Concepts et méthodes de la géostatistique

Jacques RIVOIRARD

C–158

Octobre 1995

CENTRE DE GÉOSTATISTIQUE
35, RUE SAINT–HONORÉ, 77305 FONTAINEBLEAU (France)

ECOLE DES MINES
DE PARIS

CENTRE DE GÉOSTATISTIQUE,
Ecole Nationale Supérieure des Mines de Paris,
35 rue Saint–Honoré, 77305 FONTAINEBLEAU, France

Jacques RIVOIRARD. *Concepts et méthodes de la géostatistique.* Octobre 1995. 95 p.
C–158

© ENSMP 1995

CONCEPTS ET MÉTHODES
DE LA GÉOSTATISTIQUE

1 INTRODUCTION 2
Les variables régionalisées 2
Moyennes et variances 3

2 LES METHODES TRANSITIVES 6
Evaluer une surface 6
Evaluer une abondance 8
Variance et covariogramme transitif 8
Le covariogramme géométrique 9
Annexe: Estimation locale en représentation externe 11
Exercices 13

3 QUELQUES STATISTIQUES ET LEURS EFFETS 16
Teneurs... et teneurs 16
Relation entre un bloc et son échantillonnage 19
Support et dispersion 20
Application immédiate à l'échantillonnage d'un gisement 21
Exercices 22

4 LE VARIOGRAMME 28
Derrière les variances, le variogramme 28
Le variogramme expérimental 29
Le variogramme exhaustif 30
Le modèle de variogramme 31
Propriétés 32
Exercices 34

5 LE CADRE INTRINSEQUE 36
Structure d'un phénomène 36
Les Fonctions Aléatoires (FA) 36
Les Fonctions Aléatoires Intrinsèques 37
Variances d'estimation et variances de dispersion 38
Principes d'approximation 40
Le krigeage 41
Exercices 42

6 LE CADRE STATIONNAIRE 48
Définition d'un modèle stationnaire 48
Une FA stationnaire est aussi intrinsèque 48
Variances et krigeage 49
Exercices 51

7 MODELES NON STATIONNAIRES 54
Erreurs d'ajustement 54
La dérive du modèle 55
Les FAIk 56
Le krigeage en FAIk 58
Exercices 59

8 GEOSTATISTIQUE MULTIVARIABLE 62
Le cokrigeage 62
Les outils structuraux croisés 62
Modèles de corégionalisation 64
Régressions entre variables 66
Exercices 68

9 GEOSTATISTIQUE NON-LINEAIRE 70
Dépassement de seuil et loi conditionnelle gaussienne 70
Le cas lognormal 71
Espérance conditionnelle en modèle gaussien anamorphosé 72
Changement de support: le modèle gaussien discret 72
Autres outils non-linéaires, recours aux indicatrices 75
L'estimateur KD 76
Annexe: Les polynômes d'Hermite 77
Exercices 79

10 LES SIMULATIONS 82
Qu'est-ce-qu'une simulation? Pourquoi des simulations? 82
Jetons aléatoires, dilution de points poissonniens 82
Les méthodes spectrales 84
Les bandes tournantes 85
FA gaussienne et conditionnement d'une simulation 88
Autres modèles 89
Exercices 89

11 CONCLUSION 92

12 REFERENCES ET BIBLIOGRAPHIE 94

CONCEPTS ET MÉTHODES

DE LA GÉOSTATISTIQUE

Avertissement: Le but de cet ouvrage est de présenter rapidement et synthétiquement les concepts et méthodes courants de la géostatistique. Si la présentation et l'enchaînement des différentes parties sont parfois inhabituels, le contenu n'est pas original et ne fait pour l'essentiel qu'emprunter à la littérature géostatistique existante (notamment Matheron, 1970 et 1978, voir bibliographie). Attention: il ne s'agit pas d'un guide illustré de la géostatistique.

1 INTRODUCTION

C'est dans les années 50, avec des noms comme H.S. Sichel, D.G. Krige, H.J. DeWijs, qu'ont commencé à voir le jour, pour les besoins des évaluations minières, les méthodes appelées plus tard géostatistiques, développées et conceptualisées en grande partie par G. Matheron. Depuis, cet ensemble de méthodes n'a cessé de s'enrichir, au contact de nouveaux problèmes d'évaluation et de nouveaux domaines d'application: pétrole, hydrogéologie, sciences du sol, agriculture, forêts, pêches, topographie, météorologie, environnement, génie civil, sciences des matériaux...

Les variables régionalisées

Dans chacun de ces domaines, on s'intéresse à une "variable régionalisée" qui évolue à l'intérieur d'un certain champ de l'espace. Le plus souvent cette variable est définie dans les 3 dimensions de l'espace: porosité d'une roche, perméabilité, teneur en métal... ou en 2 dimensions: densité d'arbres, densité de poissons (rapportée à la surface de la mer), profondeur d'un horizon géologique, puissance d'une couche... En pratique il peut être très utile de faire chuter la dimension de l'espace de travail: c'est le mineur qui, ayant affaire à une veine ou à un filon, travaillera à 2D plutôt qu'à 3: il intégrera une dimension et s'intéressera à la "variable montée": l'accumulation métal, somme des teneurs le long de chaque sondage traversant la veine à angle droit. Ce peut être aussi le passage de 2 ou 3D à 1D: ainsi, connaissant la densité en poissons, mesurée par acoustique en continu le long de transects parallèles, on passera à l'abondance par transect.

Typiquement, du phénomène qu'il étudie, le géostatisticien ou le praticien ne connaît que les mesures de la variable en certains points ou le long de lignes, mais dispose également d'une information qualitative importante (géologie, conditions d'échantillonnage, expérience...). Il cherche à prévoir une quantité qui dépend, bien évidemment, de la variable en dehors des points de données. Or les variables régionalisées étudiées en géostatistique présentent très souvent un comportement irrégulier, aléatoire, rendant la prévision incertaine. C'est en étudiant ce comportement, en dégageant la structure, puis en faisant usage de celle-ci, que le géostatisticien parvient à une meilleure connaissance de son phénomène.

Bien sûr, ceci suppose que les données sont, d'une certaine manière, représentatives du phénomène: on n'est jamais à l'abri du risque d'erreur radicale, dû au fait que l'essentiel du phénomène a échappé

à la reconnaissance (risque inhérent à toute méthode, géostatistique ou autre). Penser par exemple à la détection d'anomalies décamétriques à partir de mesures espacées d'un kilomètre.

Malgré la diversité de ses applications, la géostatistique conserve une unité profonde, unité que nous allons découvrir en parcourant, chapitre après chapitre, les concepts et les méthodes les plus courants.

Moyennes et variances

Bien que l'accent ne soit pas mis sur l'aspect mathématique, il sera utile de rappeler quelques notions statistiques de base. Imaginons pour cela l'expérience suivante. Nous creusons 10 trous dans le sol, et comptons le nombre de pierres (précieuses...) dans chacun d'eux. Dans quatre des dix trous nous n'en trouvons aucune, mais trois autres trous contiennent chacun une pierre, nous en comptons deux dans chacun de deux autres trous, et le dernier trou donne trois pierres. L'histogramme Fig. 1 montre graphiquement les fréquences des valeurs 0, 1, 2 et 3 prises par la variable "nombre de pierres dans un trou".

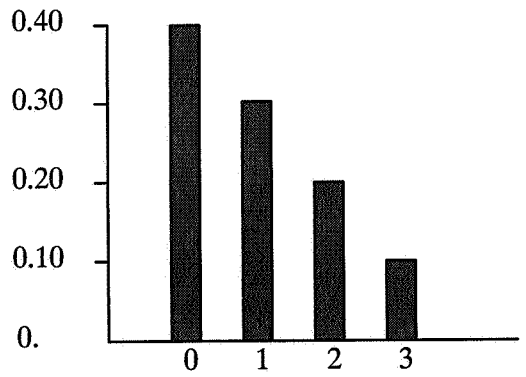

Fig. 1: Histogramme du nombre de pierres par trou.

La moyenne de cette variable vaut:

$$m=(0+0+0+0+1+1+1+2+2+3)/10=1 \text{ pierre}$$

Ainsi, en moyenne, un trou contient 1 pierre. Mais certains trous en comptent davantage, d'autres aucune. Cette variabilité se mesure par la variance des valeurs:

$$\sigma^2=[(-1)^2+(-1)^2+(-1)^2+(-1)^2+0^2+0^2+0^2+1^2+1^2+2^2]/10$$

moyenne des carrés des écarts à la moyenne. Il s'agit d'une quantité nécessairement positive (à la rigueur nulle si toutes les valeurs sont égales), qui après calcul se trouve ici égale aussi à 1. Comme la variable s'exprime en nombre de pierres, la variance, homogène au carré de la variable, s'exprime en pierres carrées, ici 1 (pierre)^2. Pour avoir une quantité homogène à la variable, on utilise l'écart−type σ, racine carrée de la variance, égal ici à $\sqrt{1}=1$ pierre. Variance ou écart−type indiquent de combien en moyenne les valeurs s'écartent de leur moyenne (sur l'histogramme, étalement des valeurs autour de leur moyenne).

Considérons maintenant le nombre de pierres par trou comme une variable aléatoire (VA en abrégé) notée X. Les fréquences précédentes deviennent des probabilités: X vaut 0 avec la probabilité 0.4 ("4 chances sur 10"), 1 avec la probabilité 0.3, 2 avec 0.2 et enfin 3 avec 0.1. En moyenne, la VA X vaut:

$$0x0.4+1x0.3+2x0.2+3x0.1=1$$

Cette moyenne s'appelle aussi, de façon pas très heureuse peut−être, l'espérance, ou la valeur probable, de X, notée E(X). On peut calculer de la même façon l'espérance du carré de la variable:

$$E(X^2)=0^2x0.4+1^2x0.3+2^2x0.2+3^2x0.1=2$$

et la variance de X:

$$\sigma_x^2=varX=E[X-E(X)]^2=(-1)^2x0.4+0^2x0.3+1^2x0.2+2^2x0.1=1$$

égale encore à $E(X^2)-[E(X)]^2$. Cette variance est égale à $E(X^2)$ si E(X) est nulle. De plus, rajouter une constante b à X donne une nouvelle espérance E(X)+b mais ne change pas la variance var(X+b)=varX. Noter qu'une variance ne peut jamais être négative.

Entre deux variables aléatoires X et Y, on utilise souvent la covariance:

$$cov(X,Y)=E[(X-E(X))(Y-E(Y))]$$

égale aussi à E(XY)−E(X)E(Y). Cette covariance ne peut dépasser, en valeur absolue, le produit des écarts−types $\sigma_x\sigma_y$, et on appelle corrélation entre X et Y le coefficient sans dimension, compris entre −1 et 1, égal à $\varrho=cov(X,Y)/\sigma_x\sigma_y$

Considérons maintenant une combinaison linéaire de X et Y:

$$aX+bY$$

Son espérance vaut:

$$aE(X)+bE(Y)$$

L'espérance de son carré s'écrit:

$$E(aX+bY)^2=a^2\,E(X^2)+b^2\,E(Y^2)+2ab\,E(XY)$$

et la variance:

$$var(aX+bY)=a^2\,varX+b^2\,varY+2ab\,cov(X,Y)$$

Ainsi les "moments d'ordre 1", E(X) et E(Y), et "d'ordre 2", $E(X^2)$, $E(Y^2)$, E(XY) (et donc variances et covariance) de X et Y, suffisent à calculer espérance et variance de leurs combinaisons linéaires. Ce seront de tels outils que l'on utilisera dans les chapitres 2 à 8 recouvrant la géostatistique dite linéaire. Rappelons à l'occasion la formule de la régression linéaire, meilleure estimation linéaire de Y connaissant X:

$$Y^*=E(Y)+p[X-E(X)]$$

avec la pente $p=cov(X,Y)/varX=\varrho\sigma_y/\sigma_x$

Revenons un instant à nos trous et à leur pierres, et supposons que ces trous soient creusés le long d'une ligne. Que le nombre de pierres trouvées le long de la ligne soit, dans cet ordre:

0 0 1 1 2 3 2 1 0 0,

0 0 1 0 0 1 1 2 2 3,

1 2 0 0 1 0 0 1 2 3

ou 1 2 0 2 1 0 0 3 1 0

ne correspond pas à la même structure spatiale. C'est précisément de l'identification de telles structures (analyse structurale) et de leur utilisation, qu'il va être question dans les chapitres suivants.

2 LES METHODES TRANSITIVES

Les méthodes transitives abordées dans ce chapitre constituent une petite partie de l'ensemble des méthodes géostatistiques, parmi lesquelles elles se distinguent, comme cela apparaîtra, par leur formalisme. Essentielles sur le plan conceptuel, elles sont toutefois d'un usage peu fréquent. Aussi irons nous rapidement sur certains aspects repris plus en détail dans les autres chapitres.

Evaluer une surface

A 2D considérons des sondages géologiques implantés selon un réseau régulier (a_1, a_2), de façon à reconnaître, en la débordant, une formation minéralisée A (Fig. 2).

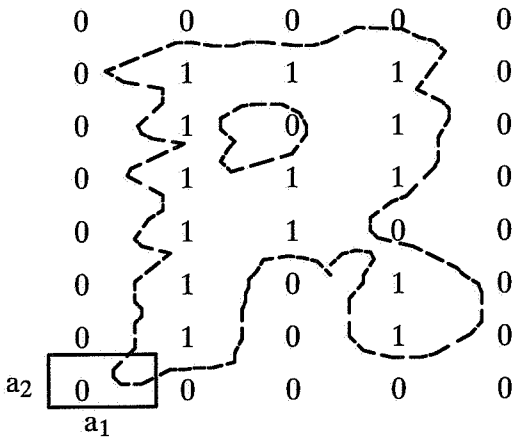

Fig. 2: Reconnaissance d'une formation par une maille régulière. Le nombre de sondages "positifs" dépend de l'implantation de l'origine de la maille.

Cette formation A occupe une surface S. Cette surface S n'est autre que la somme de tous les points appartenant à A, ce que l'on note mathématiquement:

$$S = \int\int_A dx\,dy$$

ou plus synthétiquement, x désignant un point (x,y) de l'espace 2D:

$$S = \int_A dx$$

C'est encore la somme, sur tout l'espace, de la variable indicatrice $1(x \in A)$, prenant la valeur 1 si le point x appartient à A, 0 sinon:

$$S = \int 1(x \in A)\, dx$$

Si les contours de la formation sont inconnus, la valeur exacte de S l'est également. On peut cependant en obtenir une estimation S* en comptant le nombre de sondages ayant atteint la formation:

$$S^* = N a_1 a_2,$$

estimation qui coïncide avec la surface de l'ensemble des rectangles élémentaires correspondants (Fig. 3).

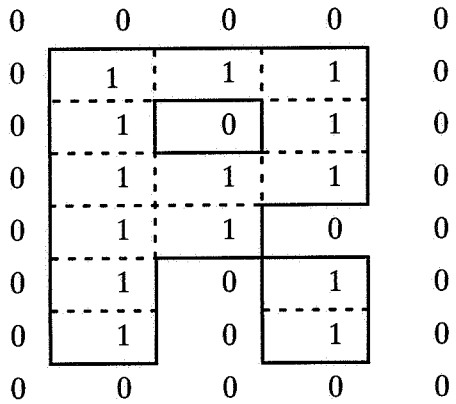

Fig. 3: La surface estimée vaut N=14 fois la surface élémentaire $a_1 a_2$. On compte $2N_1=18$ éléments linéaires en Nord–Sud, et $2N_2=10$ en Est–Ouest. D'où un écart–type relatif d'estimation σ_S/S de 10%.

La taille (a_1, a_2) de la maille élémentaire étant fixée, la valeur particulière ainsi obtenue dépend de l'implantation particulière prise par un point quelconque du réseau choisi comme origine, à l'intérieur d'un rectangle élémentaire de côtés a_1 et a_2. En considérant que cette origine est aléatoire et uniforme dans ce rectangle (autrement dit qu'elle a autant de chances d'être partout dans le rectangle), S* devient une variable aléatoire, dont la valeur probable, ou espérance, est exactement S: $E(S^*)=S$. L'erreur $S-S^*$ étant nulle en valeur probable:

$$E(S-S^*)=0$$

S* est un estimateur sans biais de S.

Le fait qu'un estimateur soit sans biais ne signifie pas qu'il soit précis. D'ailleurs on se doute, dans notre exemple, que l'estimateur sera d'autant moins précis que la maille est lâche, c'est-à-dire que a_1 et a_2 sont grands. Pour chiffrer cette imprécision, on utilise la variance de l'erreur d'estimation (en raccourci variance d'estimation):

$$\text{var}(S-S^*)=E(S-S^*)^2$$

représentant, en espérance, le carré de l'écart entre valeurs vraie et estimée. Le calcul de cette variance fait précisément l'objet des méthodes transitives.

Evaluer une abondance

Un problème analogue se pose si, au lieu d'estimer l'étendue de la minéralisation, on cherche à estimer l'abondance en métal:

$$Q=\int_A f(x)dx=\int f(x)dx$$

somme des "teneurs" f(x) en chacun des points x de la formation (ou de l'espace, f(x) étant supposée nulle en dehors de son champ A). Ici encore la somme des valeurs $f(x_i)$ mesurées aux noeuds de la maille donne un estimateur:

$$Q^*=\sum_i f(x_i)a_1a_2$$

sans biais de Q, et nous allons nous intéresser à la variance d'estimation correspondante: $\text{var}(Q-Q^*)$.

Variance et covariogramme transitif

G. Matheron donne l'expression exacte permettant de calculer cette variance:

$$\text{var}(Q-Q^*)=a_1a_2\sum_{k_1}\sum_{k_2} g(k_1a_1, k_2a_2) - \int g(h)\ dh$$

Dans cette formule, g(h) est le "covariogramme transitif" de f(x) défini par:

$$g(h)=\int f(x)\ f(x+h)dx$$

C'est une fonction du vecteur distance h, symétrique ($g(h)=g(-h)$), nulle dès que h est supérieure à la dimension du champ A dans la direction de h. Si la variable étudiée f(x) ne peut prendre de valeurs négatives, exemple de notre teneur, alors g(h) non plus, mais dans tous les cas on a:

$$|g(h)| \leq g(0)$$

et $\int g(h)\, dh = Q^2$.

Le comportement plus ou moins régulier de g(h) au voisinage de l'origine $h=0$ (continuité ou non, dérivabilité ou non) traduit la plus ou moins grande régularité de f(x) (nous reviendrons plus en détail sur ces types de régularité, mais à propos du variogramme, Chap. 4).

Dans la formule donnant la variance, k_1 et k_2 décrivent l'ensemble des entiers positifs, négatifs et nul. Les (k_1a_1, k_2a_2) représentent les vecteurs distances construits à partir de la maille (a_1, a_2). La variance d'estimation apparait alors comme l'erreur entre une valeur approchée et la valeur exacte de l'intégrale (ici à 2D) $\int g(h)dh$, erreur en général d'autant plus faible que la maille est petite, et que la fonction g, et donc la variable f, est régulière.

Dans la pratique, la fonction g(h) est inconnue. Tout au plus peut-on en calculer une estimation, pour les distances (k_1a_1, k_2a_2) multiples de la maille élémentaire, à partir des valeurs connues $f(x_i)$. Sur un tel covariogramme expérimental, et en adoptant le comportement à l'origine jugé le plus approprié, il faudra alors ajuster une courbe. C'est ce modèle (au fond une estimation du vrai covariogramme inconnu) qui fournira ensuite (une valeur estimée de) la variance d'estimation recherchée. (Mathématiquement la courbe ajustée ne peut être quelconque et doit être choisie parmi les fonctions de "type positif" de façon à assurer une variance non négative: nous retrouverons et examinerons plus en détail cette condition au chapitre 6, à propos des covariances).

Cependant la variance d'estimation dépend en grande partie de l'irrégularité du covariogramme à l'origine, et G. Matheron a développé des formules d'approximation basées sur ce degré d'irrégularité.

