

PRESENTATION DE LA GEOSTATISTIQUE

Conférence faite à l'A F T P, le 21 Mai 1967

--- RESUME ---

Ce texte ne constitue en aucune façon un exposé rigoureux et systématique de la théorie des variables régionalisées. On se contente de passer en revue les idées de base, et d'indiquer quelques unes des applications courantes utilisées dans l'industrie minière. L'objet est simplement de suggérer la possibilité d'applications analogues aux gisements pétroliers.

Les grandeurs utilisées en géologie (puissance d'une formation, teneur, perméabilité, etc ...) sont des variables régionalisées, et non aléatoires. La statistique ordinaire n'est pas capable de rendre compte des caractéristiques qualitatives de leur répartition dans l'espace. Ces caractéristiques, liées aux propriétés structurales (notamment géologiques) du phénomène naturel étudié, sont ensuite énumérées, et l'on examine la possibilité de les représenter à l'aide de schémas mathématiques, qui se classent en deux grands groupes : les représentations transitives (non probabilistes) et la théorie intrinsèque (probabiliste), cette dernière seule exposée ici.

On définit l'outil de base de la théorie intrinsèque, le variogramme, et on montre comment il représente les caractères structuraux des régionalisations. Parmi les applications de la théorie, on indique le calcul de la variance d'un échantillon dans un panneau, et de la variance d'estimation associée à un réseau de prélèvement. On mentionne ensuite brièvement le procédé du krigeage, la construction de modèle numérique et la recherche de l'optimum économique dans la reconnaissance d'un gisement.

I. GENERALITES

1.- Notion de Variable régionalisée.

En géologie, et particulièrement en géologie minière ou pétrolière, on s'intéresse à la répartition dans l'espace de certaines grandeurs, que nous appellerons des variables régionalisées. Comme exemple simple de variables régionalisées, citons parmi d'autres : la puissance d'une formation géologique, la teneur en un métal donné à l'intérieur d'un gisement minier, la perméabilité dans un champ pétrolier, etc ...

Le géologue, ou le mineur peut s'intéresser à une telle grandeur pour des raisons d'ordre purement pratique. Par exemple, il peut s'agir d'estimer les réserves d'un gisement, en tonnage et en teneur, de prévoir la production future d'un champ pétrolier, et, dans la mesure du possible, d'accompagner ces estimations d'un calcul d'erreur. Mais on peut aussi s'intéresser à ces grandeurs, plus généralement, pour essayer de comprendre, ou de rendre intelligible, le phénomène naturel qu'elles représentent. Il s'agit alors d'extraire, de la masse des données numériques brutes disponibles, les traits structuraux et les caractéristiques majeures d'un phénomène naturel.

Ces deux points de vue ne s'opposent nullement. En fait, pour résoudre un problème essentiellement pratique, comme le calcul de l'erreur possible associée à l'estimation d'un gisement minier, il est impératif de tenir compte de certains caractères structuraux de la variable régionalisée, tels que la continuité. L'exemple élémentaire suivant le montre clairement. Supposons que des mesures effectuées à intervalles réguliers le long d'une ligne sur le terrain aient donné, dans un premier cas, la séquence de valeurs numériques suivantes A :

A / 1 - 2 - 3 - 4 - 5 - 6 - 5 - 4 - 3 - 2 - 1

et, dans un deuxième cas, la séquence B :

B / 1 - 4 - 3 - 6 - 5 - 1 - 4 - 3 - 2 - 5 - 2

... / ...

Dans le cas A, on a manifestement une structure symétrique très forte, caractérisée par une décroissance très régulière de part et d'autre d'un maximum central. Dans le deuxième cas, la structure, si elle existe, est très faible, et l'impression qui se dégage est celle d'une extrême irrégularité. Bien que les onze mesures aient donné dans les deux cas les mêmes valeurs numériques (dans un ordre différent), il est clair que l'estimation d'une valeur moyenne peut être faite, dans le cas A, avec une bien meilleure précision que dans le cas B. La précision dépend, en effet, de la continuité dans l'espace de la variable régionalisée.

On voit aussi, sur cet exemple simple, que la description purement statistique du phénomène est tout-à-fait insuffisante. Le statisticien qui se contenterait de construire les histogrammes des séquences A et B concluerait, puisque ces deux histogrammes sont identiques, mais bien à tort, à l'identité des deux phénomènes naturels observés. Dans le traitement statistique ordinaire, où l'on se contente de classer sous forme d'histogrammes les échantillons disponibles, on fait par là même abstraction de l'endroit où ils ont été prélevés. On détruit les structures spatiales. Or il est clair qu'il ne suffit pas de savoir avec quelle fréquence une teneur donnée se répète dans un gisement minier. Il est au moins aussi important de savoir de quelle manière les teneurs se succèdent dans l'espace, et, notamment, quelles sont la taille et la position des zones exploitables.

