs.

- l'ordre de grandeur de l'erreur possible, par contre, est lié au périmètre 2p de la surface moyenne  $S_m$ . L'aire de la bande critique  $B = S_0 - S_p$ , comprise entre les deux surfaces limites optimiste et pessimiste, est en effet égale à :

$$(2) \quad B = 2 p a .$$

C'est à partir de ces deux indications qualitatives qu'il convient de rechercher l'expression d'une fourchette utilisable en pratique. Pour y parvenir, il convient d'assimiler la valeur inconnue de l'aire réelle  $S$  à une variable aléatoire. Théoriquement incorrecte, une telle assimilation permet seule de donner une solution pratique au problème de l'effet de bordure. La valeur probable d'une telle variable aléatoire doit être prise égale à  $S_m = n a^2$ , tandis que sa variance  $\frac{2}{S}$  doit être liée au périmètre  $2p$ .

Nous avons suggéré <sup>1</sup> l'expression suivante pour la variance relative:

$$(3) \quad \frac{\frac{2}{S}}{S^2} = \frac{1}{n^2} \frac{p^2}{S_m}$$

où figure le coefficient purement géométrique  $\frac{p^2}{S}$ , toujours supérieur à  $\pi$ , et où le nombre  $n$  de sondages positifs intervient par l'inverse de son carré. Une telle formule implique que l'erreur possible soit proportionnelle au périmètre et inversement proportionnelle à  $n$ . Plus précisément, si l'on suppose que la variable  $S$  obéit approximativement à une loi gaussienne, on pourra dire qu'il y a 95 chances sur 100 pour que la valeur réelle de  $S$  soit comprise dans la fourchette  $n a^2 \pm 2 \frac{2}{S}$ . On déduit facilement de (3) la demi largeur  $2 \frac{2}{S}$  de la fourchette, soit :

$$(4) \quad 2 \frac{2}{S} = \frac{2 p a}{\sqrt{n}} = \frac{B}{\sqrt{n}}$$

Cette demi largeur est égale à l'aire  $B = 2 p a$  de la bande critique  $S_m - S_0$  divisée par la racine du nombre  $n$  de sondages positifs.

Nous ne pouvons pas espérer que des formules aussi simples que (3) et (4) représentent mieux qu'une approximation grossière de la réalité. Seule l'expérience peut nous indiquer si elles sont, ou non, utilisables en pratique. Il n'est pas possible de déterminer expérimentalement la variance de l'hypothétique variable aléatoire  $S$ . Par contre, un contour  $S$

étant donné à priori, il est facile de compter le nombre  $n$  de sondages positifs, que donnerait la reconnaissance de  $S$  à la maille  $a$ . Ce nombre  $n$  apparaît comme une fonction de l'orientation du réseau relativement à  $S$ , et de l'implantation de l'un des sondages à l'intérieur d'un carré de côté  $a$ . Il constitue une variable régionalisée doublement périodique - la double période étant définie par un carré fondamental  $a^2$ . La variance de  $n$ , et de l'estimation  $S_m = n a^2$  de la surface  $S$ , lorsque l'origine du réseau se déplace dans le carré fondamental, est accessible à l'expérience et peut être comparée avec la formule théorique (3).

Plutôt que la variance  $\sigma_S^2$  de la formule (3), on déterminera expérimentalement la variance  $\sigma_n^2$  du nombre  $n$  de sondages positifs, dont la valeur théorique est :

$$(5) \quad \sigma_n^2 = \frac{p^2}{S}$$

valeur indépendante de  $n$ , c'est-à-dire de la maille.