Le covariogramme géométrique

L'estimation de la surface S, évoquée au début, apparait comme un cas particulier de l'estimation de Q: il suffit de prendre, comme variable

f(x), l'indicatrice $1(x \in A)$ du champ A. La variance se calcule alors avec le "covariogramme géométrique" de l'ensemble A:

$$K(h) = \int 1(x \in A)\, 1(x+h \in A)\, dx$$

mesure de l'ensemble des points x tels que x appartient à A et x+h appartient aussi à A, autrement dit des points qui appartiennent à A ainsi qu'au translaté de A par −h. Comme $K(h)=K(-h)$, le covariogramme géométrique apparait comme la mesure de l'intersection du champ A et de son translaté A_h par h (Fig. 4):

$$K(h) = |A \cap A_h|$$

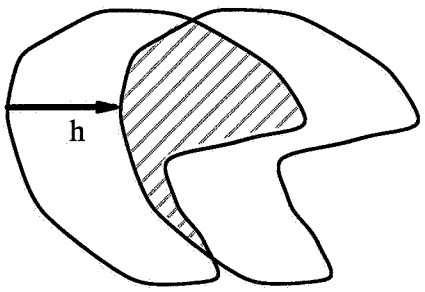

Fig. 4: Intersection d'un ensemble et de son translaté par h.

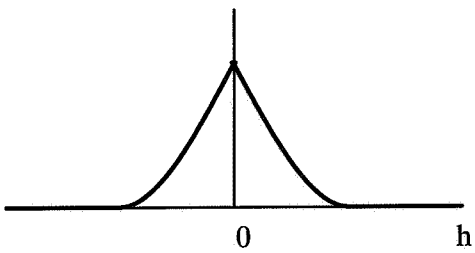

Fig. 5: Exemple de covariogramme (covariogramme géométrique d'une sphère à 3 dimensions).

On a les relations:

$$K(0) = S$$

$$\int K(h)dh = S^2$$

et $0 \leq K(h) \leq K(0)$.

Au voisinage de l'origine, un covariogramme géométrique est linéaire, Fig. 5 (avec une pente donnant la variation diamétrale de A selon la direction considérée). De cette observation, et sous des conditions courantes, G. Matheron déduit une formule d'approximation pour la variance relative:

$$\frac{\sigma_S^2}{S^2} = \frac{1}{N^2}\left[\frac{N_2}{6} + 0.061\,\frac{(N_1)^2}{N_2}\right]$$

dans laquelle $2N_1$ et $2N_2$ (avec $N_1 \geq N_2$) sont les nombres des éléments linéaires a_1 et a_2, ou a_2 et a_1, dans le contour (extérieur et intérieur) délimité par les rectangles élémentaires. Voir Fig. 3.

L'exposé précédent donne un aperçu volontairement sommaire des méthodes transitives. Mentionnons seulement que ces méthodes ne sont évidemment pas limitées au cas bidimensionnel et que le réseau d'échantillonnage n'a pas à être absolument régulier. Soulignons surtout qu'on s'intéresse à l'estimation d'une surface, d'un volume, ou d'une abondance à l'intérieur d'un certain contour, reconnus seulement par un réseau de prélèvements implanté aléatoirement par-dessus l'objet. Il s'agit là d'estimation globale, à la différence de ce que serait une estimation locale des différents points ou parties du champ (cartographie). Le covariogramme transitif ne fait d'ailleurs pas la distinction entre l'aspect géométrique du champ de la variable et le comportement de la variable à l'intérieur de ce champ: c'est ce qu'on appelle une représentation externe. Le fait qu'aucune hypothèse a priori n'est requise sur la manière dont la variable évolue à l'intérieur de son champ est l'avantage de ces méthodes transitives, en même temps qu'il en fixe la limitation. Dans les chapitres suivants au contraire, nous nous intéresserons précisément au comportement de la variable à l'intérieur de son champ (supposé alors connu).

Annexe: Estimation locale en représentation externe

Bien qu'il ne décrive pas le comportement local de la variable, le covariogramme peut être utilisé pour une estimation locale (cartographie), précisément dans des cas où les frontières du champ de la variable sont inconnues et où seules les valeurs échantillonnées nulles témoignent que l'on n'est plus à l'intérieur du champ. La construction d'un tel estimateur local est assez instructive, quant à la signification de son optimalité. La démarche s'étend à un

échantillonnage irrégulier (mais l'estimation préalable du covariogramme est plus délicate).

En chaque point x de l'espace où l'on souhaite estimer la valeur de la variable, notée ici z(x), on se propose de le faire à l'aide d'une moyenne pondérée des valeurs échantillonnées (zéros compris) à l'intérieur d'un voisinage donné de x (soit en des points $x+h_i$):

$$z(x)^* = \sum \lambda_i \, z(x+h_i)$$

On choisit ensuite les "poids" λ_i de cette combinaison linéaire de telle sorte que, s'il était possible de translater partout dans l'espace la configuration formée par les points x et $x+h_i$, la somme des erreurs quadratiques entre valeurs vraies et estimées:

$$\int [z(x)-z(x)^*]^2 \, dx$$

soit minimisée (optimalité).

Développant cette somme, on obtient:

$$\int [z(x) - \sum \lambda_i \, z(x + h_i)]^2 dx$$

$$= \int [z(x)^2 - 2 \sum_i \lambda_i \, z(x)z(x + h_i) + \sum_{i,j} \lambda_i\lambda_j \, z(x + h_i)z(x + h_j)]dx$$

$$= g(0) - 2 \sum_i \lambda_i \, g(h_i) + \sum_{i,j} \lambda_i\lambda_j \, g(h_i - h_j)$$

Les poids optimaux cherchés sont ceux qui minimisent cette somme. On montre (en annulant les dérivées par rapport aux λ_i) que ces poids sont obtenus comme solutions du système linéaire:

$$\sum_j \lambda_j g(h_i - h_j) = g(h_i) \quad \text{pour tout i}$$

On peut, aussi, imposer aux poids la condition supplémentaire suivante: que, en translatant la configuration formée par les points x et $x+h_i$, la somme des valeurs estimées $\int z(x)^* dx$ égale l'abondance réelle $\int z(x)dx$. Comme:

$$\int z(x)^* dx = \int \sum \lambda_i \, z(x+h_i) dx = \sum \lambda_i \int z(x+h_i) dx = \sum \lambda_i \int z(x) dx,$$

cette condition s'écrit $\sum \lambda_i = 1$ (somme des poids =100%).

Trouver les poids qui minimisent la somme des erreurs quadratiques sous cette condition supplémentaire nécessite l'intervention d'une inconnue supplémentaire μ (paramètre de Lagrange). On montre alors que les λ_i (et μ) sont solutions du système linéaire:

$$\begin{cases} \sum_j \lambda_j g(h_i - h_j) - \mu = g(h_i) \quad \text{pour tout i} \\ \sum_i \lambda_i = 1 \end{cases}$$

Que l'on impose ou non la condition supplémentaire, la somme des erreurs quadratiques minimisée fournit la précision (ou plutôt l'imprécision) de l'estimation. Il est intéressant de constater que c'est à un tel critère global que l'on fait appel pour optimiser une estimation locale: car c'est bien de la valeur en un point x donné qu'il s'agit.

En pratique, on détermine, pour chaque point à estimer, les poids optimaux à affecter à chacune des données présentes dans le voisinage autour de ce point (voisinage "glissant").

Exercices

▶ *Exercice* 2.1. Covariogramme géométrique 1D du segment [0,a]?

Réponse : $a-|h|$ pour $|h|<a$, 0 sinon.

▶ *Exercice* 2.2. Covariogramme géométrique 1D de la réunion des segments [0,a] et [2a,3a]?

Réponse : $2(a-|h|)$ pour $|h|<a$, $|h|-a$ pour $a \leq |h|<2a$, $3a-|h|$ pour $2a \leq |h|<3a$, 0 sinon.

▶ *Exercice* 2.3. Covariogramme transitif 1D de la fonction f(x) égale à 1 sur le segment [0,a], −1 sur [−a,0[, et 0 ailleurs?

Réponse : $2a-3|h|$ pour $|h|<a$, $|h|-2a$ pour $a \leq |h|<2a$, 0 sinon.

▶ *Exercice* 2.4. Covariogramme transitif 1D de la fonction f(x) égale à x pour $|x|<a/2$ et à 0 sinon?

Réponse : $(a^3/12)[1-3(|h|/a)+2(|h|/a)^3]$ pour $|h|<a$, 0 sinon.

▶ *Exercice* 2.5. Covariogramme géométrique d'une sphère 3D de diamètre a?

Réponse : $(\pi a^3/6)[1-(3/2)|h|/a+(1/2)(|h|/a)^3]$ pour $|h|<a$, 0 sinon, représenté Fig. 5.

▶ *Exercice* 2.6. Estimation d'un disque.

On cherche à estimer une surface par une maille orthogonale (a_1,a_2). A effort d'échantillonnage égal (a_1a_2=constant), dans quel rapport doivent se trouver a_1 et a_2 pour que la variance d'estimation soit minimale, s'il s'agit d'un disque?

Réponse : On peut penser que, le disque n'offrant pas de direction privilégiée, la maille carrée ($a_1/a_2=1$) sera en général meilleure qu'une maille rectangulaire. C'est d'ailleurs pratiquement le résultat auquel conduit la formule d'approximation, si l'on prend par exemple $a_1 \leq a_2$ et $N_1a_1=N_2a_2$=variation diamétrale.

▶ *Exercice* 2.7. Estimation d'une surface.

On cherche à estimer une surface de forme allongée par une maille orthogonale (a_1,a_2) dont une direction est celle de l'allongement. A effort d'échantillonnage égal (a_1a_2=constant), dans quel rapport doivent se trouver a_1 et a_2 pour que la variance d'estimation soit minimale, s'il s'agit d'une surface à contour elliptique?

Réponse : Ce rapport doit être celui des axes principaux de l'ellipse. En effet, par une transformation linéaire, qui conserve la variance relative σ_S^2/S^2, on se ramène à un disque et à une maille carrée, voir exercice précédent.

▶ *Exercice* 2.8. Individus dans une population.

Une population d'individus (poissons par exemple) se distribue dans l'espace selon une densité f(x), supposée nulle en dehors d'un champ fini. L'abondance de la population vaut $Q=\int f(x)dx$.

On considère la position X d'un individu pris au hasard dans cette population. Montrer que X se distribue dans l'espace avec comme densité de probabilité f(x)/Q.

En déduire que la position moyenne d'un tel individu n'est autre que le centre de gravité de la population:

$$x_{CG} = \int x \frac{f(x)}{Q} dx$$

et que sa variance, valeur moyenne de $(X-x_{CG})^2$, égale l'inertie:

$$I = \int (x-x_{CG})^2 \frac{f(x)}{Q} dx$$

On considère maintenant deux individus tirés au sort indépendamment. Soit h le vecteur distance (aléatoire) les séparant. Montrer que la densité de probabilité de h n'est autre que $g(h)/Q^2$, avec $g(h)=\int f(x)f(x+h)dx$, covariogramme de f(x).

En déduire la relation:

$$I = \frac{1}{2} \int h^2 \frac{g(h)}{Q^2} dh$$

3 QUELQUES STATISTIQUES ET LEURS EFFETS

Teneurs... et teneurs

Imaginons, pour simplifier, un gisement métallique à 2D carré (exemple factice, néanmoins fortement inspiré d'un champ de mil réel étudié par L. Sandjivy). Ce gisement pèse un certain tonnage T, contient une certaine quantité de métal Q, et sa teneur (concentration) moyenne en métal, rapport de Q à T, se trouve égale à 2.54%.

Divisons ce gisement en 4 parties de tonnage identique T/4, et de teneurs (en %):

3.08	2.22
2.85	2.00

La somme de leurs quantités de métal:

$$3.08\%T/4 + 2.22\%T/4 + 2.85\%T/4 + 2.00\%T/4$$

donnant la quantité de métal du gisement entier: 2.54%T,

on a bien: $2.54=(3.08+2.22+2.85+2.00)/4$

Autrement dit, la teneur moyenne du gisement est égale à la moyenne des teneurs de ses parties. (On dit que la teneur est une variable additive, ce qui est le cas de beaucoup de variables étudiées en géostatistique, mais non pas de toutes: la perméabilité d'un volume n'est que dans des cas très particuliers la moyenne arithmétique de la perméabilité de ses parties).

Poursuivons la division, cette fois–ci de chacune des parties, en 9 blocs identiques, de sorte que la teneur de chaque partie est la moyenne de celles de ses blocs:

2.99	3.21	4.13	3.28	1.78	2.01
3.32	2.94	3.08	1.84	1.17	1.49
3.99	2.65	1.38	3.91	1.52	3.01
2.40	0.54	1.52	2.55	1.90	3.35
4.80	1.19	2.03	2.01	1.61	2.05
5.17	4.42	3.57	2.33	1.29	0.88

La distribution des teneurs de blocs s'étale (min: 0.54%, max: 5.17%), comme le montre l'histogramme donnant la proportion de valeurs par classe Fig. 6.

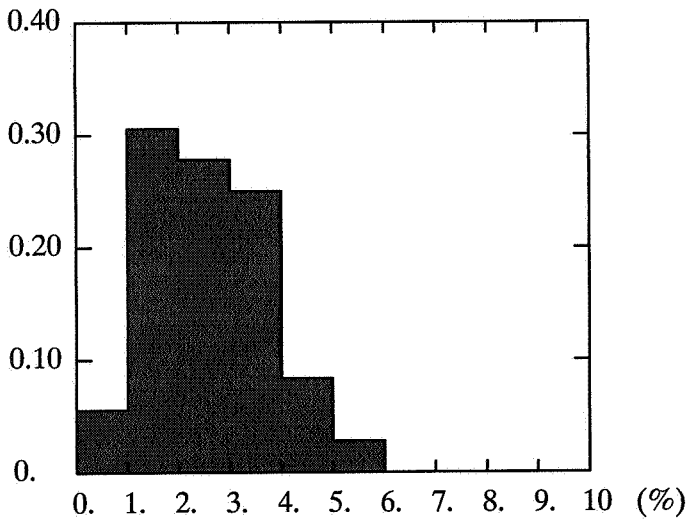

Fig. 6: Histogramme des teneurs de blocs.

Enfin, si l'on divise encore chacun des blocs en 25 petits morceaux identiques, de la taille d'éventuels échantillons (Fig. 7), l'histogramme s'étale encore davantage (min: 0.01%, max: 9.25%): Fig. 8.

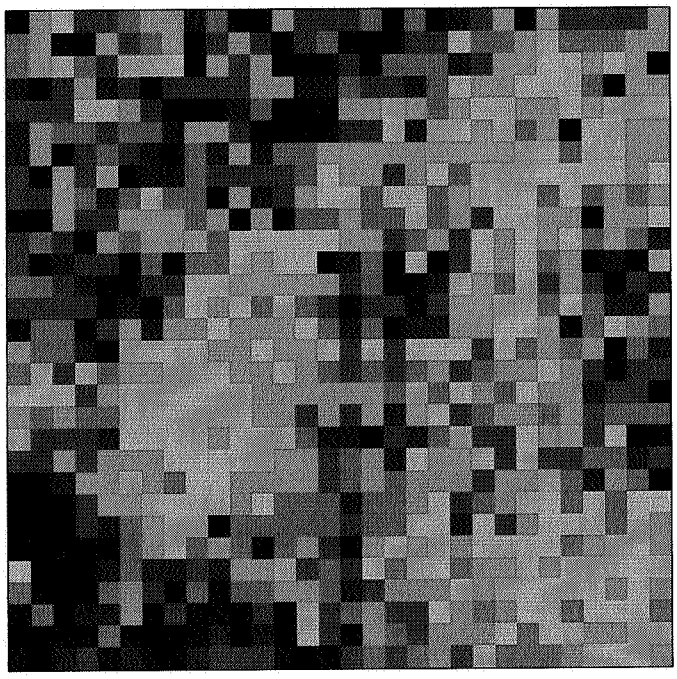

Fig. 7: Les petits morceaux.

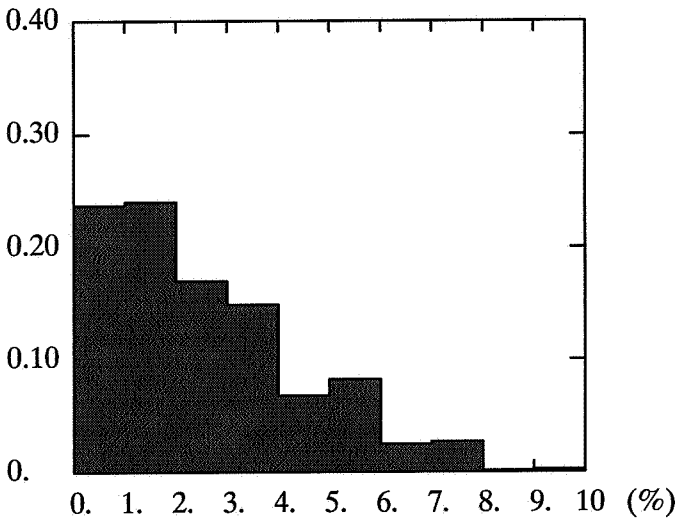

Fig. 8: Histogramme des teneurs des petits morceaux.

Relation entre un bloc et son échantillonnage

Considérons maintenant un bloc v de teneur donnée z(v): sa teneur est la moyenne de celles de ses morceaux. Si par conséquent on prélève au hasard l'un quelconque x de ses morceaux, on doit s'attendre à ce que, en espérance ou en valeur probable, la teneur de ce morceau z(x) égale celle du bloc, ce qui s'écrit:

$$E[z(x)\,|\,z(v)]=z(v)$$

Il en serait d'ailleurs de même si z(x) était remplacé par la moyenne de plusieurs morceaux prélevés non préférentiellement dans le bloc.

Cette relation montre que, si l'on échantillonne ainsi chacun des blocs, on doit s'attendre à une régression z(x) fonction de z(v) proche de la première bissectrice, voir Fig. 9.

L'autre régression par contre, celle de z(v) connaissant la teneur de l'échantillon z(x), est différente. Elle montre que la valeur probable d'un bloc est moins riche que son échantillon si celui–ci est riche, plus riche s'il est pauvre. La raison profonde de ce phénomène est que les échantillons présentent, moins souvent que les blocs, des teneurs intermédiaires, plus fréquemment au contraire des valeurs riches ou pauvres, et donnent donc une image plus contrastée que les blocs de la distribution des teneurs.

Pour l'exploitant minier qui, avant d'exploiter un bloc, ne connaît que la teneur de son échantillonnage, ceci peut représenter un piège redoutable. S'il décide en effet d'exploiter les seuls blocs échantillonnés les plus riches, il doit s'attendre à récupérer en moyenne, non pas la teneur prévue par l'échantillonnage, mais une teneur plus basse. Les blocs échantillonnés comme riches ne sont pas aussi riches que ne l'indique leur échantillonnage. Corrélativement les blocs abandonnés ne sont pas aussi pauvres que leur échantillonnage ne le laisse supposer. C'est en grande partie ce résultat essentiel, mis en évidence par le Sud–Africain D.G. Krige dans les années cinquante, qui allait servir, principalement grâce à G. Matheron, de point de départ au développement des concepts et méthodes géostatistiques. Un certain nombre de résultats remarquables peuvent d'ores et déjà être extraits de notre petit exemple.