Il apparaît ainsi que les variables régionalisées ne peuvent pas être assimilées aux variables aléatoires, dont l'étude est l'objet de la statistique habituelle. De fait, la notion de variable aléatoire ne peut recevoir un contenu concret que moyennant les deux conditions suivantes :

- 1/ Possibilité, au moins théorique, de répéter indéfiniment l'épreuve qui attribue une valeur numérique définie à la variable aléatoire.
- 2/ Indépendance mutuelle de ces épreuves, le résultat de l'une d'elle ne devant être en aucune manière influencé par le résultat de celles qui l'ont précédée.

Or une variable régionalisée ne peut vérifier ces deux conditions. Si l'épreuve consiste, par exemple, à prélever un échantillon de caractéristiques définies au point de coordonnées xyz , à l'intérieur d'un gisement, la teneur d'un tel

échantillon est unique, physiquement bien déterminée et nullement aléatoire. Il n'y a aucune possibilité de répétition, et la condition 1 n'est pas vérifiée. Il est cependant possible de prélever un nouvel échantillon non plus au même point, mais en un point voisin du précédent : bien qu'il ne s'agisse plus, à strictement parler, de la même épreuve, on pourrait à la rigueur admettre que la condition 1 est approximativement respectée. Mais alors la condition 2 ne le sera en aucune façon. Si le premier échantillon a été prélevé dans une zone riche, le deuxième, implanté dans son voisinage, aura tendance, en moyenne, à être riche également, et cette tendance sera d'autant plus marquée que la minéralisation sera plus continue.

2/ Caractéristiques qualitatives des variables régionalisées.

Les variables régionalisées, qui se présentent à l'observation, possèdent des caractéristiques qualitatives, étroitement liées à la structure du phénomène naturel qu'elles représentent. Parmi ces caractéristiques, que la statistique ordinaire est incapable d'exprimer, et qui doivent obligatoirement être prises en charge par la théorie des variables régionalisées, ou géostatistique, nous citerons les suivantes :

a/ Localisation

Une variable régionalisée ne prend pas ses valeurs n'importe où, mais dans une région bien déterminée de l'espace, que l'on appelle champ géométrique. Le champ est en général une formation géologique. Dans le cas d'une teneur, ce sera l'espace minéralisé du gisement lui-même. Mais on devra parfois limiter l'étude de la variable à une portion ou panneau de son champ géométrique naturel. De même, la variable sera parfois définie comme une fonction de point $f(M)$. Le plus souvent, cependant, on ne s'intéressera pas aux valeurs ponctuelles, mais aux valeurs moyennes de la variable à l'intérieur d'un petit domaine, ou support géométrique. Pour une teneur, le support sera le volume v de l'échantillon prélevé. Le support v doit être défini de manière précise, avec ses dimensions, sa forme géométrique, et son orientation dans l'espace. Si l'on change le support v , on obtient une nouvelle régionalisation, qui présente des analogies avec la première, mais ne lui est pas identique. C'est ainsi que des blocs de quelques mètres cubes ne se distribuent pas dans l'espace de la même manière que des carottes de quelques kilo. L'une des tâches de la géostatistique consistera à prévoir les paramètres de la variable dé-

finie sur un support v et dans un champ V , connaissant par exemple les caractéristiques de la variable ponctuelle dans un champ différent V' .

b/ Continuité

Un deuxième caractère essentiel est le degré de plus ou moins grande continuité dans l'espace de la régionalisation. Dans certains cas, par exemple pour des variables possédant une signification purement géométrique, comme la puissance d'une formation, on observera la continuité stricte des mathématiciens, qui se définit avec des ϵ et des η . Le plus souvent, on n'observera qu'une continuité plus lâche, dite continuité en moyenne. Dans ce cas, lorsque le point M tend vers M_0 , c'est seulement la valeur moyenne de $[f(M) - f(M_0)]^2$ qui tend vers zéro. La répartition dans l'espace prend déjà une allure plus irrégulière et discontinue. Enfin, il arrive que la continuité en moyenne ne soit même pas vérifiée. Dans ce cas d'extrême irrégularité, nous parlons d'effet de pépite. Les gisements d'or pépitique en fournissent un exemple classique.

c/ Anisotropies

En troisième lieu, une régionalisation peut être anisotrope. Il peut exister, par exemple, une direction privilégiée, le long de laquelle les valeurs se modifient lentement, tandis qu'elles varient beaucoup plus vite dans la direction perpendiculaire. Ce genre de phénomènes, connus sous le nom de runs, de zonalités, etc... est en général lié à l'existence de certaines structures géologiques.