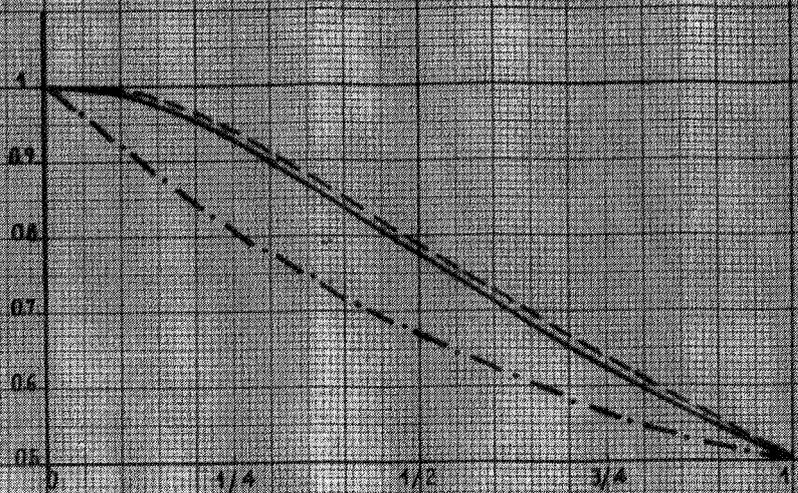
Cinq contours ont été ainsi étudiés expérimentalement : un contour A, relativement simple, représenté Figure 2 ; un contour A', identique au précédent après rotation de  $45^\circ$  par rapport au réseau de sondages (Fig 3) ; un contour B qui est un cercle de rayon 20,4 ; et deux autres contours C et D, relativement compliqués, représentés en figures 4 et 5. Pour chacun de ces contours, on a expérimenté un certain nombre de mailles carrées : mailles 3, 4, 6, 8, 10, 12 etc.., l'unité utilisée étant indiquée sur les figures 2, 3, 4 et 5. Pour chaque maille, on a compté le nombre  $n$  de sondages positifs obtenus en déplaçant le sondage origine, à l'intérieur du carré fondamental, par translations successives égales à l'unité. L'opération est très facile à réaliser matériellement, en superposant au plan du contour un calque transparent, sur lequel a été porté le réseau correspondant à la maille carrée adoptée. Pour une maille 3, on a ainsi 9 valeurs de  $n$ , pour une maille 4, 16 valeurs de  $n$  etc....

Nous reproduisons ci-contre les histogrammes des valeurs de  $n$  obtenues pour chaque maille et chaque contour.





Figure 4



$1/(1+x^2)$

$\lambda F(x)$

$1/(1+x)$

Contour D (Figure 5)											
Maille 3		Maille 4		Maille 6		Maille 8		Maille 12		Maille 24	
n	f	n	f	n	f	n	f	n	f	n	f
40	I	21	I	7	I	3	I	0	I	0	262
41		22	3	8	I	4	8	1	34	I	227
42	I	23	2	9	I	5	I4	2	35	2	87
43		24	2	10	7	6	I6	3	I9		
44	3	25		11	I2	7	II	4	35		
45		26	2	12	8	8	6	5	20		
46	I	27	2	13	5	9	6				
47	3	28	2	14	I	10	I				
		29	2			11	I				

Chacune de ces distributions peut être caractérisée par les paramètres statistiques habituels : moyenne arithmétique  $m_n$ , qui n'est autre que la surface rapportée à la maille, et variance  $\sigma_n^2$ , dont la valeur expérimentale doit être comparée à la valeur théorique  $t^2$ .

Ces différents résultats sont présentés sur le tableau suivant :

	Maille	Moyenne $m_n$	Variance expérimentale $\sigma_n^2$	$t^2/S$	$(L+P)^2$
					L1
Contour A	( 3	59.8	6.25	12.8	4.6
	) 4	33.6	3.36		
	( 6	14.9	3.94		
	) 8	8.4	3.20		
	( 10	5.4	1.37		
Contour A'	( 3	60.0	8.67	13.8	4.1
	) 4	33.8	4.45		
	( 6	15.0	1.89		
	) 8	8.4	1.22		
	( 10	5.4	0.82		