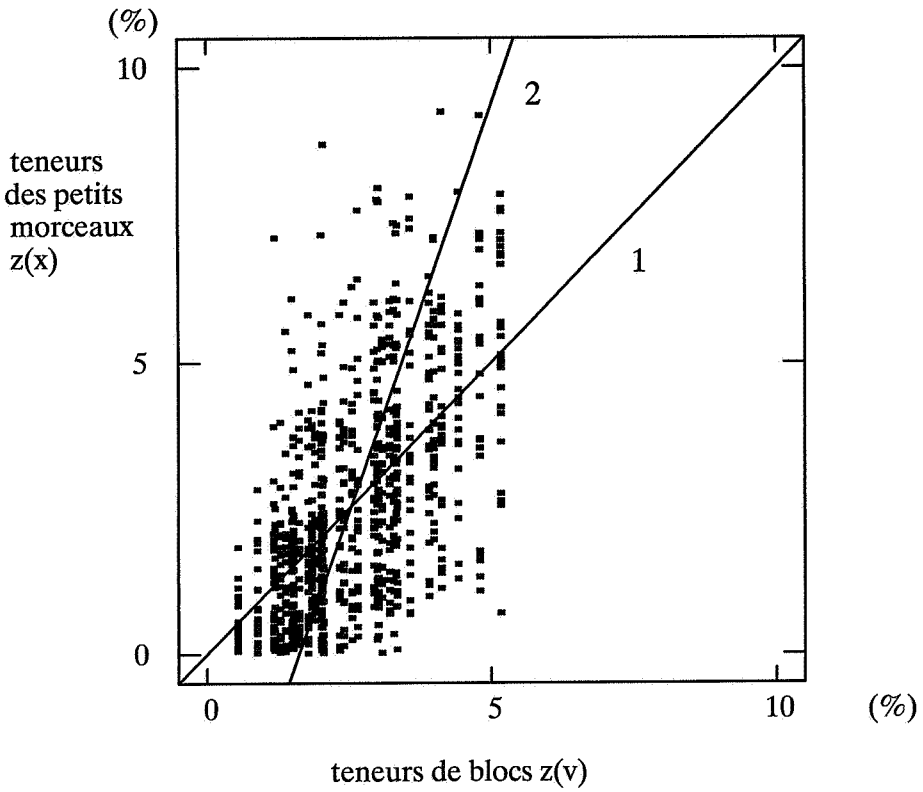

Fig. 9: Nuage de corrélation et les deux régressions. La régression (1) donne la teneur probable $E[z(x)|z(v)]=z(v)$ d'un échantillon dans un bloc de teneur z(v). La seconde régression (supposée ici linéaire) donne la valeur estimée d'un bloc dont on connait la teneur d'un échantillon.

Support et dispersion

La variable étudiée, ici une teneur, ne se présente pas de la même façon selon que l'on s'intéresse aux teneurs des blocs ou à celles des petits morceaux. Ainsi les caractéristiques, statistiques en particulier, de la teneur dépendent du support, volume physique générique, ici bloc v ou petit morceau x, sur lequel elle est définie.

Que l'on divise le gisement en blocs ou en petits morceaux, la moyenne des teneurs est la même (puisque c'est celle du gisement). Mais les petits morceaux présentent, on l'a vu, une proportion moins élevée de valeurs intermédiaires que les blocs, et une proportion plus

importante de valeurs fortes ou faibles. Cette variabilité statistique se mesure par la variance (voir Chap. 1). Si l'on considère notre gisement divisé en blocs, et ceux–ci divisés en petits morceaux, alors la variance des teneurs de tous les morceaux est supérieure à celle des blocs.

On peut même aller plus loin en considérant, en moyenne pour tous les blocs, la variance des petits morceaux à l'intérieur de chacun des blocs. Cette *variance de dispersion des morceaux dans un bloc*, comme on l'appellera, mesure la variabilité avec laquelle se distribuent, en moyenne, les morceaux à l'intérieur d'un même bloc.

On peut démontrer facilement que si, à cette variance de dispersion des morceaux x dans les blocs v, on rajoute la variance (de dispersion) des blocs v (dans le gisement G), on obtient exactement la variance de dispersion des petits morceaux x dans le gisement G, relation d'additivité que nous écrirons:

$$s^2(x|v) + s^2(v|G) = s^2(x|G)$$

De manière équivalente, la chute de variabilité, lors du passage des petits morceaux aux blocs, est égale à la variabilité des petits morceaux dans les blocs (variabilité intérieure aux blocs):

$$s^2(x|G) - s^2(v|G) = s^2(x|v)$$

Application immédiate à l'échantillonnage d'un gisement

L'échantillonnage aléatoire stratifié d'un gisement divisé en N blocs (ou strates) identiques v_i consiste à prendre un échantillon (un de nos petits morceaux) au hasard indépendamment à l'intérieur de chaque bloc. La moyenne $z(V) = \frac{1}{N}\sum z(v_i)$ inconnue du gisement est alors estimée par la moyenne $\frac{1}{N}\sum z(x_i)$ des prélèvements. L'erreur d'estimation:

$$z(V) - \frac{1}{N}\sum z(x_i) = \frac{1}{N}\sum [z(v_i) - z(x_i)]$$

est de valeur probable nulle, et l'estimation est d'autant moins bonne que la variance de cette erreur d'estimation (la "variance d'estimation"):

$$E[[z(V) - \frac{1}{N}\sum z(x_i)]^2] = E[\frac{1}{N}\sum [z(v_i) - z(x_i)]]^2$$

qui se trouve égale à $s^2(x|v)/N$, est élevée.

Par comparaison, l'échantillonnage aléatoire pur consiste à prendre N échantillons (petits morceaux) au hasard indépendamment dans le gisement. Cette fois-ci la variance d'estimation égale $s^2(x|G)/N$.

Comme $s^2(x|v) \leq s^2(x|G)$, on voit que l'échantillonnage aléatoire stratifié ne peut être que meilleur que l'échantillonnage aléatoire pur.

Exercices

▶ *Exercice* 3.1. Entre 2 et 3D.

Dans la pratique on assimile souvent à un point le support élémentaire de l'échantillon. Tout autre support (bloc...) est alors considéré comme une somme de tels points, ce qui peut, ou non, être une approximation. Considérons l'exemple suivant d'un gisement devant être exploité à ciel ouvert par bancs de 5 mètres. Ce gisement est reconnu par sondages verticaux.

Dans quelle mesure peut-on représenter par des points les "échantillons" de 5 mètres, intersections des sondages et des bancs?

Réponse : Dans cette optique toute traversée verticale d'un banc se réduit à 1 point, mettons son point central, le banc se réduit à son plan médian, et le gisement à la succession de ces plans. Il s'agit d'une représentation rigoureuse pourvu qu'on ne manipule que des blocs à côtés verticaux et dont la hauteur correspond à un ou plusieurs bancs. Ceci exclut des blocs de 5 m de haut à cheval sur deux bancs ou des blocs de 2.5 ou 7.5 mètres de haut.

En regroupant par 2 les bancs de 5 m, on obtient des bancs de 10 m. Montrer que la variance de dispersion des échantillons de 5 m dans les composites de 10 m peut s'écrire comme la moyenne des:

$$(1/4)[z_1 - z_2]^2$$

où les (z_1, z_2) sont les couples de teneurs de 5 m par composite de 10 m.

Réponse : Partir de la variance de dispersion d'un point parmi 2 points, soit la moyenne de:

$$(1/2)[(z_1 - (z_1+z_2)/2)^2 + (z_2 - (z_1+z_2)/2)^2].$$

▶ *Exercice* 3.2. Courbes de sélectivité fonction du tonnage.

Un gisement est divisé en blocs de même tonnage (support v), eux−mêmes divisés en petits morceaux.

Effet de support: On sélectionne les blocs les plus riches jusqu'à obtenir une fraction T donnée du gisement. Montrer que ces blocs contiennent une quantité de métal inférieure à celle que donnerait la sélection des morceaux les plus riches pour le même tonnage T.

Réponse : Considérer la fraction T constituée par les morceaux appartenant aux blocs sélectionnés. En échangeant les morceaux les plus pauvres de cette sélection contre les morceaux les plus riches à l'extérieur de ces blocs, le métal ne peut qu'augmenter.

Effet d'information: Lors de l'exploitation, les teneurs vraies des blocs sont inconnues et la sélection se fait sur des teneurs estimées. Conséquence pour le métal récupéré à tonnage T fixé?

Réponse : Certains blocs, en vérité pauvres, sont récupérés à la place d'autres blocs, riches, qui sont abandonnés. Le métal ne peut qu'être moindre que dans le cas idéal de teneurs parfaitement connues.

Tous ces résultats sont valables pour toute valeur de T: effet de support et effet d'information ne peuvent que dégrader la courbe métal/tonnage.

▶ *Exercice* 3.3. Courbes de sélectivité fonction de la coupure.

Même contexte que précédemment, mais on s'intéresse cette fois aux réserves au dessus d'une teneur de coupure. Soient T(z) et Q(z) le tonnage minerai et le métal apportés par les blocs passant la coupure z. On définit le bénéfice conventionnel par:

$$B(z)=Q(z)-zT(z)$$

Effet de support: Montrer qu'à coupure donnée, le bénéfice obtenu par sélection des blocs est inférieur à celui que donnerait la sélection des morceaux passant la coupure.

Réponse : Un bloc passant la coupure contient nécessairement des morceaux passant la coupure. Mais la contribution au bénéfice des morceaux ne la passant pas est négative: en les rejetant, le bénéfice augmente. Et il augmente encore en rajoutant les morceaux riches extérieurs à ces blocs.

Effet d'information: Montrer qu'à coupure donnée, le bénéfice ne peut que se dégrader si la sélection se fait sur une estimation des teneurs de blocs.

Réponse : Les blocs sélectionnés à tort apportent une contribution négative au bénéfice, sans compter le bénéfice perdu par les blocs rejetés à tort.

Finalement, pour B(z) également, il y a dégradation avec effets de support et d'information.

▶ *Exercice* 3.4. Gains et pertes en teneur dans une exploitation minière sélective.

Un gisement est constitué de la réunion d'un certain nombre de blocs de même tonnage (support v). Il sera commode de raisonner sur la teneur générique d'un tel bloc Z(v) (variable aléatoire pouvant prendre comme valeur chacune des teneurs des blocs du gisement). En espérance, c'est-à-dire en moyenne, cette teneur Z(v) égale la teneur moyenne m du gisement:

$$E[Z(v)]=m$$

et sa variance varZ(v) coïncide avec la variance de dispersion des blocs.

Le gisement a une teneur moyenne basse. Pour augmenter la teneur, on décide de ne récupérer que les blocs les plus riches. La teneur Z(v) de chacun de ces blocs est malheureusement inconnue, et la sélection se fait par application d'une teneur de coupure z sur un estimateur Z(v)* de Z(v). En moyenne on s'attend à récupérer une teneur:

$$E[Z(v)^*|Z(v)^*>z]$$

("espérance de Z(v)* pour les valeurs de Z(v)* dépassant z"), et donc à réaliser un gain en teneur égal à:

$$E[Z(v)^*|Z(v)^*>z]-m$$

On fait l'hypothèse que l'estimateur des teneurs de blocs est sans biais:

$$E[Z(v)^*]=m=E[Z(v)]$$

et que la régression de Z(v) connaissant Z(v)* est linéaire. Elle s'écrit donc:

$$E[Z(v)|Z(v)^*]=pZ(v)^*+(1-p)m,$$

avec une pente donnée par $p=cov(Z(v),Z(v)^*)/varZ(v)^*=\varrho\sigma_v/\sigma_{v^*}$ (où ϱ est la corrélation, inférieure à 1, voir Chap. 1).

Montrer que le gain en teneur effectivement réalisé, $E[Z(v)|Z(v)^*>z]-m$, égale le gain prévu multiplié par la pente de la

régression. En déduire qu'un estimateur Z(v)* plus variant que Z(v) ($\sigma_{v^*}^2 > \sigma_v^2$), par exemple une moyenne des échantillons intérieurs au bloc, surestime la teneur récupérée. Les promesses de sélectivité provenant de la trop grande dispersion des valeurs estimées ne peuvent être tenues. Dans la pratique, il est donc préférable de se rapprocher d'un estimateur sans "biais conditionnel":

$$E[Z(v)|Z(v)^*]=Z(v)^*$$

Ce problème minier important est à l'origine du krigeage (voir Chap. 5).

▶ *Exercice* 3.5. Effet de contraintes.

On considère un gisement minier constitué d'une pile verticale de blocs de tonnage 1 et de teneurs (en %):

5
5
5
5
0
0
0
10

Déterminer les projets d'exploitation maximisant le métal récupéré Q, pour un tonnage extrait T=1, 2, ..., 8, dans chacune des conditions suivantes:

– sélection libre des blocs;

– extraction en une passe verticale unique;

– extraction à ciel ouvert (avec talus verticaux).

Tracer et comparer les courbes Q(T). Remarquer la détérioration de la courbe quand les contraintes se font plus fortes. Dans quel cas la courbe est–elle concave?

On cherche, pour chaque valeur du paramètre z, les projets maximisant Q–zT. Quels projets sont–ils abandonnés et pourquoi?

(En pratique la combinatoire des projets possibles est rapidement inextricable et ce dernier paramétrage apporte une simplification importante: voir par exemple l'exercice suivant.)

▶ *Exercice* 3.6. Toit et mur de teneur.

Un profil vertical de teneurs z(x) étant donné (Fig. 10), on cherche à déterminer les projets optimaux d'extraction en une passe verticale unique. Pour cela on considère les projets maximisant Q−zT pour chaque valeur du paramètre z. Quelle condition simple doivent alors vérifier nécessairement les côtes b du toit et a du mur d'un tel projet? Remarquer que ce résultat permet une identification rapide des projets possibles, avant le tri final.

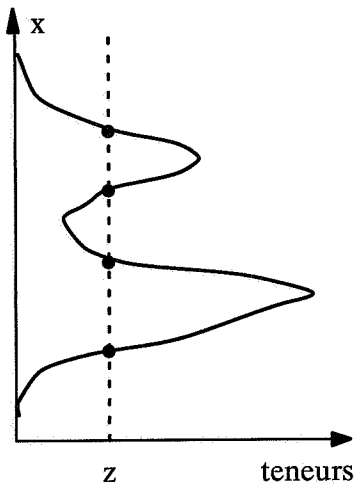

Fig. 10: Profil vertical de teneurs.

Réponse : Partir de T=b–a et Q=$\int_a^b z(x)dx$. Si les teneurs z(x) sont continues le long du profil, la quantité Q–zT admet z(b)–z et –z+z(a) comme dérivées par rapport à b et a. Ecrivant que ces dérivées s'annulent en un maximum, on obtient: z(a)=z et z(b)=z. Dans le cas général, on trouve que chacun des points a et b sert de transition entre teneurs inférieures à z et teneurs supérieures à z. En bref, a et b sont nécessairement des points où la teneur franchit la valeur du paramètre z.

▶ *Exercice* 3.7. Carrières optimales.

On cherche à exploiter à ciel ouvert un gisement peu profond. L'extraction d'un bloc suppose l'extraction préalable de tous les blocs situés au−dessus dans un cône (pointe en bas) défini par les contraintes de pente. L'intersection ou la réunion de deux projets de

fosse est encore une fosse. A chaque projet de fosse correspond un tonnage remué T et une quantité de métal Q. Appelons projets optimaux pour le paramètre z ceux qui maximisent $B_z=Q-zT$. Montrer que les projets optimaux, quand z varie, sont des fosses emboîtées les unes dans les autres. (Ce résultat est à la base d'une méthode de détermination des carrières optimales.)

Réponse : Considérons deux fosses F_z et $F_{z'}$ optimales respectivement pour z et z'>z. Supposons que $F_{z'}$ ne soit pas emboîtée dans F_z et appelons $\cap$ l'intersection des deux fosses. La contribution à $B_{z'}$ de la fosse $F_{z'}$ amputée de $\cap$ est positive ou nulle, sinon $\cap$ serait, pour z', une fosse meilleure que $F_{z'}$, qui ne serait plus optimale. Comme z'>z, cette fosse amputée apporterait une contribution strictement positive à B_z: pour la valeur z, l'union des deux fosses serait alors meilleure que la fosse F_z, qui ne serait plus optimale.

4 LE VARIOGRAMME

Derrière les variances, le variogramme

L'évolution de la variance en fonction du support, vue au chapitre précédent, est liée, on le devine, à l'agencement dans l'espace des teneurs de nos morceaux élémentaires: que des morceaux voisins aient des teneurs voisines ou non, cela se répercutera directement sur les teneurs des blocs.

D'ailleurs la variance de dispersion des petits morceaux x_i, en nombre N, dans le gisement:

$$\frac{1}{N}\sum_{i}[z(x_i)-z(V)]^2$$

peut s'écrire aussi comme demi−moyenne du carré des écarts entre tous les couples $(z(x_i), z(x_j))$ possibles:

$$\frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[z(x_i)-z(x_j)]^2$$

Classons ces couples de petits morceaux selon la distance qui les sépare, et notons N(h) le nombre de couples décalés de h. La variance apparait alors comme la moyenne (pondérée par le nombre de couples) d'une quantité dépendant de la distance:

$$\frac{1}{\sum_{h} N(h)}\sum_{h} N(h)\gamma(h)$$

avec: $$\gamma(h) = \frac{1}{2N(h)}\sum_{x_i-x_j=h}[z(x_i)-z(x_j)]^2$$

C'est cette quantité $\gamma(h)$, représentant la demi−variabilité moyenne à distance h, qu'on appelle variogramme (Fig. 11). Le variogramme est donc un outil particulier, mesurant la structure spatiale, et lié étroitement aux variances.

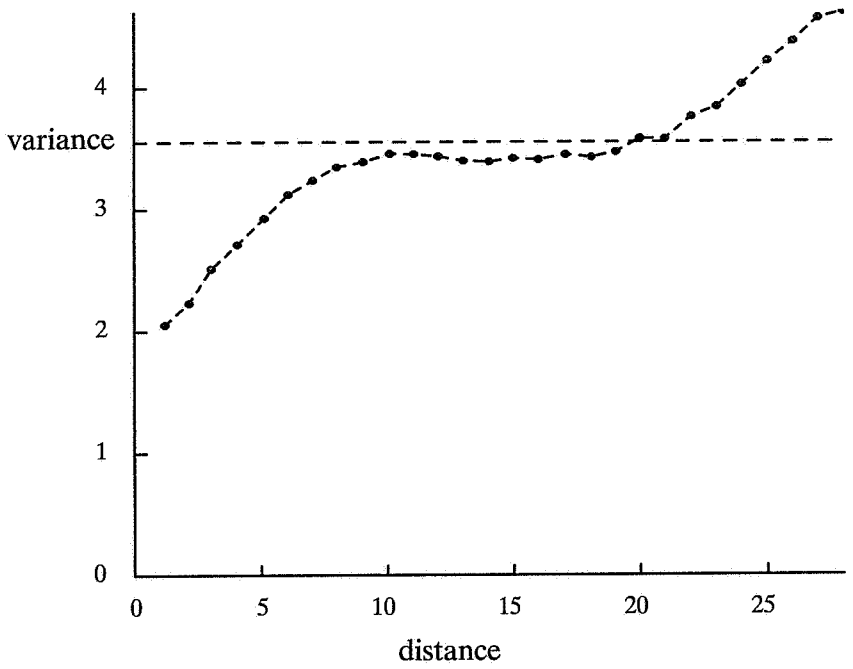

Fig. 11: Variogramme des petits morceaux.

Le variogramme expérimental

En pratique les variables auxquelles on s'intéresse sont très souvent mesurées sur un support élémentaire petit, quasi−ponctuel, que l'on assimile à un point. Pour une distance donnée, le variogramme se calcule en considérant les couples de points séparés par cette distance, en faisant pour chaque couple le carré de la différence des valeurs, et en en prenant la demi−moyenne.

Reprenant l'exemple du Chap. 1, voici 10 valeurs mesurées tous les 10 mètres le long d'une ligne:

1 2 0 0 1 0 0 1 2 3

Le variogramme à distance 10 mètres (d'ailleurs égal à celui à distance −10 mètres), est égal à:

$$\gamma(10)=(1/2)[(1-2)^2+(2-0)^2+(0-0)^2+(0-1)^2+ (1-0)^2+(0-0)^2+(0-1)^2+(1-2)^2+(2-3)^2]/9$$

et on trouve ainsi:

distance	variogramme	nombre de paires
0	0.	10
10	0.56	9
20	1.00	8
30	1.07	7
40	1.25	6
50	1.40	5
60	1.88	4
70	1.50	3
80	0.50	2
90	2.00	1

Le variogramme est ainsi calculé pour les multiples du pas des données. Si les données ne sont pas régulièrement espacées, les distances sont regroupées par classe: par exemple les classes correspondant aux multiples d'un pas choisi, à une tolérance d'un demi−pas près.