d/ Phénomènes de transition

D'autres types de structures peuvent se manifester, liées à l'apparition dans le champ géométrique de la variable, d'un réseau de discontinuités : on parle alors, d'une manière générale, de phénomènes de transition. Comme exemples simples, on peut citer, pour les formations sédimentaires, la stratification et le relai lenticulaire des bancs. Il peut arriver que la variable, constante ou quasi constante à l'intérieur des bancs, subisse des variations brusques au passage d'un banc à l'autre. A ces discontinuités verticales, matérialisées dans l'espace par les joints stratigraphiques, viennent souvent s'ajouter des discontinuités horizontales, dues à la terminaison lenticulaire et au relai des bancs individuels. Ce réseau complexe

de discontinuités réalise une partition de l'espace minéralisé en compartiments plus ou moins indépendants les uns des autres (correspondant, par exemple, à d'anciens micro-bassins de sédimentation, où les conditions de dépôt ont évolué de façon plus ou moins autonome).

On peut noter que l'effet de pépite d'origine granulométrique est lié, lui aussi, à un phénomène de transition. Le réseau de discontinuité est ici celui qui sépare, dans l'espace, les différents grains de stérile et de minéral. La nature du phénomène est la même, mais l'échelle est bien différente. L'effet de pépite apparaît comme une forme de phénomène de transition liée à la présence de micro-structures discontinues dans le champ de la régionalisation.

3/ Les schémas théoriques.

Pour élaborer une représentation mathématique des variables régionalisées, capable de rendre compte des caractères structuraux énumérés ci-dessus, on peut penser utiliser la théorie probabiliste des processus stochastiques. Il faut bien voir, cependant, que les naturalistes soulèvent ici une objection très forte, liée à l'unicité des phénomènes naturels et à l'impossibilité de l'inférence statistique.

Rappelons, en effet, rapidement la notion de processus stochastique. Partant de la notion d'une variable aléatoire simple X , définie par sa loi de distribution probabiliste $F(x)$, les mathématiciens ont introduit ensuite des variables aléatoires à plusieurs composantes $(x_1, x_2 \dots x_n)$, décrites par des lois de distribution simultanée $F(x_1, x_2 \dots x_n)$, puis ils ont examiné ce qui se produit lorsque n devient infini. Un processus stochastique est, si l'on veut, une variable aléatoire à une infinité de composantes. Dans le cas qui nous occupe, ces composantes, en nombre infini, seraient les valeurs prises par la variable régionalisée en chacun des points de son champ géométrique. On voit quelle est l'objection des naturalistes. Les valeurs effectivement prises par la variable en tous les points de son champ doivent être considérées comme ayant été obtenues simultanément par un tirage au sort en une épreuve unique effectuée selon la loi de probabilité à une infinité de composantes qui définit le processus. Mais le passage inverse n'est pas possible. L'épreuve ayant été unique, on ne peut pas remonter des valeurs expérimentales à la distribution théorique, pas plus qu'on ne peut inférer la loi de distribution $F(x)$ d'une variable aléatoire ordinaire X lorsque l'on dispose seulement du résultat d'une épreuve unique, ayant donné une valeur numérique particulière (par exemple, $x = 98$).

Dès lors que l'inférence statistique n'est pas fondée, il est clair que seules les valeurs numériques effectivement prises par la variable, c'est-à-dire le phénomène naturel lui-même, possèdent une réalité physique, et tout recours à une interprétation probabiliste est parfaitement arbitraire. Tel est effectivement le point de vue adopté, en géostatistique, dans un premier groupe de méthodes appelé "représentation transitive". Ces méthodes, qui consistent en gros à faire l'analyse harmonique des données expérimentales, ne font appel à aucune hypothèse probabiliste. Elles mettent en jeu un formalisme mathématique assez complexe, et nous n'en parlerons pas davantage ici.

Mais il y a un cas particulier où l'objection des naturalistes peut être levée. C'est le cas dit "stationnaire". En termes abstraits, on dit qu'un processus est stationnaire si la loi $F(x_1, x_2, \dots, x_n; M_1, M_2, \dots, M_n)$ des valeurs prises simultanément par la variable en un nombre n quelconque de points arbitraires $M_1, M_2, \dots, M_n$ est invariante par une translation d'ensemble de ces points dans l'espace. Autrement dit, on postule l'homogénéité du phénomène dans l'espace. Si l'on a de solides raisons de penser que les conditions physiques chimiques du dépôt, ou de la genèse d'une minéralisation sont, effectivement, restées homogènes dans l'espace du gisement, une telle hypothèse peut être adoptée, au moins en première approximation. En vertu de son caractère stationnaire, on peut dire que le phénomène s'est en quelque sorte répété lui-même dans l'espace, et on conçoit que l'inférence statistique redevienne possible. C'est en formulant de manière un peu plus stricte ce genre d'hypothèse que l'on construit le deuxième grand groupe de méthodes de la géostatistique, qui porte le nom de "Théorie des schémas intrinsèques".