- VIII -

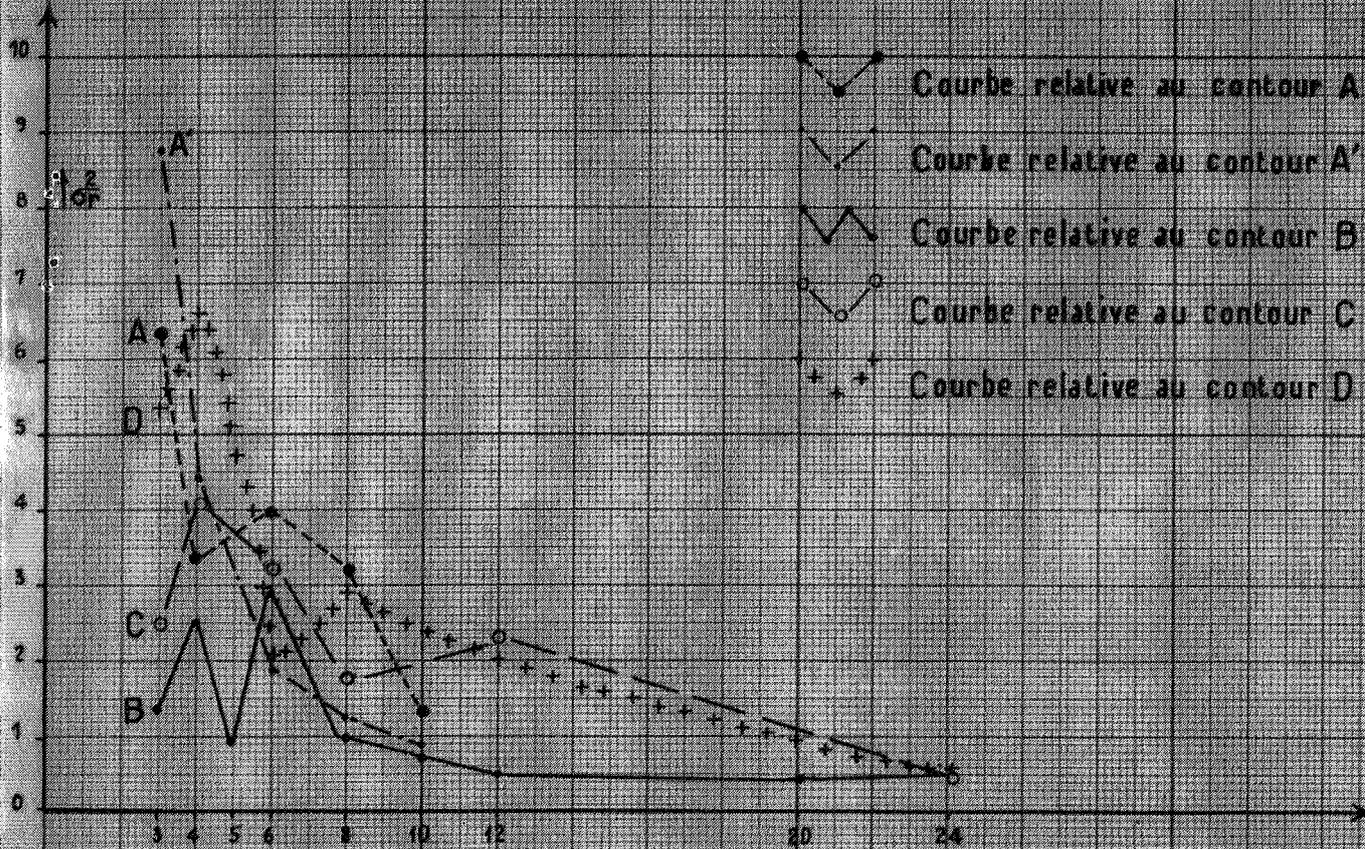
Maille	Moyenne $m_n$	Variance $\frac{2}{6n}$ expérimentale	$t^2/S$	$\frac{(L+E)^2}{LE}$	
Contour B	( 3	143.2	1.28	5.1	4.0
	) 4	80.6	2.49		
	( 5	51.6	0.81		
	) 6	35.8	2.88		
	( 8	20.1	0.99		
	) 10	12.9	0.74		
	( 12	8.9	0.50		
Contour C	) 20	3.2	0.49	33.8	4.15
	( 24	2.2	0.53		
	( 3	66.7	2.39		
	) 4	37.7	4.06		
	( 6	16.8	3.21		
	) 8	9.4	1.72		
	( 12	4.2	2.24		
Contour D	) 24	1.0	0.47	64.5	4.05
	( 3	44.6	5.36		
	) 4	25.1	7.05		
	( 6	11.1	2.00		
	) 8	6.3	2.82		
	( 12	2.78	2.00		
) 24	0.69	0.51			