A plusieurs dimensions, le variogramme dépend d'une distance h qui est en fait un vecteur distance. On peut alors parler de variogramme directionnel lorsqu'on calcule un variogramme pour une distance h ayant une direction donnée. Dans l'exemple du chapitre précédent, le variogramme "Est−Ouest" serait le variogramme moyen de chacune des lignes Est−Ouest. Lorsque les données ne sont pas implantées régulièrement, on convient d'une tolérance sur la direction comme sur les distances.

Le variogramme exhaustif

Le variogramme d'une variable régionalisée est d'autant mieux connu que la densité d'information est élevée. En tout état de cause, le "vrai" variogramme est celui qui serait calculé à partir d'une connaissance exhaustive de la variable sur son champ. En pratique, un tel variogramme est inconnu, et le variogramme calculé expérimentalement sur les données disponibles en fournit une estimation.

Parfois le champ est reconnu avec une première maille, puis celle−ci est resserrée sur une partie du champ où les valeurs sont plus élevées et plus variables. Il convient alors de distinguer le variogramme calculé à partir de la maille large, du variogramme de la partie mieux reconnue, calculé avec la maille resserrée. Mélanger d'emblée toutes ces données peut produire un résultat hétérogène et dénué de signification.

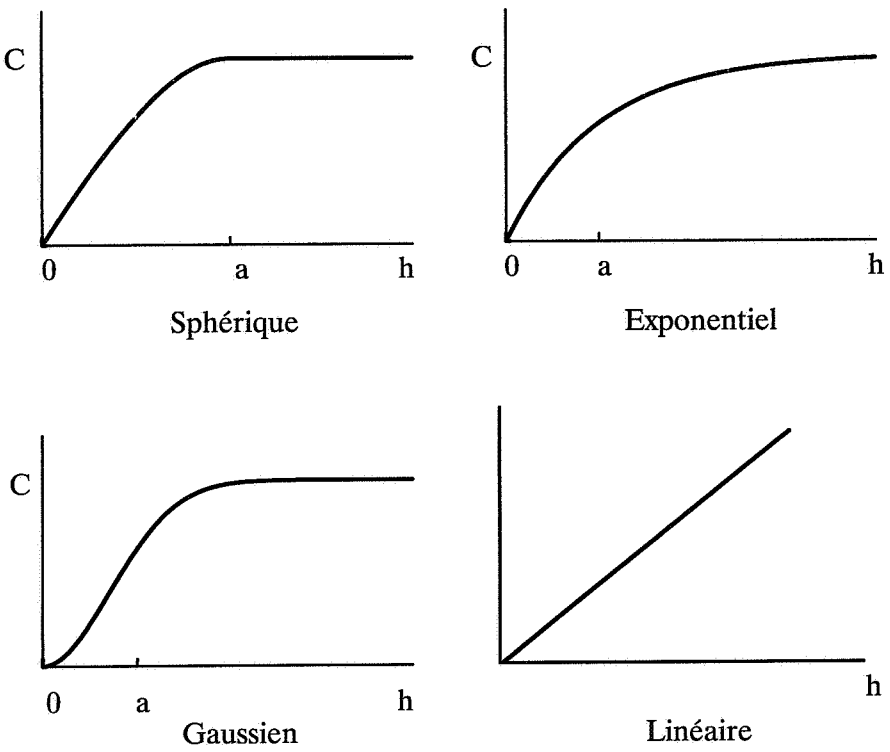

Fig. 12: Quelques modèles courants de variogrammes:

Sphérique: $\gamma(h) = C\left[\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|}{a}\right)^3\right]$ pour $|h|<a$, $\gamma(h)=C$ sinon.

Exponentiel: $\gamma(h) = C[1 - \exp - (|h|/a)]$

Gaussien: $\gamma(h) = C[1-\exp(-h^2/a^2)]$

Linéaire (les puissances $|h|^{\alpha}$ sont des modèles autorisés pour $\alpha<2$).

Le modèle de variogramme

En pratique, le géostatisticien dégage les caractéristiques importantes du variogramme expérimental, et peut alors lui ajuster une fonction mathématique (le variogramme théorique ou modèle de variogramme), qui peut être considérée comme une estimation du vrai variogramme.

N'importe quelle fonction de la distance h n'est pas habilitée à être un modèle de variogramme. Par exemple la fonction valant 0 pour $|h|<a$ et $|h|-a$ sinon, bien que positive et vérifiant $\gamma(h)=\gamma(-h)$, ne peut être un variogramme: la variabilité ne saurait être rigoureusement nulle à petite distance, sans l'être à toute distance. Par ailleurs on peut montrer qu'une fonction comme celle valant $|h|$ pour $|h|<a$, et a sinon, est un variogramme à 1D mais non à 2D ou plus. Nous verrons, chapitre suivant, les conditions mathématiques que doit vérifier une fonction pour servir de variogramme. En pratique, on fait appel habituellement à quelques modèles répertoriés (exemples Fig. 12), ou à une somme de tels modèles.

Propriétés

Il y a isotropie lorsque le variogramme est le même pour toutes les directions (autrement dit lorsque le variogramme dépend du module $|h|$ de la distance h, et non de la direction de cette distance). Dans le cas contraire il y a anisotropie. Souvent, anisotropie dite géométrique, il est possible de se ramener à une figure isotrope par simple transformation linéaire des coordonnées. Voir exercices 5.5 et 6.1.

Le comportement du variogramme au voisinage de l'origine est révélateur du degré de régularité de la variable: Fig. 13. Par définition le variogramme est nul entre un point et lui-même: $\gamma(0)=0$. Parfois il existe un décrochement à l'origine: lorsque la distance h tend vers 0, le variogramme tend vers une valeur C_0 non nulle (il est donc discontinu en 0). Cet *effet de pépite*, terme qui rappelle le domaine minier, correspond à une variable z(x) discontinue, dans le sens où, en moyenne, $[z(x+h)-z(x)]^2$ ne tend pas vers 0 (absence de continuité "en moyenne quadratique"). En pratique il est dû à une microstructure inaccessible à cause de l'espacement des données, ou à des erreurs de mesures. Lorsque, pour h non nulle, le variogramme se maintient à C_0, on dit que la structure est pépitique pure ou aléatoire pure: la variable n'est pas structurée spatialement, c'est un bruit blanc.

Au contraire un variogramme sans effet de pépite correspond à une variable continue. Si de plus la variable est dérivable, dans le sens où, en moyenne, $\{[z(x+h)-z(x)]/h\}^2$ tend vers une valeur finie, alors le variogramme a un comportement parabolique (en h^2) à l'origine. Ceci correspond à une variable hautement régulière.

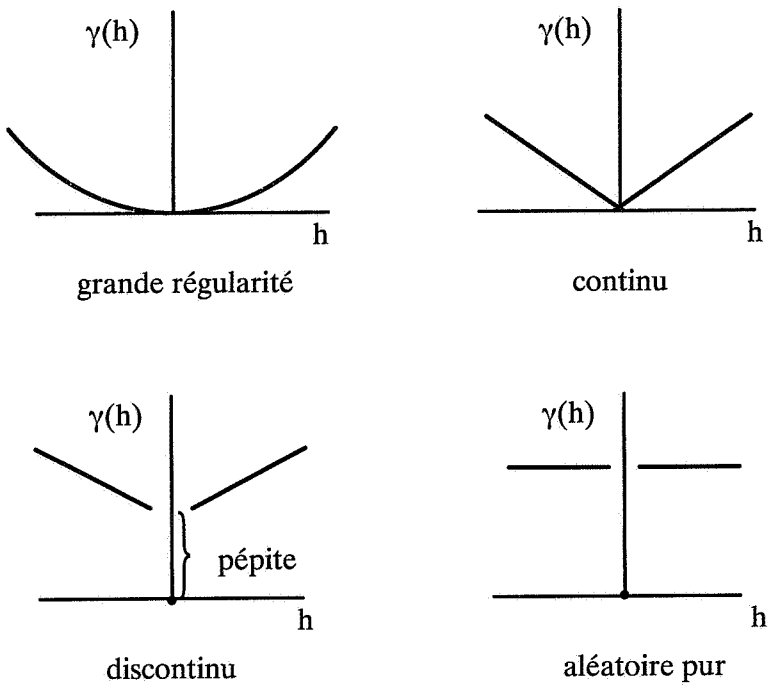

Fig. 13: Comportement du variogramme à l'origine.

En pratique cependant, on observe très souvent un comportement linéaire à l'origine, ou un comportement en $|h|^{\alpha}$, avec $\alpha<2$. La variable est alors continue mais non dérivable (toujours "en moyenne quadratique", voir par exemple exercice 4.1). Remarquer alors que, étant identique pour un vecteur distance h et son opposé $-h$: $\gamma(-h)=\gamma(h)$, le variogramme lui-même présente, pour $\alpha \leq 1$, une angularité en 0 et n'y est pas dérivable.

En général la variabilité moyenne entre deux points augmente avec la distance qui les sépare. Certains variogrammes continuent ainsi à augmenter jusqu'aux grandes distances. Mais souvent le variogramme atteint, à une distance appelée *portée*, un *palier* sur lequel il se stabilise (Fig. 14 a et b). Cette portée fixe la limite de la zone d'influence d'un échantillon.

Parfois, après une première croissance, le variogramme redescend et présente un minimum local: cet "effet de trou" est l'indice d'une pseudo-période, distance à laquelle la variabilité se trouve abaissée (Fig. 14c).

Enfin le variogramme peut montrer l'existence de structures emboîtées à différentes échelles (Fig. 14d).

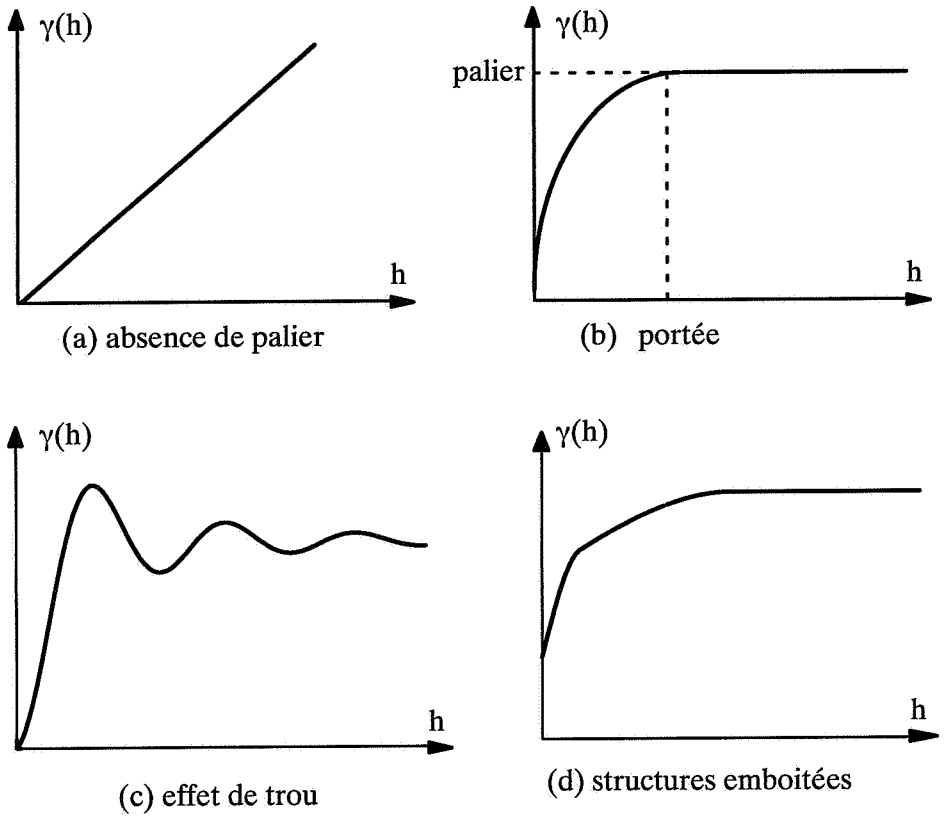

Fig. 14: Quelques traits du variogramme.

Exercices

▶ *Exercice* 4.1. Mosaïque.

On considère, à 1D, un champ divisé en un nombre fini de compartiments, et une variable régionalisée constante à l'intérieur de chaque compartiment: mathématiquement il s'agit donc d'une fonction constante par morceaux.

Cette variable est supposée connue sur les N points d'une grille régulière à maille très fine, bien inférieure à la dimension des compartiments.

Montrer que, dans le calcul du variogramme au pas numéro h, le nombre de couples est égal à N−h. Montrer qu'aux premiers pas, les couples à cheval sur deux compartiments sont au nombre de Mh, où M est le nombre de discontinuités à l'intérieur du champ. En déduire

qu'aux premiers pas, le variogramme est proportionnel à $Mh/(N-h)$, soit $(M/N)h$ en première approximation (comportement linéaire à l'origine): la variable est donc continue en moyenne quadratique, mais non dérivable en moyenne quadratique.

▶ *Exercice* 4.2. Marche aléatoire.

Il s'agit de construire une marche aléatoire simple, à partir de tirages à pile ou face successifs d'une pièce de monnaie. Partant de $Z_0=0$, on construira de proche en proche $Z_1, Z_2, ..., Z_i, ...$ en ajoutant au résultat précédent, soit Z_{i-1}, la variable Y_i, égale à 1 si face apparaît et à -1 si c'est pile:

$$Z_i=Z_{i-1}+Y_i$$

Calculer le variogramme aux premiers pas. Quel est son comportement?

Réponse : Le variogramme a un comportement linéaire... pourvu que le nombre de tirages ne soit pas trop faible.

5 LE CADRE INTRINSEQUE

Structure d'un phénomène

La structure spatiale, au sens large, d'une variable régionalisée est tout simplement la manière dont se comporte la variable dans son champ. Fait évidemment partie au premier chef de la structure spatiale le partage éventuel du champ en champs plus petits bien identifiés et de caractéristiques différentes: par exemple distinction, dans certains gisements métalliques, entre une zone à minerai oxydé et une zone à minerai sulfuré.

C'est ensuite, à l'intérieur du champ ou sous–champ choisi, que l'on pourra décrire la structure à l'aide, par exemple, du variogramme. Il est toujours possible, sur un jeu de données distribuées dans l'espace, de calculer un variogramme expérimental. Il suffit, pour un vecteur distance h donné, de considérer tous les couples de valeurs distants de h. On accède alors à une mesure expérimentale de la variabilité moyenne à distance h. Une telle statistique, cependant, fait jouer le même rôle à chacun des couples à distance h: elle n'aura véritablement de sens que si les différentes paires de valeurs obéissent, d'une certaine façon, à la même "loi".

En particulier, cela signifie que le comportement de ces couples n'est pas affecté par la plus ou moins grande proximité des frontières du champ. Au fond, le champ peut, dans cette optique, être considéré comme une fenêtre à travers laquelle on observe un phénomène régi par une certaine loi de comportement, phénomène qu'on pourrait imaginer se déployer jusqu'à l'infini.

Les Fonctions Aléatoires (FA)

Pour décrire un tel phénomène avec sa variabilité, il est très commode de recourir à un modèle probabiliste de fonction aléatoire. Supposons, dans un tirage de dé, que sorte la face portant le numéro 5. Cette valeur 5 peut être considérée comme la "réalisation" d'une variable aléatoire Z pouvant prendre l'un quelconque des nombres de 1 à 6.

Considérons maintenant la suite de tirages en des instants t_1, t_2, t_3, t_4. Soit 5, 2, 1, 4 le résultat. Cet ensemble de valeurs peut être également considéré comme une réalisation de l'ensemble des VA indexées par le temps $Z(t_1)$, $Z(t_2)$, $Z(t_3)$, $Z(t_4)$. On peut imaginer d'autres résultats, mais il n'en demeure pas moins que c'est celui–ci, et lui seul, qui est sorti réellement à ces instants.

De même, en géostatistique, on considérera l'ensemble des valeurs $z(x_1)$, $z(x_2)$, ..., $z(x_n)$ de la variable régionalisée en des points x_1, x_2, ..., x_n du champ, comme une réalisation d'un ensemble de VA $Z(x_1)$, $Z(x_2)$, ..., $Z(x_n)$. Plus synthétiquement on considérera l'ensemble des valeurs $z(x)$ comme une réalisation d'une VA $Z(x)$ indexée par la position dans l'espace, ce que l'on appelle une fonction aléatoire (FA en abrégé).

Les Fonctions Aléatoires Intrinsèques

C'est le modèle qui permet, si l'on veut, de donner un sens au variogramme. On appelle modèle de fonction aléatoire intrinsèque (en abrégé FAI) une fonction aléatoire $Z(x)$ dont les accroissements $Z(x+h)-Z(x)$ entre points à distance h vérifient les propriétés:

(i) $E[Z(x+h)-Z(x)]=0$

(ii) $var[Z(x+h)-Z(x)]$, alors égale à $E[(Z(x+h)-Z(x))^2]$, $=2\gamma(h)$

La première condition traduit l'absence de "dérive", c'est-à-dire le fait qu'il n'existe aucune variation systématique entre deux points (hypothèse par exemple incompatible, si l'on étudie la profondeur du fond marin en bordure de côte, avec l'approfondissement en s'éloignant des côtes).

La seconde condition signifie que la variabilité entre deux points x et x+h dépend du vecteur distance h entre ces deux points, mais non de la position x: elle est la même qu'entre y et y+h. D'où le variogramme $\gamma(h)$, version théorique du variogramme expérimental.

Considérons maintenant des points quelconques de l'espace x_1, x_2, ..., x_n et des coefficients λ_1, λ_2, ..., λ_n. On montre que, sous la condition: $\Sigma\lambda_i=0$, la combinaison linéaire (dite alors autorisée, en abrégé CLA):

$$\sum_i \lambda_i Z(x_i)$$

possède une espérance nulle, et une variance qui peut s'exprimer à l'aide du variogramme:

$$\text{var} \sum_i \lambda_i Z(x_i) = - \sum_i \sum_j \lambda_i \lambda_j \gamma(x_i - x_j)$$

Le modèle de variogramme doit assurer la positivité d'une telle variance. Mathématiquement, ceci signifie que son opposé $-\gamma(h)$ (à

cause du signe − dans l'expression) doit appartenir aux fonctions de type positif conditionnel (celles qui garantissent précisément la positivité d'une telle expression sous la condition $\Sigma\lambda_i=0$). C'est en particulier le cas des modèles présentés chapitre précédent.

Variances d'estimation et variances de dispersion

Outre son pouvoir descriptif, un des intérêts du variogramme va être précisément de pouvoir calculer des variances.

Exemple trivial: estimation de Z(x) par Z(x+h). L'erreur d'estimation Z(x+h)−Z(x) est une combinaison linéaire autorisée (somme des poids: 1−1=0), et sa variance vaut précisément $2\gamma(h)$.

Autre exemple: estimation de la moyenne de Z sur un bloc v, soit $Z(v)=\frac{1}{V}\int_v Z(y)dy$ par Z(x). L'erreur Z(v)−Z(x) est une combinaison linéaire autorisée (somme: $\frac{1}{V}\int_v dy - 1 = 1-1 = 0$) et sa variance égale:

$$2\gamma(x,v)-\gamma(v,v)$$

avec $\gamma(x,v) = \frac{1}{V}\int_v \gamma(x-y)dy$: valeur moyenne de γ entre le point x et un point décrivant v,

et $\gamma(v,v)$: valeur moyenne de γ entre 2 points décrivant v indépendamment.

Cette variance est celle de l'erreur faite en estimant Z(v) par Z(x), ou bien d'ailleurs Z(x) par Z(v). Une telle variance, variance d'une erreur d'estimation, est appelée *variance d'estimation*.

Considérons maintenant, non plus un point fixé, mais un point x intérieur et aléatoire uniforme dans v. En moyenne pour toutes les positions possibles de ce point, on obtient simplement:

$$\text{var}[Z(x)-Z(v)]=\gamma(v,v)$$

Comme le point x peut prendre toutes les positions à l'intérieur de v, il s'agit là de la *variance de dispersion* d'un point dans un bloc de support v.