La théorie intrinsèque, en tant qu'elle postule l'homogénéité du phénomène dans l'espace et se formule en termes probabilistes, s'oppose fortement aux représentations transitives. Elle est à la fois moins générale et plus puissante. Malgré cette opposition, on constate une remarquable convergence entre les résultats des deux méthodes, et une parenté profonde de leurs formalismes mathématiques. La théorie intrinsèque, en réalité, n'utilise pas le contenu probabiliste des hypothèses faites, mais seulement leur caractère stationnaire (intrinsèque), et c'est pourquoi, malgré la différence de formulation, elle reste de même nature que les représentations transitives. C'est cette théorie intrinsèque qui va être sommairement exposée dans le reste de cet article.

II.- LA THEORIE INTRINSEQUE

Nous définirons, en premier lieu, l'outil mathématique de base de cette théorie, qui est le variogramme, en examinant son aptitude à représenter les caractères énumérés au paragraphe 2. Nous exposerons ensuite avec quelques détails deux applications importantes : calcul d'une variance en fonction du support et du champ, et variance d'estimation, et, plus brièvement, quelques autres (krigeage, construction de modèles numériques ...). Les procédés technologiques de mise en oeuvre pratique, faute de place, ne seront pas développés : le lecteur pourra se référer à mon *Traité de Géostatistique Appliquée* ⁽¹⁾.

4/ Le variogramme

L'outil mathématique de base est une fonction $\gamma(h)$, appelée demi-variogramme ou loi de dispersion intrinsèque. Si l'on désigne par $f(M)$ une variable régionalisée à support ponctuel, définie dans un champ V , le demi-variogramme se définit de la manière suivante :

$$(1) \quad \gamma(h) = \frac{1}{2V} \int_V [f(M+h) - f(M)]^2 dV$$

Ainsi le variogramme est égal à la valeur moyenne dans le champ V du carré de la différence des valeurs prises par la variable en deux points distants de h . On prendra garde que l'argument h est un vecteur, de composantes h_1, h_2, h_3 , et que la notation condensée $\gamma(h)$ représente, en réalité, une fonction $\gamma(h_1, h_2, h_3)$ de ces trois composantes. L'hypothèse ⁽²⁾ de base de la théorie est que la fonction

-
- (1) - Editions TECHNIP, Paris. Tome I. (1962), Tome II (Le Krigeage) sous presse, Tome III (l'effet de pépite et les phénomènes de transition) en préparation.
- (2) - Dans la formulation probabiliste, en réalité, l'intégrale (1) est une intégrale stochastique, susceptible de fluctuations aléatoires autour d'une certaine valeur probable. C'est cette valeur probable qui définit la "vraie" loi de dispersion intrinsèque. Ces fluctuations permettent donc de tolérer certaines variations du variogramme, lorsque l'on modifie le champ V , sans abandonner l'hypothèse intrinsèque. Il faut bien voir, cependant, que cette vraie loi a un caractère hypothétique, et que seules les intégrales (1) possèdent une réalité physique.

$\gamma(h)$ ne dépend pas du champ, ou de la portion de champ V retenue pour son calcul, autrement dit que cette fonction possède le caractère intrinsèque.

En général, le variogramme est une fonction croissante de h , comme il est naturel, puisque les valeurs prises en deux points distincts sont, en moyenne, d'autant plus différentes que ces points sont plus éloignés l'un de l'autre. Il donne ainsi un contenu précis à la notion traditionnelle de zone d'influence d'un échantillon : la croissance plus ou moins rapide du variogramme représente, en effet, la manière plus ou moins rapide dont se détériore l'influence d'un échantillon donné sur des zones de plus en plus lointaines du gisement. Examinons, plus précisément, dans quelle mesure la fonction $\gamma(h)$ est susceptible de représenter les caractères structuraux énumérés au paragraphe 2. Les caractères de localisation seront examinés au paragraphe suivant. Restent la continuité, les anisotropies, et les phénomènes de transition.

a/ Continuité.

La continuité d'une régionalisation est liée étroitement au comportement du variogramme au voisinage de l'origine. On peut, en gros, distinguer quatre types (Figure 1). Une variable possédant la continuité stricte, par exemple une variable à caractère purement géométrique, comme la puissance d'une formation, est associée à un variogramme présentant un comportement parabolique à l'origine : pour une fonction $f(x)$, continue et dérivable, en effet, $f(x+h) - f(x)$ est équivalent à $h f'(x)$, de sorte que le variogramme défini en (1) se comporte comme h^2 lorsque h est petit.