La figure 6 représente, pour chacun des contours, la variation de la variance expérimentale en fonction de la maille :

- On est, tout d'abord, frappé par le caractère non régularisé de ces courbes. Il ne faut pas espérer pouvoir les représenter de façon rigoureuse par une loi simple. On note, cependant, sur toutes les courbes, une tendance générale à la décroissance pour les grandes valeurs de la maille  $a$ . Cette tendance s'explique assez bien si l'on remarque que, lorsque la maille devient assez grande pour que la surface  $S$  soit tout entière comprise dans un seul carré  $a^2$ , les seules valeurs possibles de  $n$  sont 0 et 1. La variance  $\frac{2}{6n}$  est alors  $\frac{S}{a^2} \left(1 - \frac{S}{a^2}\right)$ ,

Figure 6

Variance  $\sigma_p^2$  en fonction de la maille  $a$



inférieure à l'unité, et, à fortiori, inférieure à  $\frac{p^2}{S} > 3.1416$ . Par raison de continuité, lorsque  $n$  est faible, on doit s'attendre à observer des variances peu élevées. A partir de  $n = 10$  ou  $20$ , les variances deviennent plus élevées et varient de façon assez capricieuse. Toutes les valeurs observées, sans exception, sont inférieures à la valeur théorique du  $p^2/S$ , et parfois même - dans le cas des contours C et D - très notablement inférieure. Ces deux contours C et D, particulièrement irréguliers, possèdent des coefficients  $p^2/S$  très forts, tandis que les variances  $\sigma_n^2$  conservent des valeurs très comparables à celles que l'on observe pour A, A' ou B. L'influence du  $p^2/S$  n'est donc pas aussi déterminante que le suggère la formule (5).

Les concavités et les apophyses du contour, qui majorent démesurément le  $p^2/S$  de C ou D, n'ont pas grand effet sur les variances. On est alors tenté de remplacer le  $p^2/S$  du contour par celui du rectangle circonscrit de côtés  $L$  et  $l$  parallèles aux directions principales du réseau, soit  $\frac{(L+1)^2}{L \cdot l}$ . Les valeurs de ce coefficient sont toutes comprises entre  $4$  et  $5$ , même pour les contours C et D, et sont du même ordre de grandeur que les variances observées. Il conviendrait sans doute d'adopter, en pratique, un coefficient intermédiaire entre  $p^2/S$  et  $\frac{(L+1)^2}{L \cdot l}$ .

En conclusion, il apparaît que les formules théoriques 3, 4 ou 5 donnent dans tous les cas une estimation par excès de la vraie valeur de la variance.

Ce sont des formules sûres, puisqu'elles conduisent à surestimer l'ordre de grandeur de l'erreur possible. Elles deviennent exagérément pessimistes lorsque le nombre  $n$  de sondages utiles est de quelques unités, ou lorsque le contour de la surface présente des saillants et des rentrants, et un  $p^2/S$  élevé. Dans le premier cas, on peut remplacer le  $p^2/S$  par un coefficient de l'ordre de 1 ou 2 ; dans le deuxième on obtient un ordre de grandeurs plus raisonnable en substituant au  $p^2/S$  l'expression analogue  $\frac{(L+1)^2}{L \cdot l}$  calculée sur le rectangle circonscrit. Mais on peut alors se tromper par défaut.