Notons bien la distinction, fondamentale en géostatistique, entre une variance d'estimation, variance de l'erreur lorsqu'on estime une

quantité par une autre, et variance de dispersion, indiquant la variabilité des valeurs à l'intérieur d'un certain domaine. Voyons cela plus généralement.

Variances de dispersion : Si l'on considère maintenant un domaine V, réunion de blocs de support égal à v, eux−mêmes divisibles en petits morceaux assimilés à des points, on a, comme au chapitre 3, additivité des variances:

variance de dispersion d'un point dans V (soit $\gamma(V,V)$)
= variance de dispersion d'un point dans v (soit $\gamma(v,v)$)
+ variance de dispersion de v dans V

ou, de façon équivalente:

variance de dispersion de v dans V$=\gamma(V,V)-\gamma(v,v)$.

En pratique $\gamma(v,v)$, variabilité intérieure aux blocs v, représente la chute de variance lorsqu'on passe du support ponctuel au support bloc, ce dont nous verrons une utilisation au chapitre 9. Ce passage s'accompagne donc d'une régularisation.

Naturellement, la variabilité spatiale est elle aussi modifiée par la régularisation sur bloc. Le variogramme $\gamma_v(h)=\frac{1}{2}E[(Z(v_h) - Z(v))^2]$ de la variable régularisée sur bloc est donné par:

$$\gamma_v(h)=\gamma(v,v_h)-\gamma(v,v)$$

où h est le décalage entre les blocs v et v_h, et $\gamma(v,v_h)$ la valeur moyenne du variogramme ponctuel entre un point parcourant v et un autre décrivant v_h. Lors de la régularisation, la portée se trouve augmentée de la dimension du bloc dans la direction considérée. Le comportement à l'origine devient plus régulier. Ce sont naturellement les structures aux plus courtes distances qui sont affectées le plus par la régularisation (absorption de la composante pépitique en particulier). Voir exercice 5.11.

Variances d'estimation : Supposons que l'on se propose d'estimer la valeur moyenne $Z(V)=\frac{1}{V}\int_V Z(y)dy$ sur un domaine V, à partir de la moyenne $\frac{1}{n}\sum_i Z(x_i)$ des valeurs en des points x_1, x_2, ... x_n. L'erreur $Z(v)-\frac{1}{n}\sum_i Z(x_i)$ est une combinaison linéaire dont la somme des

coefficients $\frac{1}{V}\int_v dy-(1/n)n$ vaut $1-1=0$, et sa variance est donnée par:

$$-\gamma(v,v) + \frac{2}{n}\sum_i \gamma(x_i,v) - \frac{1}{n^2}\sum_{i,j} \gamma(x_i-x_j)$$

On voit qu'elle ne dépend que de la structure variographique et de la configuration géométrique du domaine à estimer et de son échantillonnage.

Principes d'approximation

La formule précédente est très utile pour l'estimation globale. Des approximations ont été développées, qui permettent, dans certains cas courants, de calculer facilement cette variance.

A 1D: On considère une ligne de longueur L=Na reconnue par une maille régulière centrée (un point x_i au centre de chaque segment l_i de longueur a). L'erreur d'estimation:

$$Z(L)-\frac{1}{N}\sum Z(x_i)=\frac{1}{N}\sum Z(l_i)-\frac{1}{N}\sum Z(x_i)=\frac{1}{N}\sum[Z(l_i)-Z(x_i)]$$

est donc la somme des erreurs élémentaires d'estimation de chaque segment uniquement par son centre. En considérant, principe d'approximation, que ces erreurs sont sans corrélation, on déduit que la variance d'estimation de la ligne égale 1/N fois la variance d'estimation élémentaire d'un segment par son centre:

$$E\left(\frac{1}{N}\sum[Z(l_i)-Z(x_i)]^2\right)=\frac{1}{N}\sum E[[Z(l_i)-Z(x_i)]^2] = \frac{1}{N}\sigma_1^2$$

Attention: si l'on estimait chaque segment à l'aide des données de son voisinage, les corrélations entre erreurs de segments voisins ne pourraient plus être négligées et le principe d'approximation cesserait d'être applicable.

A 2D, maille carrée: résultat analogue avec la variance d'estimation élémentaire d'un carré par son centre.

A 2D, lignes et tranches: Les données sont resserrées le long de lignes j parallèles et équidistantes de longueur L_j. Pour le calcul de la variance, on considère que chaque ligne L_j est la ligne médiane d'une tranche v_j, la réunion de ces tranches formant le gisement. L'erreur

d'estimation de la moyenne du gisement par la moyenne des échantillons se décompose ainsi (en introduisant tranches, lignes et segments):

$$Z(V)-\frac{1}{N}\sum Z(x_i)=\frac{1}{\sum L_j}\sum_j L_j[Z(v_j)-Z(L_j)]+\frac{1}{N}\sum[Z(l_i)-Z(x_i)]$$

Le principe d'approximation consiste à négliger les différentes corrélations entre erreurs d'estimation des lignes par leurs échantillons et des tranches par leur ligne, et on obtient finalement:

$$\frac{1}{\sum L_j^2}\sum L_j^2\sigma_{v_j}{}^2+\frac{1}{N}\sigma_1^2$$

où σ_1^2 est la variance d'estimation élémentaire vue précédemment à 1D et $\sigma_{v_j}{}^2$ celle de la tranche v_j par sa ligne.

De 2D à 1D: Dans le cas d'un échantillonnage continu (exemple de la densité en poissons mesurée par acoustique le long de transects), les lignes sont parfaitement connues et il ne subsiste plus que le terme de tranche. Il est alors encore plus simple de se ramener à 1D en considérant comme variable l'abondance cumulée le long de chaque transect. Cette remarque importante vaut naturellement aussi pour les méthodes transitives.

Le krigeage

Nous avons signalé (Chap. 3, en particulier exercice 3.4) le risque que présente, dans une exploitation minière sélective, l'estimation des valeurs inconnues de blocs par leur échantillonnage. Pour éviter ce danger dû à la variabilité naturellement trop élevée de l'échantillonnage, D.G. Krige proposait de prendre un estimateur basé sur une régression (sous des hypothèses lognormales adaptées aux gisements d'or sur lesquels il travaillait).

A sa suite, G. Matheron propose l'estimateur baptisé par lui krigeage, estimateur linéaire ne requérant que la connaissance du variogramme. La valeur d'un bloc Z(v) se voit alors estimée par une moyenne pondérée des données présentes à l'intérieur d'un voisinage de ce bloc. Dans un modèle de FAI, l'erreur $Z(v)-Z(v)^*$ faite en estimant Z(v) par la moyenne pondérée $\sum\lambda_i Z(x_i)$ est une combinaison linéaire, dont la somme des coefficients $1-\sum\lambda_i$ est

nulle pourvu que la somme des poids affectés aux échantillons xi soit égale à 100%:

$$\sum \lambda_i = 1$$

La variance d'estimation peut s'écrire alors:

$$-\gamma(v,v) + 2\sum_i \lambda_i \gamma(x_i, v) - \sum_{i,j} \lambda_i \lambda_j \gamma(x_i - x_j)$$

Le meilleur estimateur sera obtenu en choisissant les poids λ_i qui conduisent à la variance la plus faible (sous la condition $\sum \lambda_i = 1$). Ces poids apparaissent alors comme solution du système linéaire:

$$\begin{cases} \sum_j \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_i, v) \quad \text{pour tout i} \\ \sum_i \lambda_i = 1 \end{cases}$$

Ainsi le krigeage permet-il de donner, aux divers échantillons, les poids les meilleurs, compte tenu de la configuration géométrique de l'ensemble et de la structure. Le plus souvent les échantillons éloignés du bloc reçoivent un poids quasi-nul, les échantillons plus proches leur faisant écran, ce qui facilite le choix du voisinage. On appelle "variance de krigeage" la variance de l'erreur d'estimation ainsi minimisée.

Lorsque v est réduit à un point, on a un krigeage ponctuel. Et on peut montrer que, le voisinage étant fixé, le krigeage d'un bloc coïncide avec la moyenne de celui des différents points du bloc. En dehors des applications minières, le krigeage, appliqué à l'estimation de chacun des points d'une grille, apparait comme une méthode de cartographie. Mais d'autres modèles que les FAI, basés sur des hypothèses de stationnarité différentes, sont également utilisables, comme on le verra plus loin.

Exercices

▶ *Exercice* 5.1. Effet d'une dérive linéaire sur le variogramme.

Dans un modèle intrinsèque, espérance et variance des accroissements Z(x+h)−Z(x) ne dépendent pas de x: elles sont dites stationnaires. De plus E[Z(x+h)−Z(x)]=0, ce qui traduit l'absence de dérive.

Considérons maintenant un modèle dans lequel:

(i) il existe une dérive linéaire: $E[Z(x)]=ax+b$;

(ii) $var[Z(x+h)-Z(x)]$ ne dépend que de h.

Montrer:

– que espérance et variance des accroissements sont encore stationnaires;

– que la dérive se traduit par une composante parabolique croissant comme h^2 sur le variogramme que l'on calculerait expérimentalement selon la formule $(1/2)E[(Z(x+h)-Z(x))^2]$.

Remarque: on démontre que la variance d'accroissements stationnaires $var[Z(x+h)-Z(x)]$ augmente nécessairement moins vite que h^2 aux grandes distances. Une croissance parabolique sur $(1/2)E[(Z(x+h)-Z(x))^2]$ atteste donc de l'existence d'une dérive.

▶ *Exercice* 5.2. Dispersion d'un point parmi n.

Z(x) est une FA intrinsèque de variogramme $\gamma(h)$. Soient n points x_1, x_2, ..., x_n. Variance de dispersion d'un point parmi ces n points?

Réponse : $\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(x_i-x_j)$

Cas d'un effet de pépite pur ($\gamma(h)=C_0$ pour $h\neq 0$)?

Réponse : $(1-1/n)C_0$.

▶ *Exercice* 5.3. Estimation en un point par une moyenne sur n points.

Z(x) est une FA intrinsèque de variogramme $\gamma(h)$. Variance d'estimation en un point x_0 par la moyenne en n points x_1, x_2, ..., x_n?

Réponse : $\frac{2}{n}\sum_{i=1}^{n}\gamma(x_i-x_0) - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(x_i-x_j)$

Cas d'un effet de pépite pur?

Réponse : $(1+1/n)C_0$.

▶ *Exercice* 5.4. Régularisation.

Z(x) est une FA intrinsèque de variogramme isotrope linéaire. On considère sa régularisée sur le support constitué par un segment

vertical de longueur L. Comment la régularisation affecte−t−elle le variogramme vertical?

Réponse : La régularisation selon la verticale augmente la continuité dans cette direction: le calcul montre qu'une croissance parabolique apparait à l'origine sur le variogramme vertical.

Le variogramme vertical est−il égal au variogramme "monté" (on appelle ainsi le variogramme horizontal de la régularisée verticale)?

Réponse : Non, la variable régularisée n'est plus isotrope.

▶ *Exercice* 5.5. Anisotropie géométrique.

Considérons un variogramme 2D $\gamma(h_x, h_y)$ qui ne dépend que de $\sqrt{h_x^2 + 4h_y^2}$. Quelles sont les directions de continuité et de variabilité maximales?

Réponse : Continuité maximale selon x, variabilité maximale selon y.

Quelle transformation simple permettrait de se ramener à une figure isotrope?

Réponse : Multiplier les y par 2.

On cherche à estimer un carré par une ligne médiane, orientée soit selon x, soit selon y. Laquelle conduit à la meilleure estimation?

Réponse : La continuité étant selon x, c'est en la coupant, avec une ligne selon y, qu'on aura la meilleure estimation. On peut s'en convaincre en se ramenant au cas isotrope, le rectangle obtenu étant mieux estimé par sa médiane la plus longue.

▶ *Exercice* 5.6. Segment estimé par son centre ou ses extrémités.

On considère un variogramme puissance $\gamma(h) = |h|^{\alpha}$.

Variogramme moyen entre (un point parcourant) le segment et une extrémité?

Variogramme moyen sur le segment (c'est−à−dire entre deux points parcourant indépendamment le segment)?

Variance d'estimation du segment par son centre? par ses extrémités?

Comparer selon les valeurs de α.

Réponse : Pour $\alpha>1$, forte continuité, le point central seul donne une meilleure précision que les deux extrémités. C'est le contraire pour $\alpha<1$. Et pour $\alpha=1$, les deux configurations sont équivalentes.

▶ *Exercice* 5.7. Echantillonnage d'une ligne.

Comparer, pour un variogramme puissance, l'estimation d'une ligne L=Na par N points selon les trois stratégies suivantes:

– maille régulière centrée;

– échantillonnage aléatoire stratifié (cf Chap. 3);

– échantillonnage aléatoire pur.

▶ *Exercice* 5.8. Krigeage en présence de lacunes.

Z(x) est une FAI de variogramme isotrope $\gamma(h)$ à 2D. On cherche à estimer la valeur en un point O, à partir de quatre points A, B, C et D répartis régulièrement sur un cercle de rayon h centré en O. Poids de krigeage affecté à chacun de ces quatre points?

Réponse : 0.25.

On enlève le point D. Comparer les poids de krigeage affectés aux points A, B et C. Cas d'un variogramme purement pépitique? d'un variogramme linéaire?

Réponse : $\lambda_A=\lambda_C=\gamma(h\sqrt{}/2)/[4\gamma(h\sqrt{}/2)-\gamma(2h)]$, $\lambda_B=1-2\lambda_A$. Cas pépitique pur: tous les poids sont égaux à 0.333. Cas linéaire: les points les plus proches de la lacune D reçoivent chacun un poids 0.387, plus fort que le poids 0.226 du point opposé à la lacune.

▶ *Exercice* 5.9. Krigeage en présence de redondances.

Z(x) est une FAI de variogramme isotrope $\gamma(h)$ à 2D. On cherche à estimer la valeur en un point O, à partir de quatre points A, B, C et D situés sur un cercle de rayon h centré en O. Les points C et D sont très proches et forment le sommet d'un triangle presque équilatéral ayant A et B comme autres sommets. Poids de krigeage affecté à chacun de ces quatre points? Cas pépitique pur? Cas d'un variogramme sans effet de pépite?

Réponse : $\lambda_C=\lambda_D=(1/2)/[3-C_0/\gamma(h\sqrt{}/3)]$, $\lambda_A=\lambda_B=0.5-\lambda_C$. Ces poids dépendent de l'importance de la composante pépitique C_0. Ils sont tous égaux à 1/4 pour un effet de pépite pur. En l'absence d'effet de pépite, les points C et D sont redondants et reçoivent au total le même poids (1/3) qu'aurait reçu C en l'absence de D.

▶ *Exercice* 5.10. Krigeage d'un point connaissant deux autres points.

Z(x) est une FAI de variogramme $\gamma(h)$ à 1D. On veut kriger la valeur Z_0 en un point x_0, à partir des valeurs Z_1 et Z_2 en deux points $x_1=x_0+h_1$ et $x_2=x_0+h_2$. Poids de krigeage?

Réponse : $\lambda_1=(1/2)[1+(\gamma(h_2)-\gamma(h1))/\gamma(h)]$,

et $\lambda_2=(1/2)[1-(\gamma(h_2)-\gamma(h_1))/\gamma(h)]$, avec $h=h_2-h_1$.

Que deviennent ces poids pour x_1 et x_2 de part et d'autre de x_0 ($h_1>0$ et $h_2<0$) dans le cas d'un variogramme linéaire?

Réponse : Ils correspondent à une interpolation linéaire: $\lambda_1=|h_2|/(|h_1|+|h_2|)$, $\lambda_2=|h_1|/(|h_1|+|h_2|)$.

On suppose dorénavant x_1 et x_2 du même côté de x_0 ($h_1>0$ et $h_2-h_1=h>0$). Le poids de Z_2 peut-il être supérieur à celui de Z_1?

Réponse : Oui en cas d'effet de trou, non pour un variogramme non décroissant.

Montrer que pour un variogramme linéaire, le poids de Z_2 est nul: le point x_1 fait écran au point x_2. Montrer que le poids de Z_2 est positif pour un variogramme concave, et négatif pour un variogramme convexe.

Cas d'une FA dérivable: Montrer que le krigeage de Z_0 est une combinaison linéaire de $(Z_1+Z_2)/2$ et de $(Z_2-Z_1)/h$, avec des poids que l'on déterminera.

Réponse : Poids 1 et $-h[\gamma(h_2)-\gamma(h_1)]/2\gamma(h)$.

Que devient ce krigeage lorsque le point x_2 se rapproche de x_1?

Réponse : Z_2 tend vers Z_1, et donc $(Z_1+Z_2)/2$ tend vers Z_1; $(Z_2-Z_1)/h$ tend vers la dérivée de Z en x_1, avec un poids $-\gamma'(h_1)/\gamma''(0)$. Le krigeage est donc une combinaison de Z et de sa dérivée au point x_1.

▶ *Exercice* 5.11. Krigeage en présence d'un effet de pépite.

En un point de donnée, la meilleure estimation linéaire est égale à la valeur elle-même: le krigeage est donc un interpolateur exact. Montrer qu'en présence d'un effet de pépite, l'interpolateur de krigeage est discontinu à l'endroit des points de données.

Réponse : Les poids de krigeage dépendent linéairement du second membre du système de krigeage. Quand on se rapproche d'un point de donnée x_j, le second membre de l'équation correspondante, $\gamma(x-x_j)$, tend vers l'effet de pépite, et non vers $\gamma(x_j-x_j)=\gamma(0)=0$.

L'estimation obtenue par continuité:

– peut être interprétée comme le krigeage d'une microrégularisée de la variable (dont la structure ne diffère que par la disparition de l'effet de pépite);

– correspond au krigeage avec filtrage de l'erreur, dans le cas où l'effet de pépite est attribué à des erreurs de mesure indépendantes et de moyenne nulle.

6 LE CADRE STATIONNAIRE

Assez fréquemment le variogramme présente, à partir d'une distance appelée portée, un palier auquel il se stabilise. Ceci correspond à un phénomène dont les variations sont limitées et se font autour d'un niveau moyen m.

On va choisir alors de faire coïncider, dans un cadre probabiliste, ce paramètre de moyenne m avec l'espérance (valeur probable) de la fonction aléatoire Z(x) en tout point x:

$$E[Z(x)]=m$$

Etant constante, cette valeur probable ne dépend pas du point x (elle est "stationnaire"), d'où ici encore absence de dérive.

Définition d'un modèle stationnaire

On appellera modèle stationnaire (d'ordre 2, voir Chap. 1) un modèle dans lequel espérance et covariance sont stationnaires:

(i) l'espérance est constante $E[Z(x)]=m$;

(ii) la covariance entre deux points x et x+h, la même donc qu'entre deux points y et y+h, n'est fonction que de la distance h:

$$\text{cov}[Z(x), Z(x+h)]=E[(Z(x)-m)(Z(x+h)-m)]=C(h)$$

La valeur de C pour h=0:

$$C(0)=\text{cov}[Z(x), Z(x)]=\text{var } Z(x)=E[(Z(x)-m)^2]$$

donne la variance (stationnaire), fixant la variabilité de Z(x) autour de sa moyenne m.

Une FA stationnaire est aussi intrinsèque

Une FA stationnaire est également intrinsèque. En effet il est facile de vérifier que les accroissements $Z(x+h)-Z(x)$ ont une espérance nulle et une variance ne dépendant que de h. On obtient la relation liant variogramme et covariance:

$$C(0)-C(h)=\gamma(h)$$

Pour une distance donnée, plus la corrélation spatiale, mesurée par la covariance, est forte, plus la variabilité entre points, donnée par le variogramme, est faible: Fig. 15.