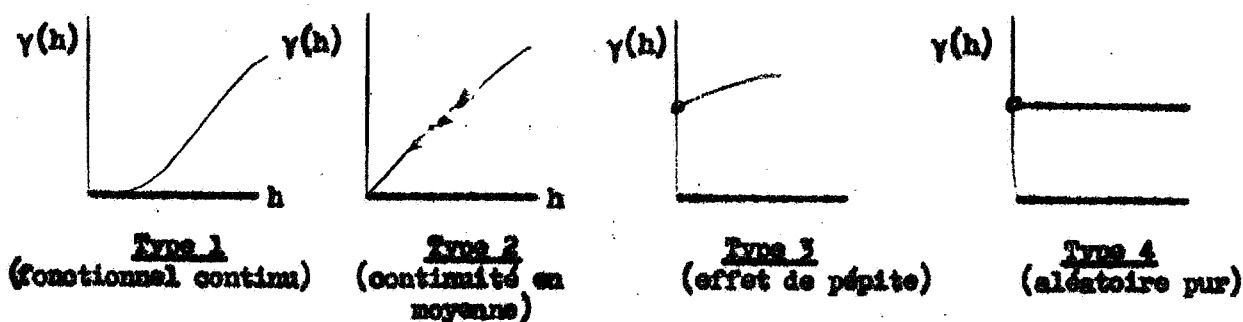


Figure 1

A une variable ne possédant que la continuité en moyenne, est associé un variogramme continu à l'origine, mais y présentant une tangente oblique : ce type 2 est parfois appelé type linéaire. Pour une variable ne possédant même pas la continuité en moyenne, c'est-à-dire présentant un effet de pépite, on observe un troisième type de variogramme, comportant une discontinuité à l'origine. Le type 4 (discontinuité pure à l'origine, et variogramme plat) est un cas limite correspondant à la notion usuelle de variable aléatoire. Entre les deux types extrêmes 1 et 4, correspondant l'un au fonctionnel continu, l'autre à l'aléatoire pur, on voit apparaître une gamme d'états intermédiaires, fréquemment observés dans la nature, dont l'étude est l'objet propre de la géostatistique.

b/ Anisotropies

Comme l'argument h est un vecteur, le variogramme doit être construit selon plusieurs directions de l'espace. C'est en étudiant comment se déforme la fonction intrinsèque $\gamma(h)$ lorsque l'on change la direction du vecteur h que l'on peut mettre en évidence et caractériser des anisotropies éventuelles. On peut observer différents types d'anisotropie, liés à l'existence de directions privilégiées ou de zonalités, qui peuvent en général s'interpréter selon des critères géologiques. La géostatistique apporte ici au géologue un outil nouveau pour déceler, ou étudier de manière précise, certains caractères structuraux (directions de courant dans une formation sédimentaire, par exemple).

c/ Phénomènes de transition.

Les phénomènes de transition, comme la stratification ou le relai lentillaire, se manifestent par un variogramme du type représenté sur la figure 2,

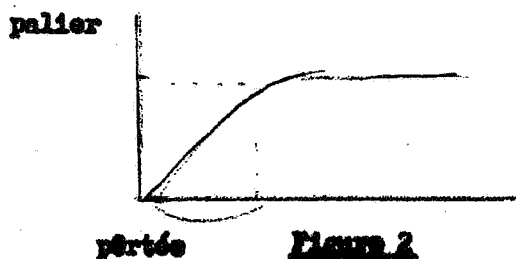


Figure 2

caractérisé par une portée et un palier.

Dans le cas d'un phénomène de relai, par exemple, la portée représente la dimension moyenne des micro-bassins autonomes de sédimentation. En étudiant comment se déforme cette dimension lorsque l'on fait varier la

direction du vecteur h , on peut se faire une idée sur la forme géométrique des micro-bassins, et mettre en évidence leur direction d'allongement. Dans le bassin ferrifère ⁽¹⁾ de Lorraine, on a pu ainsi établir que ces directions d'allongement étaient parallèles aux anciennes directions de courant.

Très souvent, un effet de pépité (type 3 de la fig. 1) peut se résoudre, si l'on change l'échelle d'observation, en un phénomène de transition du type de la figure 2. Il n'y a pas, en réalité, discontinuité à l'origine, mais une transition très rapide de portée très faible. En Lorraine ⁽¹⁾, par exemple, les variogrammes expérimentaux sont du type de transition, avec des portées de quelques centaines de mètres, mais présentent de plus un effet de pépité à l'origine. En utilisant des prélèvements distants de 50 cms, on a pu montrer que cet effet de pépité correspondait, en réalité, à une transition très rapide, avec une portée de l'ordre de 3 ou 4 mètres, de sorte que l'on voit se superposer des structures et des phénomènes de transition d'échelles et d'ampleur bien différentes.