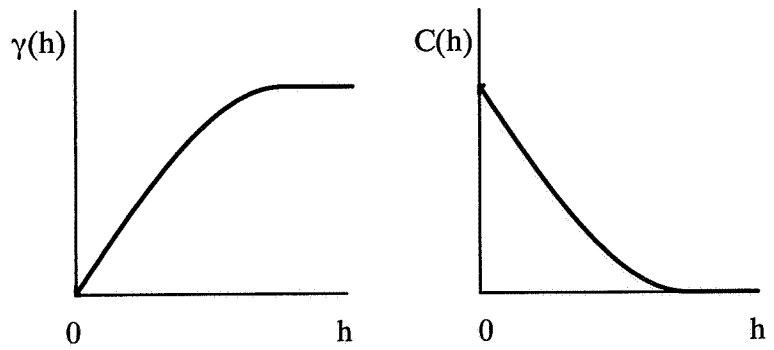

Fig. 15: Variogramme et covariance associés.

Réciproquement, le modèle intrinsèque, plus général, n'est pas forcément stationnaire (exemple: variogramme linéaire). Affranchi de la stationnarité, et ne nécessitant pas l'intervention du paramètre m, le variogramme apparaît comme un outil structural plus général que la covariance stationnaire. En contrepartie, l'intérêt du modèle stationnaire provient précisément de l'apport du paramètre m. Au passage il convient de remarquer que ce paramètre m représente effectivement la moyenne de Z(x) sur un champ V, pourvu que celui-ci soit suffisamment grand par rapport à la structure spatiale. Expérimentalement, on prendra par exemple pour m la moyenne des échantillons prélevés régulièrement sur un tel champ.

Variances et krigeage

Dans le modèle stationnaire, toute combinaison linéaire de la forme:

$$\sum_i \lambda_i Z(x_i)$$

a une espérance:

$$E[\sum \lambda_i Z(x_i)] = [\sum \lambda_i] m$$

et admet une variance donnée par:

$$\sum_i \sum_j \lambda_i \lambda_j \, \mathrm{cov}[Z(x_i), Z(x_j)] = \sum_i \sum_j \lambda_i \lambda_j \, C(x_i - x_j)$$

Mathématiquement, la fonction C(h) doit être définie positive pour assurer la positivité de ces variances.

On en déduit en particulier la variance de blocs v:

$$\text{var } Z(v)=\text{cov}[Z(v),Z(v)]=\text{cov}[\frac{1}{v}\int_v Z(x)dx, \frac{1}{v}\int_v Z(y)dy]$$

$$=\frac{1}{v^2}\int_v\int_v C(x-y)dxdy=C(v,v)$$

ainsi que la covariance régularisée entre blocs v et v_h décalés de h:

$$\text{cov}[Z(v),Z(v_h)]=C(v,v_h)$$

Pour estimer la valeur d'un bloc (ou d'un point) dans le modèle stationnaire, on peut utiliser, lorsqu'elle est connue, la moyenne m, en plus des échantillons voisins:

$$Z(v)^*=\sum_i \lambda_i Z(x_i)+ \lambda_m\, m$$

et l'erreur moyenne:

$$E[Z(v)-Z(v)^*]=m-[\sum_i \lambda_i+\lambda_m]m$$

sera nulle si l'on donne à la moyenne le poids $\lambda_m=1-\sum_i \lambda_i$ assurant une somme des poids $\sum_i \lambda_i +\lambda_m$ égale à 100%.

Le meilleur estimateur linéaire (krigeage) sera alors donné par les poids λ_i qui minimisent la variance de l'erreur d'estimation:

$$C(v,v) - 2\sum_i \lambda_i C(x_i,v) + \sum_{i,j} \lambda_i\lambda_j C(x_i-x_j)$$

Il suffit alors d'annuler les dérivées partielles de cette variance vis à vis des λ_i pour obtenir le système linéaire dont les λ_i sont solutions:

$$\sum_j \lambda_j C(x_i-x_j) = C(x_i,v) \quad \text{pour tout i}$$

En pratique on observe que le poids complémentaire λ_m donné à la moyenne est d'autant plus fort que les échantillons sont rares dans le voisinage. Ainsi le krigeage va pallier à un manque local d'information en donnant plus de poids au paramètre m, ce qui est conforme à la signification stationnaire de ce paramètre.

Exercices

▶ *Exercice* 6.1. Anisotropie zonale.

On considère à 2D une FA stationnaire dont le variogramme est la somme de deux variogrammes à portée finie, l'un isotrope, l'autre ne dépendant que de la distance verticale, soit (en notations 2D):

$$\gamma(h_x, h_y)=\gamma_1(\sqrt{h_x^2 + h_y^2})+\gamma_2(h_y)$$

Montrer que la covariance C(h) peut également s'écrire comme somme de deux covariances C_1 et C_2 que l'on déterminera.

Quel est le coefficient de corrélation entre deux points situés à grande distance sur une même verticale? sur une même horizontale?

Réponse : 0 pour la verticale, $C_1(0)/[C_1(0)+C_2(0)]$ pour l'horizontale. Tout se passe comme si γ_2 avait une portée infinie en horizontal.

Variance de dispersion d'un point sur une horizontale? sur une ligne non horizontale?

Réponse : $C_1(0)$ pour une ligne horizontale, $C_1(0)+C_2(0)$ sinon.

▶ *Exercice* 6.2. Variance et covariance dans un champ V.

Z(x) est une FA stationnaire de moyenne m et covariance C(h), dont on considère la restriction à un champ V.

Montrer que la variance de dispersion d'un point dans V vaut:

$$\gamma(V,V)=C(0)-C(V,V)$$

et que la covariance entre deux points à distance h dans V est:

$$C(h)-C(V,V)$$

Noter que ce n'est que pour un champ "grand" par rapport à la structure (C(V,V) nul, Z(V)=m) que ces variance et covariance s'identifient aux variance C(0) et covariance C(h) a priori du modèle.

▶ *Exercice* 6.3. Covariance produit (ou factorisée).

Soit à 2D une covariance s'écrivant comme produit de deux covariances 1D:

$$C(h_x, h_y)=C_x(h_x)C_y(h_y)$$

Montrer que la structure selon la direction x n'est pas affectée par une régularisation selon y.

Réponse : La régularisation selon y ne joue que sur C_y: la covariance selon x reste donc, à un facteur près, égale à $C_x(h_x)$.

Une telle covariance peut-elle être isotrope?

Réponse : Oui, alors $C(h)=\exp(-(h/a)^2)$.

On considère la covariance exponentielle factorisée:

$$C(h)=\exp(-|h_x|/a_x)\exp(-|h_y|/a_y)$$

Montrer que cette covariance est exponentielle selon toute direction. On suppose $a_x=a_y=a$. Que devient le paramètre de portée dans la direction diagonale y=x?

Réponse : $a/\sqrt{2}$.

▶ *Exercice* 6.4. Krigeage de la moyenne. Moyenne connue et inconnue.

On considère une FA stationnaire, donc également intrinsèque, Z(x). Ecrire le système de krigeage intrinsèque d'un champ V à partir de points x_i. Lorsque V devient très grand, Z(V) tend vers le paramètre de moyenne m, et le krigeage de Z(V) tend vers le krigeage de la moyenne (estimateur linéaire optimal de la moyenne) m^K. Ecrire le système de krigeage de la moyenne.

On considère maintenant le krigeage d'un point x_0 de deux façons: soit le krigeage stationnaire faisant appel à la moyenne (appelé "krigeage à moyenne connue" ou "krigeage simple"); soit le krigeage intrinsèque ("krigeage à moyenne inconnue" ou "krigeage ordinaire"). Montrer que le krigeage ordinaire:

$$\sum \lambda_i^{KO} Z(x_i)$$

n'est autre qu'un krigeage simple:

$$\sum \lambda_i^{KS} Z(x_i) + \lambda_m m$$

dans lequel la moyenne m est remplacée par son krigeage.

Réponse : On vérifie que les poids obtenus par ce dernier procédé sont bien la solution, supposée unique, du système de krigeage ordinaire.

Remarque: on a une relation correspondante entre les variances de krigeage des krigeages ordinaire, simple, et de la moyenne:

$$\sigma_{KO}^2=\sigma_{KS}^2+\lambda_m^2\,\sigma_m^2$$

▶ *Exercice* 6.5. Krigeages très simples.

Z(x) est une FA stationnaire de moyenne m et covariance C(h). Krigeage et variance de krigeage de $Z(x_0)$ par $Z(x_1)$?

Réponse : $Z(x_0)^K = \lambda Z(x_1)+(1-\lambda)m$ avec $\lambda=C(x_0-x_1)/C(0)$, variance $1-\lambda^2$.

Krigeage de Z(v) par Z(x)?

Réponse : Idem avec $\lambda=C(x,v)/C(0)$.

Krigeage de Z(x) par Z(v)?

Réponse : Idem avec $\lambda=C(x,v)/C(v,v)$.

Krigeage de Z(x) par Z(v), avec x aléatoire dans v?

Réponse : $\lambda=C(v,v)/C(v,v)=1$, donc $Z(x)^K=Z(v)$.

▶ *Exercice* 6.6. Krigeage d'un point connaissant deux autres points.

Z(x) est une FA stationnaire de moyenne m et covariance C(h) à 1D. On veut kriger la valeur Z_0 en un point x_0, à partir des valeurs Z_1 et Z_2 en deux points $x_1=x_0+h_1$ et $x_2=x_0+h_2$ situés du même côté de x_0 ($h_1>0$ et $h_2-h_1=h>0$). Poids de krigeage?

Réponse : $$\lambda_1=\frac{1}{2}\left[\frac{C(h_1)+C(h_2)}{C(0)+C(h)}+\frac{C(h_1)-C(h_2)}{C(0)-C(h)}\right]$$

et $$\lambda_2=\frac{1}{2}\left[\frac{C(h_1)+C(h_2)}{C(0)+C(h)}-\frac{C(h_1)-C(h_2)}{C(0)-C(h)}\right].$$

Vérifier que, pour une covariance C(h) décroissante en $|h|$, le poids de x_1 est supérieur à celui de x_2.

Montrer que, dans le cas d'une covariance exponentielle, le poids de x_2 est nul: le point x_1 fait écran à x_2.

Poids de la moyenne? Remarquer que ce poids est nécessairement positif pour une covariance décroissante.

Réponse : $$\lambda_m=1-\frac{C(h_1)+C(h_2)}{C(0)+C(h)}.$$

7 MODELES NON STATIONNAIRES

Erreurs d'ajustement

On s'intéresse ici à des phénomènes qui, à l'échelle du champ de travail, manifestent une dérive: exemple déjà mentionné de la profondeur du fond marin en s'éloignant des côtes; ou encore profondeur d'un horizon géologique imperméable en forme de dôme servant de toit à un réservoir de pétrole.

Si un tel horizon est très régulier, faire passer une courbe mathématique régulière (polynôme de faible degré par rapport aux coordonnées, fonction spline) par les points de données (les quelques valeurs de la profondeur connues par puits) peut suffire à en donner une fidèle représentation.

Si ce n'est pas possible, une solution consiste à ajuster celle-ci de façon à passer au plus près des points de données. On choisira par exemple, technique des moindres carrés, les coefficients du polynôme qui minimisent le carré des erreurs (différences entre valeurs mesurées et valeurs prédites par la courbe) aux points de données. La courbe passera alors d'autant plus près des points de données que ce carré des erreurs sera petit. Ce carré des erreurs peut donc être utilisé comme critère d'ajustement. Cependant, par le fait même qu'il est minimisé aux points de données, ce critère surestime la qualité de l'estimation en dehors de ces points.

Par ailleurs l'ajustement par une courbe bien lisse, et l'existence d'un écart entre valeurs ajustées et valeurs mesurées aux points de données, peut facilement laisser croire que les valeurs mesurées sont entachées d'erreurs, et que l'ajustement est plus proche de la réalité que ne le sont ces mesures. Ce qui revient, paradoxalement, à donner priorité au modèle ajusté par rapport aux données. Mais supposons qu'effectivement le phénomène lui-même obéisse à une courbe simple $m(x)$, et que les mesures $z(x_i)$ aux points de données x_i soient entachées d'erreurs e_i indépendantes et de même loi:

$$z(x_i)=m(x_i)+e_i$$

Il faut alors noter, à propos de l'ajustement $m(x)^*$ qui est fait de $m(x)$, que les erreurs quadratiques:

$$[z(x_i)-m(x_i)^*]^2$$

justement parce que minimisées aux points de données, ont tendance à sous-estimer les erreurs quadratiques vraies des mesures:

$$[z(x_i)-m(x_i)]^2 = e_i^2$$

La dérive du modèle

A moins qu'elle ne soit guidée par une raison physique, la distinction, sur les valeurs mesurées du phénomène, entre une partie régulière m(x) et une autre sans structure spatiale, peut sembler simpliste. Que se passe-t-il si l'on fait l'hypothèse, plus large, que le phénomène peut être représenté par la somme Z(x) d'une partie régulière m(x), la dérive, et d'un résidu Y(x) présentant lui-même une structure spatiale:

$$Z(x)=m(x)+Y(x)?$$

Dans un modèle de FA, il sera commode de considérer m(x) comme l'espérance, non stationnaire puisque dépendant de x, de Z(x):

$$E[Z(x)]=m(x)$$

et, dans les cas les plus simples, le résidu, de moyenne nulle:

$$E[Y(x)]=E[Z(x)]-m(x)=0$$

sera considéré comme stationnaire avec une certaine covariance C(h).

Dans certains cas, la dérive sera définie sans ambiguïté. Ce pourra être, par définition, la partie invariante dans le temps pour un phénomène se répétant dans le temps. Ou bien la composante à grande distance, connue, pour un phénomène présentant par ailleurs des fluctuations à courte distance (la notion de distance dépendant naturellement de l'échelle de travail). Dans ces cas-là, il suffit de soustraire la dérive à Z(x), pour se ramener au cas stationnaire avec $Y(x)=Z(x)-m(x)$.

Supposons maintenant que la dérive m(x) ne soit pas connue, mais admettons qu'elle puisse s'exprimer sous la forme d'un polynôme, de degré k donné, des coordonnées:

$$m(x)= \sum a_l f^l(x)$$

où les $f^l(x)$ sont les monômes de degré inférieur ou égal à k (1, x, y, x^2, y^2, xy à 2 dimensions pour k=2), et les a_l des coefficients inconnus. L'erreur d'estimation de $Z(x_0)$ par une combinaison linéaire $\sum \lambda_i Z(x_i)$ a comme espérance:

$$E[Z(x_0)- \sum \lambda_i Z(x_i)]$$

Celle–ci sera nulle (non biais), quels que soient les a_l, moyennant les conditions:

$$\sum_i \lambda_i f^l(x_i) = f^l(x_0) \qquad \text{pour tout } l$$

La variance de l'erreur, qui s'écrit:

$$C(x_0 - x_0) - 2\sum_i \lambda_i C(x_i - x_0) + \sum_{i,j} \lambda_i \lambda_j C(x_i - x_j)$$

ne dépend pas des a_l. L'estimateur optimal de Z(x), appelé Krigeage Universel, est alors obtenu en minimisant cette variance sous les conditions garantissant le non biais. Les poids sont solutions du système de krigeage (où les μ_l sont des paramètres de Lagrange):

$$\begin{cases} \sum_j \lambda_j C(x_i - x_j) - \sum_l \mu_l f^l(x_i) = C(x_i - x_0) & \text{pour tout } i \\ \sum_i \lambda_i f^l(x_i) = f^l(x_0) & \text{pour tout } l \end{cases}$$

On peut, de façon analogue, procéder à l'estimation optimale $m(x)^*$ de la dérive m(x) par une combinaison linéaire des $Z(x_i)$.

Cette méthode, toutefois, repose sur la connaissance de la covariance des résidus C(h). Or, pour les mêmes raisons que pour l'ajustement polynomial, les résidus estimés aux points de données $Z(x_i) - m(x_i)^*$ ont tendance à être plus proches de zéro que les "vrais" résidus inconnus $Z(x_i) - m(x_i)$, et à présenter une structure spatiale moins marquée que ceux–ci (pouvant d'ailleurs laisser croire à tort à une absence de structure spatiale). Nous voilà ainsi dans un cercle vicieux où, de la variable Z(x), nous aurons du mal à faire la distinction, en pratique, entre la part qui devrait revenir à la dérive, et celle du résidu.

Les FAIk

Un moyen de contourner la difficulté est d'utiliser un modèle de fonction aléatoire intrinsèque d'ordre k (en abrégé FAIk). Reprenons tout d'abord l'exemple d'une fonction aléatoire Z(x), somme d'une dérive polynomiale m(x) et d'un résidu Y(x) stationnaire. Si m(x) est constant, égal à m, Z(x) est stationnaire. Mais elle est également, comme on l'a vu au chapitre précédent, intrinsèque. Simplement, le paramètre de moyenne m disparaît des accroissements $Z(x+h) - Z(x)$ ou des CLA:

$$\sum \lambda_i Z(x_i) \text{ avec } \sum \lambda_i = 0$$

que manipule le modèle intrinsèque.

De la même façon, si la dérive m(x) est linéaire, par exemple ax+b à 1D, elle est "filtrée" si l'on considère des accroissements d'accroissements:

$$[Z(x+2h)-Z(x+h)]-[Z(x+h)-Z(x)]=Z(x+2h)-2Z(x+h)+Z(x)$$

(car $a(x+2h)-2a(x+h)+ax=0$), ou toute CL:

$$\sum \lambda_i Z(x_i) \text{ telle que: } \sum \lambda_i = 0 \text{ et } \sum \lambda_i x_i = 0$$

Plus généralement une dérive polynomiale d'ordre k sera filtrée en considérant les accroissements généralisés d'ordre k:

$$\sum \lambda_i Z(x_i) \text{ avec } \sum_i \lambda_i f^l(x_i) = 0$$

où les $f^l(x)$ sont les monômes de degré inférieur ou égal à k.

On appelle alors FAIk une fonction aléatoire dont les accroissements généralisés d'ordre k (combinaisons linéaires autorisées à l'ordre k) sont stationnaires et de moyenne nulle. La variance de ces accroissements s'exprime à l'aide d'un nouvel outil structural, la covariance généralisée K(h) (définie à un polynôme pair de degré 2k près):

$$\text{var}[\sum \lambda_i Z(x_i)] = \sum_i \sum_j \lambda_i \lambda_j K(x_i - x_j)$$

Le cas k=0 correspond au cas intrinsèque déjà rencontré, avec K(h) égal à $-\gamma(h)$ (à une constante additive près, disparaissant de toute façon des variances). De même que la famille des variogrammes γ est plus vaste que celle des covariances stationnaires C(h), de même celle des K(h) augmente avec k (fonctions k−conditionnellement définies positives).

Pour k>0, la manipulation d'accroissements généralisés rend l'analyse structurale (choix du degré k et ajustement d'un modèle de covariance généralisée) beaucoup moins directe et plus délicate que dans le cas du variogramme (voir exercice 7.5). On utilise souvent pour K(h) des modèles polynomiaux basés sur $-|h|$ (pour $k \geq 0$), $|h|^3$ (pour $k \geq 1$), $-|h|^5$ (pour $k \geq 2$)... qui se prêtent bien à un ajustement automatique. Notons également le modèle $h^2 \text{Log}|h|$ qui, pour une

FAI1 à 2 dimensions, conduit à une interpolation par krigeage identique à une spline plaque mince.

Le krigeage en FAIk

Une fois le degré k et la covariance généralisée K(h) choisis, le krigeage, estimateur linéaire optimal de Z(x) par les $Z(x_i)$, s'obtient en écrivant que l'erreur:

$$Z(x) - \sum \lambda_i Z(x_i)$$

est une CLA:

$$\sum_i \lambda_i f^l(x_i) = f^l(x_0) \qquad \text{pour tout l}$$

et en en minimisant la variance:

$$K(x_0 - x_0) - 2\sum_i \lambda_i K(x_i - x_0) + \sum_{i,j} \lambda_i \lambda_j K(x_i - x_j)$$

D'où le système de krigeage:

$$\begin{cases} \sum_j \lambda_j K(x_i - x_j) - \sum_l \mu_l f^l(x_i) = K(x_i - x_0) & \text{pour tout i} \\ \sum_i \lambda_i f^l(x_i) = f^l(x_0) & \text{pour tout l} \end{cases}$$

dont on notera la ressemblance avec celui du krigeage universel. Dans une FAIk, les CLA, donc les poids et variance de krigeage, ne changent pas si l'on rajoute ou retranche un polynôme de degré k: ils filtrent en particulier toute dérive de cette forme, qui n'apparait plus explicitement dans le modèle.