5/ Les variances

Pour étudier l'influence des caractères de localisation, nous nous contenterons de montrer comment la connaissance du $\gamma(h)$ d'une variable ponctuelle permet de prévoir la variance d'échantillonnage non ponctuels v dans une portion quelconque V du champ géométrique. Si $f(M)$ désigne la variable ponctuelle, la variable non ponctuelle de support v se définit comme la valeur moyenne de $f(M)$ dans un échantillon de taille v prélevé au point M , soit :

$$y(M) = \frac{1}{v} \int_V f(M+u) dv$$

le vecteur u décrivant le volume v centré en M . Si m désigne la moyenne de f , ou de y , dans le champ V , la variance de $y(M)$ se définit naturellement comme la valeur moyenne dans V du carré de la différence $y(M)-m$, soit :

$$(2) \quad \sigma^2(v, V) = \frac{1}{V} \int_V [y(M)-m]^2 dV$$

- (1) - L'étude géostatistique du Bassin de Lorraine a été entreprise par un groupe d'élèves de l'Ecole des Mines de Nancy. Les résultats expérimentaux mentionnés ici figurent dans différents rapports de stage.

On montre ⁽¹⁾ alors que cette variance ne dépend que du variogramme $\gamma(h)$, et peut se calculer par la formule :

$$(3) \sigma^2(v, V) = \frac{1}{V^2} \int_V dv \int_{V'} \gamma(MM') dv' - \frac{1}{V^2} \int_V dv \int_V \gamma(MM') dv'$$

$$= F(V) - F(v)$$

Cette formule synthétique met en évidence la différence $F(V) - F(v)$ de la valeur moyenne du variogramme $\gamma(MM')$ dans V et v , $F(V)$ par exemple se calculant en faisant décrire le volume V à chacune des extrémités M et M' du vecteur MM' (intégrale sextuple).

On voit apparaître ici une différence fondamentale avec la statistique habituelle. L'échantillon v peut être considéré comme la réunion d'un très grand nombre N de volumes élémentaires jointifs dv . Si ces volumes élémentaires étaient indépendants les uns des autres, la variance de v serait en raison inverse de leur nombre N : on devrait s'attendre à observer une variance en $\frac{1}{V}$, ou $\frac{1}{V} - \frac{1}{V}$. En fait, dans un gisement minier, les volées de 10 tonnes n'ont jamais une variance mille fois plus faible que des carottes de sondages de 10 kilo. Cela prouve que les volumes élémentaires dv ne sont nullement indépendants, mais s'insèrent dans un réseau de corrélations spatiales, dont la nature est liée à la plus ou moins grande continuité de la minéralisation dans l'espace, et que le variogramme a justement pour rôle d'exprimer. Les dv ont des tenseurs beaucoup moins différentes, en moyenne, que s'ils étaient indépendants, et c'est pourquoi les volées de 10 tonnes ont une variance bien supérieure au millièrne de celle des carottes. Tel est la signification de la formule (3).

(1)- Dans la formulation probabiliste, (2) est une intégrale stochastique, dont la valeur probable donne la "vraie" variance de v dans V . Ce point de vue permet de tolérer certaines fluctuations des variances expérimentales vis-à-vis de la formule théorique (3), et d'expliquer pourquoi des échantillons de même taille peuvent avoir des variances un peu différentes dans différents panneaux V égaux entre eux.

De cette formule, on tire immédiatement une intéressante relation d'additivité :

$$(4) \quad \sigma^2(v|V) = \sigma^2(v|V') + \sigma^2(V'|V)$$

qui exprime que la variance de l'échantillon v dans le gisement V est égale à la somme des variances de v dans un panneau V' et du panneau V' dans le gisement V .

La covariance dans V de deux échantillons distincts v et v' se définit de la même manière, et s'obtient par une formule synthétique analogue à (3), que nous n'écrirons pas. Mais on voit que le coefficient de corrélation dans V de v et v' n'est absolument pas indépendant de V . En particulier, si $\gamma(h)$ est une fonction indéfiniment croissante (cas relativement fréquent), ce coefficient de corrélation tend vers $+1$, quels que soient v et v' , lorsque l'on augmente indéfiniment les dimensions du panneau V . Cette absence de caractère intrinsèque enlève beaucoup de son intérêt au coefficient de corrélation, dont la géostatistique ne fait pratiquement pas usage. C'est, en définitive, le variogramme lui-même qui représente au mieux le réseau des corrélations spatiales.