On peut ainsi réaliser une cartographie du toit d'un réservoir pétrolier à partir des profondeurs de ce toit connues à l'emplacement des puits. L'allure en dôme que pourrait présenter la dérive se trouve implicitement prise en compte par le filtrage des polynômes de degré 2. Si les données puits sont peu nombreuses, la cartographie sera faite en tout point à l'aide de l'ensemble des puits ("voisinage unique"), et le dôme correspond à l'allure générale. Si les données sont au contraire nombreuses, la cartographie sera faite en sélectionnant les puits présents dans un voisinage autour de chaque point estimé ("voisinage glissant"), sous l'hypothèse que, localement, la dérive est par exemple un polynôme de degré 2.

Exercices

▶ *Exercice* 7.1. Modèle additif.

Soit une fonction aléatoire définie par $Z(x)=m(x)+Y(x)$, avec $Y(x)$ stationnaire de moyenne 1 et covariance $C(h)$. Que valent la dérive $E[Z(x)]$, la variance de $Z(x)$, et les covariances centrée $cov[Z(x),Z(x+h)]$ et non centrée $E[Z(x)Z(x+h)]$? Ces covariances sont-elles stationnaires?

Montrer que si $m(x)$ est un polynôme de degré k, $Z(x)$ est une FAIk.

▶ *Exercice* 7.2. Modèle multiplicatif.

Soit une fonction aléatoire définie par $Z(x)=m(x)Y(x)$, avec $Y(x)$ stationnaire de moyenne 1 et covariance $C(h)$. Alors $m(x)=E[Z(x)]$ représente la dérive de $Z(x)$. Que valent la variance de $Z(x)$ et la covariance $cov[Z(x),Z(x+h)]$? Sont-elles stationnaires?

On suppose la dérive linéaire. Montrer que les CLA à l'ordre 1 de $Z(x)$ ont bien une espérance nulle, mais que leur variance n'est pas stationnaire. En déduire que $Z(x)$ N'EST PAS une FAI1.

▶ *Exercice* 7.3. Nombre minimum de points dans une CLA2 à 2D.

Quel est le nombre minimum de points nécessaires pour construire une CLA2 à 2D? Que devient ce minimum pour des points situés sur une courbe quadratique ou s'ils sont alignés?

> *Réponse :* Les coefficients de la combinaison linéaire doivent vérifier un système linéaire de 6 équations à seconds membres nuls, correspondant à 1, x, y, x^2, y^2, xy: il faut donc en général au minimum 7 points pour avoir des coefficients non tous nuls. Cependant si les points appartiennent à une courbe quadratique, les équations se réduisent à 5 d'entre elles, et le minimum de points à 6. Si les points sont alignés, on est ramené à une dimension: le nombre d'équations tombe à 3, d'où un minimum de 4 points.

▶ *Exercice* 7.4. Configuration spéciale.

$Z(x)$ est une FAI2 à 2D. On considère un point O, que l'on pourra prendre comme origine, et n points A_1, ..., A_n, situés sur un cercle de rayon a. Montrer qu'une combinaison linéaire de ces points, donnant le poids 1 à O, ne peut être autorisée à l'ordre deux. En déduire que l'estimation au point O à partir des seuls points A_1, ..., A_n est impossible dans ce modèle.

Réponse : Soit λ_i le poids attribué au point A_i. Pour être autorisée, la combinaison linéaire devrait vérifier, entre autres:

$$1+\sum \lambda_i=0$$

$$\sum \lambda_i x_i^2=0$$

$$\sum \lambda_i y_i^2=0$$

Or $\sum \lambda_i x_i^2+\sum \lambda_i y_i^2=\sum \lambda_i (x_i^2+y_i^2)=\sum \lambda_i a^2=a^2$ ne peut être nul.

L'erreur d'estimation en O ne peut donc filtrer les polynômes d'ordre 2.

▶ *Exercice* 7.5. Construction d'accroissements généralisés d'ordre k.

Montrer que les erreurs, aux points de données, de l'ajustement par moindres carrés d'un polynôme de degré k sont des accroissements généralisés d'ordre k. En pratique ce résultat permet de construire des accroissements généralisés d'ordre k, sur lesquels ajuster une covariance généralisée.

▶ *Exercice* 7.6. Dérive externe.

Pour cartographier le toit d'un réservoir pétrolier, on dispose souvent, en plus des valeurs de la profondeur Z(x) aux puits, d'une couverture sismique. Cette couverture est susceptible de fournir une variable auxiliaire f(x), par exemple un temps sismique, bien connue sur le champ, et qui peut avantageusement jouer le rôle de facteur de forme pour la profondeur. On admet par exemple que la dérive de la profondeur Z(x) est de la forme $a_0+a_1f(x)$, avec des coefficients ai inconnus.

On veut estimer la profondeur en un point x_0 à partir des profondeurs aux puits $Z(x_i)$. Quelles conditions imposer aux poids de cette combinaison pour assurer le non biais?

Réponse : $\sum \lambda_i=1$ et $\sum \lambda_i f(x_i)=f(x_0)$.

▶ *Exercice* 7.7. Filtrage d'une perturbation périodique.

Lors d'une campagne de mesures, on visite des sites x_1, x_2, x_3... à des instants respectifs t_1, t_2, t_3... La variable mesurée est la somme d'une variable cible et d'une perturbation temporelle sinusoïdale de fréquence connue u et de moyenne nulle:

$$W(x,t)=Z(x)+R(t)$$

Pour cartographier la variable cible, on considère des combinaisons linéaires des valeurs mesurées aux points x_i. Quelles conditions doivent vérifier les poids de façon à filtrer la perturbation? Cas de plusieurs fréquences?

Réponse : $\sum \lambda_i \cos(ut_i)=0$ et $\sum \lambda_i \sin(ut_i)=0$ pour chaque fréquence u.

8 GEOSTATISTIQUE MULTIVARIABLE

Le cokrigeage

L'étude simultanée de plusieurs variables régionalisées pose très souvent le problème de leur relation structurale. Supposons par exemple que l'on cherche à cartographier par géostatistique les profondeurs de deux horizons géologiques Z_1 et Z_2, à partir des valeurs connues en certains points. Vaut-il mieux les kriger séparément? ou est-il préférable de cartographier séparément l'une d'entre elle et la puissance $P=Z_2-Z_1$ de la couche intermédiaire? Tout dépend de la relation structurale existant entre ces variables. Dans certains cas le comportement de chacun des horizons pourra être indépendant de l'autre, justifiant ainsi leur krigeage séparé. Dans d'autres ce pourra être au contraire la puissance Z_2-Z_1 qui sera indépendante de l'une d'entre elles, justifiant alors le traitement séparé de ces 2 variables.

Mais, plus généralement, on pourra choisir d'estimer chacune des profondeurs par une combinaison linéaire de ces deux profondeurs aux points de données, soit par exemple:

$$Z_1(x)^* = \sum_i \lambda_i^1 Z_1(x_i) + \sum_j \lambda_j^2 Z_2(x_j)$$

Du fait que $Z_2=Z_1+P$, l'estimation linéaire optimale correspond également à la meilleure combinaison linéaire des $Z_1(x_i)$ et $P(x_i)$. Ainsi le choix des variables n'importe plus. Avec cet estimateur appelé cokrigeage, on assure toujours la cohérence:

$$Z_2(x)^{CK} = Z_1(x)^{CK} + P(x)^{CK}$$

Les outils structuraux croisés

La connaissance, dans le cas intrinsèque par exemple, des variogrammes simples γ_1 et γ_2 de Z_1 et Z_2 n'est alors plus suffisante. Il est nécessaire de connaître en sus le variogramme croisé entre les variables:

$$\gamma_{12}(h) = (1/2)E[(Z_1(x+h)-Z_1(x))(Z_2(x+h)-Z_2(x))]$$

d'ailleurs égal à $\gamma_{12}(-h)$.

Dans le cas où Z_1 et Z_2 seraient indépendants, ce variogramme croisé serait identiquement nul. Si par contre $P=Z_2-Z_1$ était indépendant

de Z_1, alors le variogramme croisé entre Z_1 et Z_2 (somme de celui entre Z_1 et elle-même d'une part, et de celui, nul, entre Z_1 et P d'autre part) serait égal au variogramme de Z_1.

A la différence des variogrammes simples, le variogramme croisé entre deux variables peut prendre des valeurs négatives, une augmentation de l'une des variables allant de pair avec une diminution de l'autre. S'il s'agit de concentrations en éléments différents, ceci peut traduire un mécanisme de substitution d'un élément par l'autre.

La connaissance des variogrammes simples et croisé permet de calculer la variance de toute combinaison linéaire autorisée ($\sum \lambda_i^1=0$, $\sum \lambda_j^2=0$):

$$\mathrm{var}[\sum_{i\in I} \lambda_i^1 Z_1(x_i) + \sum_{j\in J} \lambda_j^2 Z_2(x_j)] =$$

$$-\sum_{i,k\in I} \lambda_i^1\ \lambda_k^1\ \gamma_1(x_i-x_k) - \sum_{j,l\in J} \lambda_j^2\ \lambda_l^2\ \gamma_2(x_j-x_l) - 2\sum_{i\in I}\sum_{j\in J} \lambda_i^1\ \lambda_j^2\ \gamma_{12}(x_i-x_j)$$

En particulier le cokrigeage s'obtient en minimisant la variance de l'erreur d'estimation $Z_1(x)-Z_1(x)^*$, sous réserve que celle-ci est bien une combinaison linéaire autorisée, d'où les conditions:

$$\sum \lambda_i^1=1 \text{ et } \sum \lambda_j^2=0$$

A remarquer que les deux variables, ou plus, peuvent (cas isotopique) ou non (cas hétérotopique) être connues aux mêmes points. Le cokrigeage est d'ailleurs un moyen d'améliorer l'estimation d'une variable par l'apport d'une autre variable mieux échantillonnée.

Dans le cas stationnaire où les moyennes (m_1 et m_2) sont connues, l'estimateur de cokrigeage s'appuie sur ces moyennes:

$$Z_1(x)^{CK} = \sum_i \lambda_i^1 Z_1(x_i) + [1-\sum \lambda_i^1]m_1 + \sum_j \lambda_j^2 [Z_2(x_j)-m_2]$$

et, aux covariances simples C_1 et C_2 s'ajoute maintenant la covariance croisée:

$$C_{12}(h)=\mathrm{cov}[Z_1(x), Z_2(x+h)]=E[(Z_1(x)-m_1)(Z_2(x+h)-m_2)]$$

On notera que pour une distance vectorielle h, cette covariance croisée $C_{12}(h)$ (égale à $C_{21}(-h)$) n'est pas nécessairement égale à $C_{12}(-h)$. Ainsi un décalage dans l'espace entre les deux variables peut

se manifester par un maximum de corrélation à une distance h non nulle. En ce sens la covariance croisée est un outil plus riche que le variogramme croisé, alors égal à:

$$\gamma_{12}(h)=C_{12}(0)-(1/2)[C_{12}(h)+C_{12}(-h)]$$

et qui ne peut faire la distinction entre h et $-h$.

Cette covariance croisée mesure, aux écarts-types près, la corrélation spatiale entre les deux variables, c'est-à-dire non seulement le coefficient de corrélation habituel en un même point (h=0):

$$\varrho_{12}=C_{12}(0)/\sigma_1\,\sigma_2$$

mais également celui entre points à distance h.

Par ailleurs le coefficient de corrélation habituel n'est en général pas identique selon que l'on considère les deux variables mesurées ponctuellement, ou bien les mêmes variables mesurées sur un support plus grand. Ce n'est que dans le modèle de "corrélation intrinsèque" qu'un tel coefficient est indépendant du support, comme d'ailleurs du champ. Ce modèle se caractérise par des variogrammes (ou covariances) simples et croisés proportionnels les uns aux autres. Ainsi non seulement les structures simples des variables sont (à un facteur près) identiques, mais leurs structures croisées aussi sont identiques à ces structures simples. Dans ce modèle très particulier, on montre que le cokrigeage s'identifie au krigeage, pourvu que les variables soient connues aux mêmes points de données (cas isotopique).

Modèles de corégionalisation

L'ajustement des variogrammes/covariances simples et croisés doit être mathématiquement cohérent, de façon à garantir des variances positives. Les modèles linéaires de corégionalisation sont les plus couramment employés.

De tels modèles sont en fait déjà utilisés dans le cas monovariable. Considérons en effet une variable Z stationnaire, dont la structure serait somme de deux schémas sphériques de paliers 9 et 16 et portées 20 et 100:

$$9\ \mathrm{Sph}(20) + 16\ \mathrm{Sph}(100)$$

Une telle variable peut s'écrire comme une combinaison:

$$3Y_1 + 4Y_2$$

de deux composantes Y_1 et Y_2 sans corrélation spatiale, de variance unité et de structure sphérique de portées respectives 20 et 100.

Prenons maintenant deux variables Z_1 et Z_2 de structures:

$$C_1 = 25\ \mathrm{Sph}(20) + 100\ \mathrm{Sph}(100)$$

$$C_2 = 100\ \mathrm{Sph}(20) + 100\ \mathrm{Sph}(100)$$

$$C_{12} = -24\ \mathrm{Sph}(20) + 60\ \mathrm{Sph}(100)$$

Ce modèle correspond, par exemple, à la décomposition:

$$Z_1 = 3Y_1 + 4Y'_1 + 6Y_2 + 8Y'_2$$

$$Z_2 = -8Y_1 + 6Y''_1 + 10Y_2$$

où Y_1, Y'_1, Y''_1, Y_2, Y'_2 sont des composantes sans corrélation spatiale, de variance 1, et de structures sphériques de portées respectives 20, 20, 20, 100, 100.

Il s'agit donc bien d'un modèle cohérent, qui garantira des variances positives.

Ainsi, dans un modèle linéaire de corégionalisation, les variables se décomposent linéairement en composantes élémentaires, et donc leurs structures (variogramme ou covariance) simples et croisées sont des combinaisons linéaires de schémas élémentaires. Les composantes propres à chacune des variables disparaissent des structures croisées. Inversement, un schéma élémentaire donné (par exemple sphérique de portée donnée) ne peut apparaître dans la structure croisée entre deux variables s'il n'est présent dans leurs structures simples.

Un fois ajusté, un tel modèle permet le cokrigeage des variables. Mais il permet aussi de cokriger une des composantes (par exemple Y_1) à partir des données sur Z_1 et Z_2, méthode connue sous le nom d'analyse krigeante. La décomposition en composantes élémentaires n'est en général cependant pas unique, et le choix de l'une d'entre elles devient alors conventionnel. Ainsi, dans l'exemple précédent, le même modèle structural pour Z_1 et Z_2 peut être obtenu par la décomposition différente:

$$Z_1 = 4Y_1 + 3Y'_1 + 10Y_2$$

$$Z_2 = -6Y_1 + 8Y''_1 + 6Y_2 + 8Y'_2$$

Le modèle linéaire précédent, éventuellement étendu au cas non stationnaire, n'est pas le seul moyen de décrire une corégionalisation

en vue d'un cokrigeage. Par exemple, la connaissance du gradient d'un horizon géologique peut être utilisée pour améliorer la cartographie faite à partir des seules données de profondeur. Dans ce cas les composantes du gradient ont des structures simples, et des structures croisées avec la profondeur, qui dérivent de la structure de la profondeur. Voir exercice 8.2.

Régressions entre variables

Lorsqu'une variable, mettons T(x), exerce une influence sur une autre, Q(x), il peut être avantageux de considérer la régression de cette dernière à partir de l'autre, prise au même point:

$$E[Q(x)|T(x)]$$

soit $aT(x)+b$, s'il s'agit d'une régression linéaire.

Le résidu:

$$R(x)=Q(x)-aT(x)-b$$

est alors, par construction, sans corrélation avec T(x). L'écriture:

$$Q(x)=aT(x)+b+R(x)$$

ne correspond cependant à une décomposition intéressante spatialement que si R(x) n'est pas corrélé non plus à T(y) en tout autre point y (pourquoi sinon ne pas partir de la régression de T sachant Q?). La covariance (ou variogramme) croisée entre T et Q est alors égale, à un facteur près, à la structure simple de T. Et le cokrigeage, dans le cas isotopique, peut s'obtenir par krigeage séparé des variables sans corrélation spatiale T et R.

L'exemple suivant, emprunté à l'industrie minière, illustre ce propos, en même temps d'ailleurs qu'il offre un aperçu sur des domaines plus vastes que ceux des statistiques linéaires. Il s'agit d'une minéralisation dont la géométrie se répartit elle−même de manière très variable à l'intérieur d'un domaine donné. Cette géométrie est alors représentée par une variable régionalisée T(x) (cette variable, définie ponctuellement à 3D, vaudra 1 ou 0 selon que le point appartient ou non à la minéralisation; définie à 2D par intégration sur la verticale, elle est alors égale à la puissance de la minéralisation, comme disent les mineurs).

Par ailleurs la teneur Z(x) permet de passer de la variable géométrique T(x) à l'accumulation de métal $Q(x)=Z(x)T(x)$. Noter

que dans ce cas, la teneur Z(x) n'est pas additive (il faut la pondérer par T(x) pour en faire une moyenne), mais que les variables T(x) et Q(x), souvent alors appelées variables utiles, le sont.

Sous l'hypothèse que la distribution spatiale des teneurs à l'intérieur de la minéralisation est indépendante de la géométrie de celle-ci, on a effectivement la décomposition, valable spatialement, du métal en deux termes sans corrélation spatiale:

$$Q(x)=E[Q(x)|T(x)]+[Q(x)-E[Q(x)|T(x)]]$$

$$=mT(x)+R(x)$$

où m est la teneur moyenne. (L'hypothèse d'indépendance peut d'ailleurs être allégée et réduite à une indépendance en moyenne $E[Z(x)|T(x),T(y)]=m$.)

On remarquera que le résidu $R(x)=Q(x)-mT(x)=(Z(x)-m)T(x)$ dépend de T(x), bien qu'il ne lui soit pas corrélé. Et malgré l'utilité de la décomposition linéaire (cokrigeage par exemple), il est beaucoup plus parlant, d'un point de vue physique, de se référer au modèle multiplicatif:

$$Q(x)=Z(x)T(x)$$

et à la relation d'indépendance existant entre Z et T.

Dans le même esprit, mais provenant cette fois du domaine halieutique, voici un modèle un peu plus compliqué. Il s'agit de la régionalisation d'une densité en poisson en relation avec un ou plusieurs paramètres hydrographiques T(x) (température, salinité...). La préférence, par le poisson, de certaines conditions conduit à une régression $E[Z(x)|T(x)]$ non linéaire (par exemple une courbe en cloche sur la gamme préférée). Si, de plus, à paramètres fixés, l'écart-type de Z(x) est proportionnel à cette régression, on peut penser à un modèle multiplicatif:

$$Z(x)=E[Z(x)|T(x)]Y(x)$$

avec une variable indépendante Y(x) (par exemple stationnaire de moyenne 1). Celle-ci décrit alors la structure du poisson non expliquée par les paramètres hydrographiques.

Exercices

▶ *Exercice* 8.1. Analyse krigeante pour une variable.

La structure C(h) d'une FA stationnaire Z(x) de moyenne m se décompose en une structure pépitique $C_0(h)$, une structure $C_1(h)$ de portée 1, et une structure $C_2(h)$ de portée 2. La FA elle-même s'exprime alors comme somme de 3 composantes sans corrélation spatiale obéissant respectivement à ces structures:

$$Z=Y_0+Y_1+Y_2$$

Montrer que les moyennes m_0, m_1 et m_2 de ces composantes sont à priori indéterminées, en dehors de la relation nécessaire:

$$m=m_0+m_1+m_2$$

Système de krigeage de chaque composante en x_0 à partir de $Z(x_1)$, $Z(x_2)$, ..., $Z(x_n)$?

Réponse : C'est le système de krigeage de $Z(x_0)$, dans lequel le second membre $C(x_0-x_i)$ est remplacé par $C_0(x_0-x_i)$, $C_1(x_0-x_i)$ et $C_2(x_0-x_i)$ respectivement.