6/ Les variances d'estimations.

Pour estimer la teneur moyenne z d'un gisement, ou d'un panneau V , on dispose, en général, des teneurs y_i de n prélèvements implantés en des points précis du panneau V , et on prend, par exemple, leur moyenne arithmétique.

$$(5) \quad y = \frac{1}{n} \sum y_i$$

La théorie intrinsèque associe à l'erreur d'estimation $(z - y)$ une variance $D^2(z - y)$, dite variance d'estimation. Pour la définir, on imagine que le panneau réel V est extrait d'un grand champ géométrique K dans lequel règne la même loi de dispersion intrinsèque $\gamma(h)$ que dans V . Lorsque, par la pensée, on déplace l'ensemble du panneau V et des échantillons y_i qu'on y a implantés dans le champ K , l'erreur $z - y$ apparaît comme une variable régionalisée.

sée, dont la variance dans K peut être calculée : on s'aperçoit que K n'intervient pas réellement, et que cette variance, dite variance d'estimation, ne dépend que de $\gamma(h)$ et de la géométrie relative de V et du réseau de prélèvement. A l'aide des variances et covariances de z et de y dans K, on a, en effet :

$$(6) \quad D^2(y - z) = \sigma_y^2 + \sigma_z^2 - 2 \sigma_{yz}$$

et l'influence de K se traduit seulement par une même constante additive dans les expressions de σ_y^2 , σ_z^2 et σ_{yz} , constante qui s'élimine d'elle-même de la relation (6). Dans le cas d'un réseau de prélèvements ponctuels, effectués aux points $M_1, M_2 \dots M_n$ du gisement V, la formule (3) permet d'explicitier la variance d'estimation sous la forme synthétique suivante :

$$(7) \quad D^2(y - z) = \frac{2}{nV} \sum_1 \int_V \gamma(M_1 M) dM - \frac{1}{n^2} \sum_{1j} \gamma(M_1 M_j) \\ - \frac{1}{V^2} \int_V dM \int_V \gamma(M M') dM'$$

où dM et dM' représentent des éléments de volume. Le terme mixte, où figurent à la fois une sommation et une intégration divisé par deux, apparaît comme une valeur approchée de l'intégrale double aussi bien que la somme double qui constituent les deux autres termes. La variance d'estimation se présente ainsi comme l'erreur associée au calcul numérique approché de certaines intégrales. Sans entrer dans les détails, il est clair qu'elle sera d'autant plus faible que le réseau des points de prélèvements M_1 sera plus représentatif de la géométrie réelle du gisement V, et que la fonction $\gamma(h)$ elle-même sera plus régulière. Comme les irrégularités majeures d'un variogramme se produisent au voisinage de l'origine, et sont liées à la plus ou moins grande continuité de la régionalisation, on voit que cette variance sera d'autant plus faible, comme il est naturel, que la minéralisation sera elle-même plus continue.

En pratique, cependant, on n'utilisera pas la formule (7), mais des formules d'approximation beaucoup plus simples, qui reposent sur la notion de variance d'extension élémentaire. A chaque échantillon distinct, prélevé dans le gisement, est attribuée une zone d'influence qu'il est censé représenter. A l'erreur commise en attribuant à la zone d'influence la teneur de l'échantillon,

est associée une variance, dite variance d'extension élémentaire, donnée comme un cas particulier simple de la formule générale (7), donc aisée à calculer et à tabuler pour chaque type de fonction $\gamma(h)$. Dans la mesure où les erreurs commises en estimant chaque zone individuellement par son échantillon peuvent être considérées comme indépendantes (de covariance géostatistique nulle), on conçoit que la variance d'estimation (7) puisse se calculer approximativement en divisant par le nombre n des échantillons cette variance d'extension élémentaire σ_E^2 , soit :

$$D^2(y - z) = \frac{1}{n} \sigma_E^2$$

Sous cette forme simpliste, cependant, ce principe d'approximation est rarement utilisable. Il convient, en réalité, de procéder par composition d'un terme de tranche et d'un terme de section. Montrons sur un exemple de quoi il s'agit. Soit un gisement filonien tracé à n niveaux, chaque niveau étant échantillonné par rainurages équidistants, le nombre total des rainurages étant N . On raisonne de la manière suivante : si les teneurs réelles des niveaux tracés étaient connues, on obtiendrait la variance d'estimation en divisant par le nombre n des niveaux la variance d'extension $\sigma_{E_1}^2$ de chacun d'eux dans sa tranche d'influence, tranche constituée par les deux demi-étages amont et aval. Mais la teneur réelle des travaux miniers n'est pas connue. Elle est seulement estimée à partir des rainurages. L'erreur commise en confondant la teneur des tracés avec celle des rainurages est affectée d'une variance d'estimation, que l'on calcule en divisant par N la variance d'extension élémentaire $\sigma_{E_2}^2$ d'un rainurage dans le tronçon de galerie qu'il représente (terme dit de section). Enfin on montre que ces deux erreurs (extension de la teneur des rainurages aux travaux miniers, extension de la teneur des travaux miniers supposée réellement connue du gisement entier) peuvent être regardées comme à peu près indépendantes. On obtient ainsi la formule :

$$(8) \quad D^2(y - z) = \frac{1}{n} \sigma_{E_1}^2 + \frac{1}{N} \sigma_{E_2}^2$$

De ces deux termes, le premier (terme de tranche) est en général assez largement prédominant. Il serait vain d'espérer augmenter notablement la précision en multipliant les prélèvements de rainurages sans effectuer des travaux miniers supplémentaires. La variance d'estimation ne descend pas en dessous de la valeur du terme de tranche, valeur notable, et pratiquement très vite atteinte lorsque N augmente un peu.

7/ Quelques autres applications.

Parmi les applications de la théorie intrinsèque, utilisées en recherche minière, citons le procédé géostatistique du krigeage. Ce procédé consiste à trouver la meilleure estimation possible d'un panneau donné V à partir des teneurs de n échantillons $x_1, x_2 \dots x_n$, prélevés en des points déterminés à l'intérieur, et aussi à l'extérieur du panneau V . On cherche à déterminer les coefficients a_i vérifiant la "condition de moyenne"

$$\sum a_i = 1$$

tels que la combinaison linéaire

$$y = \sum_i a_i x_i$$

conduise à la variance d'estimation $D^2(y - z)$ la plus petite possible. Ces coefficients sont donnés par un système d'équations linéaires, que l'on peut toujours résoudre et tabuler. Ils ne dépendent que du variogramme et de la géométrie relative du panneau à estimer et du réseau des n prélèvements utilisés.

Ce procédé permet, entre autres, de rationaliser le calcul empirique des "moyennes mobiles". De telles moyennes mobiles sont souvent calculées à l'aide des pondérations assez arbitraires, qui ne permettent pas de leur attribuer une signification vraiment claire. Le procédé du krigage leur donne un sens précis. Elles représentent alors la meilleure estimation de la valeur moyenne de la variable dans un panneau donné, mobile dans le gisement. On peut prendre comme panneau de référence la zone d'influence d'un échantillon, mais on peut aussi, à volonté, s'intéresser à des unités plus grandes ou plus petites.

Comme autre exemple d'application, indiquons la possibilité de construire un modèle numérique de régionalisation possédant une loi de dispersion intrinsèque donnée à l'avance. Cette possibilité résulte d'un théorème général selon lequel un processus stationnaire a pour transformée de Fourier un processus à accroissements indépendants. Il est aisé de construire, dans l'espace des fréquences à 2 ou 3 dimensions, un processus à accroissements indépendants possédant la densité de variance convenable.

La transformation inverse donne le processus stationnaire, donc du variogramme souhaité. Ces opérations peuvent se faire à l'aide de calculatrices électroniques. On voit leur intérêt pratique. S'il s'agit, par exemple, de réaliser un modèle hydrodynamique permettant de prévoir, par essais sur modèle réduit, la production future d'un champ pétrolifère, on peut penser que les propriétés hydrodynamiques macroscopiques sont essentiellement liées à la manière, continue ou non, isotrope ou non, etc ..., dont se succèdent dans l'espace les perméabilités. Il se pourrait donc qu'un modèle, présentant même valeur moyenne et même variogramme des perméabilités que le gisement réel, sans reproduire les détails de structure de celui-ci, permette de procéder à des prévisions suffisamment exactes.

En dernier lieu, mentionnons un autre domaine d'application de la géostatistique, qui concerne la rationalisation de la reconnaissance et de la mise en exploitation des gisements. Puisque l'on sait chiffrer, à l'aide d'une variance d'estimation, la quantité d'information apportée par un schéma donné de reconnaissance par sondages ou travaux miniers, il est possible de déterminer la valeur économique de cette information, de la mettre en balance avec le prix de revient des recherches elles-mêmes, et de faire ainsi apparaître le volume optimum des travaux de reconnaissance qu'il convient de faire sur un gisement donné. Ces méthodes de rationalisation, qui seront prochainement publiées ⁽¹⁾, apparaissent comme un prolongement naturel de la géostatistique, et sont sans doute susceptibles de plus larges applications.

G. MATHERON

(1) - In Annales des Mines, sous le titre "Recherche d'optimum dans la reconnaissance et la mise en exploitation des gisements miniers" par G. MATHERON et Ph. FORMERY.