Montrer que dans la cartographie d'une composante, seules sont significatives les variations.

Réponse : Ceci est dû au fait que la moyenne est indéterminée.

On suppose que Y_0 correspond à des erreurs de mesure de moyenne nulle. A quoi correspond le krigeage de Y_1+Y_2?

Réponse : C'est un krigeage de Z avec filtrage des erreurs, voir aussi exercice 5.11.

▶ *Exercice* 8.2. Cokrigeage d'une variable et de sa dérivée.

Z(x) est une FAI0 à 1D, dérivable (en moyenne quadratique) et de variogramme $\gamma(h)$.

Montrer que $E[Z'(x)]=0$, $E[Z(x+h)-Z(x), Z'(x)]=\gamma'(h)$ et $E[Z'(x)Z'(x+h)]=\gamma''(h)$, et en déduire que Z'(x) est une FA stationnaire.

On connaît Z et Z' en un point x_1, et on veut estimer, par cokrigeage, la valeur de Z en x_0. Poids de cokrigeage? Comparer avec le résultat de l'exercice 5.10.

Réponse : 1 et $-\gamma'(h_1)/\gamma''(0)$, avec $h_1=x_1-x_0$.

On veut estimer cette fois $Z'(x_0)$. Poids de cokrigeage de $Z(x_1)$ et $Z'(x_1)$?

Réponse : 0 et $\gamma''(h_1)/\gamma''(0)$.

9 GEOSTATISTIQUE NON-LINEAIRE

Dépassement de seuil et loi conditionnelle gaussienne

Les méthodes non-linéaires permettent de résoudre un certain nombre de problèmes spécifiques que l'on rencontrera au cours du chapitre. Parmi ceux-ci figure le problème important des dépassements de seuil. Le fait qu'en un point x_0, la valeur estimée d'une variable soit égale à 0.6 ne nous garantit en effet pas que la vraie valeur inconnue $Y(x_0)$ ne dépasse pas le seuil de 2 par exemple. Ce seuil peut représenter une côte d'alerte pour une concentration en polluant, la teneur de coupure dans une exploitation minière sélective, etc. Dans de tels problèmes de dépassement de seuil, ce que l'on cherche, ce sera plutôt la probabilité pour que la valeur vraie inconnue de Y en un point x_0 dépasse la coupure, compte tenu des informations du voisinage (probabilité dite conditionnelle).

Une telle probabilité, inaccessible à partir des seules hypothèses de la géostatistique linéaire, se calcule au contraire très bien dans un modèle multi-gaussien. On rappelle qu'une variable suit une distribution "normale", ou gaussienne, si sa densité de probabilité a la forme de la célèbre courbe en cloche d'équation:

$$g_{\mu,\sigma^2}(t)=\frac{1}{\sqrt{2\pi}\,\sigma}\ \exp-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2$$

μ et σ^2 étant respectivement la moyenne et la variance de la variable (Fig. 16).

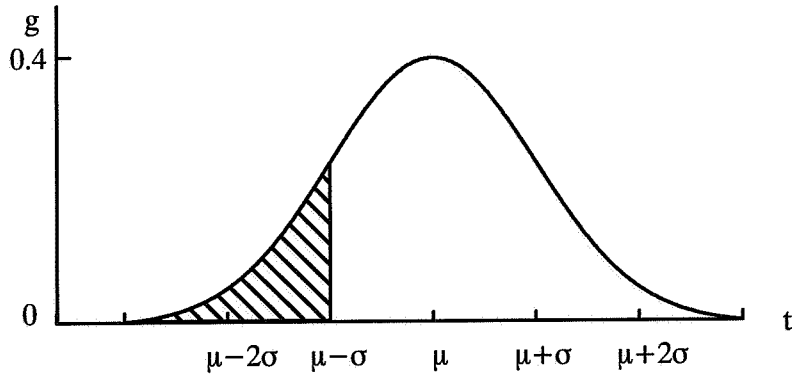

Fig. 16: Densité de probabilité gaussienne. L'aire hachurée donne la probabilité (0.16) d'être inférieur à $t=\mu-\sigma$.

Une FA est alors multi−gaussienne si, pour tout ensemble de points $x_0, x_1, ..., x_n$, toute combinaison linéaire de $Y(x_0), Y(x_1), ... Y(x_n)$, est gaussienne. A fortiori $Y(x_0)$ en un point donné x_0 suit−elle une loi gaussienne. Dans ce modèle, et c'est ce qui fait son intérêt ici, la distribution de $Y(x_0)$ conditionnellement aux valeurs prises en des points $x_1, x_2, ..., x_n$ prend une expression sympathique: c'est encore une distribution gaussienne, de moyenne la valeur krigée de $Y(x_0)$, et de variance la variance de krigeage. La probabilité pour que la valeur inconnue de $Y(x_0)$ dépasse un seuil y donné s'obtient alors en sommant la densité de probabilité à partir de y:

$$\int_y g_{y_0^K, \sigma_K^2}(t)\ dt$$

Le cas lognormal

Pour la suite de ce chapitre nous nous placerons exclusivement dans le cas d'une FA Z(x) stationnaire (de fait la géostatistique non−linéaire n'est pratiquement pas développée hors du cadre stationnaire). Alors la loi de probabilité de Z est la même pour tout point x, de sorte que l'histogramme des données en est une version expérimentale. Or en pratique, l'histogramme d'une concentration ou d'une teneur z(x) est le plus souvent dissymétrique (beaucoup de valeurs faibles et une queue de valeurs élevées), et est incompatible avec une distribution gaussienne. Parfois la distribution lognormale est appropriée: dans ce cas c'est le logarithme $Y(x)=\text{Log } Z(x)$ qui est gaussien, de sorte que Z(x) est l'exponentielle d'une gaussienne. La probabilité que $Z(x_0)$ dépasse un seuil z, connaissant les valeurs de Z en $x_1, x_2, ..., x_n$, est alors égale à la probabilité que $Y(x_0)$ dépasse la coupure $y=\text{Log } z$ connaissant les valeurs de $Y=\text{Log } Z$ en $x_1, x_2, ..., x_n$, probabilité que l'on a vu comment calculer sous des hypothèses multi−gaussiennes.

Un tel modèle peut être utilisé à d'autres fins. La variable conditionnée $Z(x_0)$ apparait comme l'exponentielle d'une gaussienne de moyenne y_0^K et de variance σ_K^2: on montre que c'est encore une variable lognormale, de moyenne (espérance conditionnelle):

$$z(x_0)^{EC} = \exp [y_0^K + \sigma_K^2/2]$$

et variance:

$$\sigma_{EC}^2 = [z(x_0)^{EC}]^2 [(\exp \sigma_K^2) - 1]$$

La moyenne fournit une estimation non linéaire de $Z(x_0)$, théoriquement (c'est−à−dire aux prix des hypothèses faites)

meilleure que ne le serait l'estimation linéaire de $Z(x_0)$ par krigeage. La variance (également variance d'estimation de $Z(x_0)$ par $Z(x_0)^{EC}$) est une variance conditionnelle qui dépend des valeurs aux points x_1, x_2, ..., x_n. Elle montre que la variabilité d'une variable lognormale augmente avec le carré de sa valeur attendue, "effet proportionnel" dont la géostatistique linéaire est impuissante à rendre compte, en particulier au niveau des variances de krigeage.

Espérance conditionnelle en modèle gaussien anamorphosé

Bien que la distribution lognormale s'utilise dans l'industrie minière de l'or en Afrique du Sud, il s'agit d'un exemple, certes pédagogique, mais très particulier, et le plus souvent le modèle gaussien est utilisé par l'intermédiaire d'une transformation plus générale, l'anamorphose gaussienne. La variable étudiée $Z(x)$ est alors considérée comme la transformée d'une variable gaussienne $Y(x)$, par une fonction non décroissante Φ (exponentielle dans le cas lognormal). On choisit de préférence pour $Y(x)$ une loi gaussienne réduite, et la connaissance de la fonction Φ est équivalente à celle de la distribution de $Z(x)$.

Si l'on connaît les valeurs de Y aux points x_1, x_2, ..., x_n, alors $Z(x_0)=\Phi[Y(x_0)]$ apparait, dans le modèle multigaussien, comme la transformée par Φ d'une gaussienne de moyenne y_0^K et de variance σ_K^2. On peut en déduire les probabilités de dépassement de seuil de $Z(x_0)$, l'estimation de $Z(x_0)$:

$$[Z(x_0)]^{EC} = \int \Phi(y)\ g_{y_0^K,\,\sigma_K^2}(y)\ dy$$

ou de fonctions de $Z(x_0)$, ainsi que les variances conditionnelles correspondantes.

Très souvent la variable d'intérêt, sur laquelle s'applique le seuillage par exemple, est définie, non sur un support ponctuel, mais sur un support plus grand, par exemple celui d'un bloc. L'estimation ponctuelle précédente n'est plus suffisante, et il est nécessaire de recourir à un modèle de changement de support.

Changement de support: le modèle gaussien discret

Imaginons l'exploitation sélective d'un gisement minier divisé en blocs égaux: une fois abattu, chaque bloc est envoyé en usine ou au stérile,

selon que sa teneur dépasse ou non une coupure fixée. Supposons de plus, cas idéal, que la teneur de chaque bloc, au moment de son exploitation, est parfaitement connue. Alors la connaissance de la distribution des teneurs de blocs détermine les quantités de minerai et de métal récupérables à toute coupure. Or comment prévoir, avant la mise en exploitation éventuelle du gisement, ces réserves récupérables?

Ce que l'on connaît, avant exploitation, ce sont des teneurs d'échantillons (carottes de sondages), autrement dit des teneurs mesurées sur un support quasi–ponctuel, ou en tout cas très petit. Si de plus, comme nous le supposerons, ces échantillons sont prélevés régulièrement dans le gisement, sont suffisamment nombreux, et si leurs analyses en teneur sont exactes, l'histogramme expérimental de leurs teneurs fournit une bonne image de la distribution statistique des teneurs sur support quasi–ponctuel à l'intérieur du gisement. Mais cette distribution présente davantage de valeurs riches et pauvres, moins de valeurs intermédiaires, que celle des teneurs de blocs, d'où un contraste entre riche et pauvre, et des promesses de sélectivité, illusoires (voir Chap. 3). Comment prévoir alors la distribution (inconnue) des teneurs de blocs, à partir de celle des échantillons? Tel est l'objet des modèles de changement de support, comme le modèle gaussien discret présenté ici.

Qu'il soit divisé en blocs ou en échantillons quasi–ponctuels, la teneur du gisement est la même: les moyennes des deux distributions sont identiques. Par ailleurs la géostatistique linéaire nous permet de calculer la variance de dispersion des blocs dans le gisement (Chap. 5). Mais moyenne et variance ne suffisent pas à déterminer la distribution des teneurs de blocs. Pour aller plus loin, nous écrirons que, connaissant la teneur d'un bloc, la teneur d'un échantillon ponctuel prélevé au hasard dans le bloc est, en valeur probable, égale à la teneur du bloc:

$$E(Z(x)\,|\,Z(v)) = Z(v)$$

Dans le modèle gaussien anamorphosé, la variable $Z(x)$ (teneur sur support ponctuel) est considérée comme la transformée, par une fonction Φ caractéristique de sa distribution, d'une variable gaussienne réduite $Y(x)$. De même la variable $Z(v)$ (teneur sur le support bloc) va être considérée comme la transformée, par une fonction Φ_v, d'une autre gaussienne réduite Y_v. Faisons l'hypothèse que, pour un bloc et un échantillon aléatoire dans ce bloc, les gaussiennes correspondantes forment un couple bigaussien de

corrélation r. Alors, à Y_v fixé, Y(x) suit une loi normale de moyenne rY_v et variance $1-r^2$. D'où la réécriture de la relation précédente:

$$\Phi_v(Y_v) = E(\Phi(Yx) \mid Y_v) = \int \Phi(y)\ g_{rYv,\,1-r^2}(y)\ dy$$

La variance de $\Phi_v(Y_v)$ augmente avec r: choisissant la valeur de ce paramètre de façon à respecter la variance des blocs $Z(v)=\Phi_v(Y_v)$, la relation détermine complètement l'anamorphose gaussienne des teneurs de blocs, et donc leur distribution:

$$F(z)=P(Z(v)<z)=P(Y_v<y)=\int^{y} g(t)dt=G(y)$$

avec $z=\Phi_v(y)$ et G fonction de répartition d'une gaussienne réduite.

La proportion de minerai récupérable à coupure z s'écrit donc $T(z)=1-G(y)$, tandis que la quantité de métal est donnée par:

$$Q(z)=\int_z t\ F(dt) = \int_y \Phi_v(y_v)\ g(y_v)\ dy_v$$

Telles sont, prévues par le modèle, les réserves récupérables à coupure z à l'échelle globale de l'ensemble du gisement.

Pour l'estimation locale, le gisement (supposé stationnaire de covariance C(h)) est "discrétisé" en ses différents blocs v_i (d'où le nom du modèle). La covariance $C(v_i,v_j)$ entre deux blocs v_i et v_j est obtenue par régularisation de la covariance ponctuelle (Chap. 5). De plus on considère que les échantillons disponibles sont aléatoires chacun indépendamment dans son bloc (on renonce à leur implantation exacte). Aux anamorphoses point et bloc près, le modèle est défini par les gaussiennes blocs et points. Sous des hypothèses bigaussiennes, on peut déterminer les covariances entre toutes ces gaussiennes. La covariance bloc-bloc $\varrho(v_i,v_j)$ se déduit de $C(v_i,v_j)$ (voir annexe). De plus on montre que si x_i et x_j sont des points différents appartenant respectivement aux blocs v_i et v_j, les covariances "point-bloc" $\varrho(x_i,v_j)$ et "point-point" $\varrho(x_i,x_j)$ sont respectivement égales à $r\varrho(v_i,v_j)$ et $r^2\varrho(v_i,v_j)$. Sous des hypothèses multi-gaussiennes, la loi conditionnelle des blocs, connaissant les échantillons de leur voisinage, permet alors l'estimation des réserves. Dans un contexte autre que minier, elle donnerait la probabilité pour que la valeur inconnue d'un bloc dépasse tel ou tel seuil. Mais l'estimation peut également être réalisée par krigeage disjonctif (voir plus loin), ne nécessitant alors que des hypothèses gaussiennes bivariables.

Autres outils non−linéaires, recours aux indicatrices

Oublions maintenant le problème important du changement de support, et revenons au problème du dépassement de seuil en un point. Ce qui est inconnu, au fond, c'est la variable indicatrice $I(Z(x) \geq z)$, égale à 1 si $Z(x)$ dépasse le seuil z, à 0 sinon. Et la probabilité conditionnelle, comprise entre 0 et 1 (qui n'est autre que l'espérance conditionnelle de cette indicatrice), en donne l'estimation. Un des avantages du modèle multi−gaussien est précisément de permettre l'accès à cette probabilité conditionnelle. Ceci est d'autant plus séduisant que le modèle multi−gaussien ne nécessite l'identification que de peu de paramètres (moyenne et covariance, outre l'anamorphose éventuelle). Cette économie de moyens n'est cependant qu'apparente, car faire usage d'un modèle multi−gaussien, c'est utiliser des hypothèses très fortes, puisqu'elles portent sur les lois multivariables (en pratique on ne peut guère espérer un contrôle allant au delà des lois bivariables). C'est un modèle particulier, qui peut se révéler inadapté. Comment alors procéder à l'estimation d'une indicatrice telle que $I(Z(x) \geq z)$?

La solution la plus simple consiste à considérer l'indicatrice $I(Z(x) \geq z)$ comme une variable isolée, que l'on pourra alors kriger à partir des valeurs 0 et 1 connues dans le voisinage pour cette variable. Si tous les poids sont positifs ou nuls, le résultat sera compris entre 0 et 1. Sinon, au gré de poids négatifs, une valeur estimée inférieure à 0 ou supérieure à 1 pourra apparaître. Cet inconvénient sera partagé par tous les estimateurs autres que la probabilité conditionnelle.

Cependant le comportement d'une indicatrice est en général loin d'être indépendant de celui des indicatrices à d'autres coupures, et il y a là risque substantiel de perte d'information. Par ailleurs, supposons que l'on cherche à évaluer les trois indicatrices $I(Z(x) \geq z)$, $I(Z(x) \geq z')$ et $I(z \leq Z(x) < z')$ pour $z' > z$. L'estimation de deux d'entre elles entraîne celle de la troisième, mais leur choix n'est pas indifférent: nous voilà ramenés à un problème analogue à celui du début du chapitre 8.

Une solution consiste à utiliser les mêmes poids pour toutes les indicatrices. Dans ce cas peu importe lesquelles on choisit pour en déduire les autres. Ce procédé se justifie d'ailleurs théoriquement dans le cadre du modèle mosaïque à valuations indépendantes (voir exercice 9.3), modèle dans lequel les indicatrices obéissent toutes à une structure commune.

Expérimentalement toutefois, on observe le plus souvent une destructuration des indicatrices aux seuils élevés, destructuration incompatible avec le modèle mosaïque. Une solution consiste, comme au chapitre 8, à procéder au cokrigeage des indicatrices. Dans le cas stationnaire auquel nous nous limitons, il nous faut alors connaître les moyennes des indicatrices:

$$E[I(Z(x)\geq z)]=P[Z(x)\geq z]=T(z)$$

c'est-à-dire la loi de probabilité, et les covariances simples et croisées:

$$\text{cov}[I(Z(x)\geq z), I(Z(x+h)\geq z')]=P[Z(x)\geq z, Z(x+h)\geq z']-T(z)T(z')$$

autrement dit les lois bivariables de (Z(x), Z(x+h)) pour tout h.

L'estimateur KD

Considérons pour simplifier une variable Y(x) prenant les valeurs 0, 1, 2 et 3. Comme par exemple:

$$I(Y(x)\geq 1)= I(Y(x)\geq 2)+I(Y(x)=1)$$

il y a équivalence entre la famille des indicatrices de classes I(Y(x)=i) et celle des indicatrices de seuils I(Y(x)≥j), et le cokrigeage de l'une équivaut à celui de l'autre. Par ailleurs la variable Y(x) peut s'écrire:

$$Y(x)=0\ I(Y(x)=0)+1\ I(Y(x)=1)+2\ I(Y(x)=2)+3\ I(Y(x)=3)$$

En effet, si Y(x) vaut 2 par exemple, alors toutes les indicatrices sont nulles, sauf I(Y(x)=2) qui vaut 1, et l'expression donne bien la valeur 2. De même, une fonction quelconque de Y(x), f[Y(x)], prenant les valeurs f_0, f_1, f_2, f_3, peut s'écrire:

$$Y(x)=f_0\ I(Y(x)=0)+f_1\ I(Y(x)=1)+f_2\ I(Y(x)=2)+f_3\ I(Y(x)=3)$$

Ainsi le cokrigeage des indicatrices permet-il d'estimer toute fonction de Y(x): c'est l'estimateur appelé krigeage disjonctif (KD). De même que le krigeage nécessite un modèle de covariance, de même le KD requiert une hypothèse sur les lois bivariables. En pratique on ne résout pas explicitement d'immenses systèmes de cokrigeage. On a recours à des lois bivariables "isofactorielles", dans lesquelles existe une famille de fonctions particulières équivalente à celle des indicatrices, les facteurs, entre lesquels il n'existe pas de corrélation spatiale. Il suffit alors de kriger séparément chacun d'eux pour en déduire le KD.

Exemple: le modèle de FA bigaussienne, défini par des lois (Y(x), Y(x+h)) bigaussiennes. Expérimentalement, ceci donne des nuages de corrélation différés (du vecteur h) de forme elliptique, d'autant plus resserrés que la corrélation est forte. De telles lois sont isofactorielles: les facteurs sont les polynômes d'Hermite (voir annexe). Une fois déterminé le développement d'une fonction f[Y(x)] en polynômes d'Hermite, son estimation KD est obtenue en krigeant chaque polynôme. Comme la structure spatiale de ceux-ci se dégrade très vite quand le degré augmente, leur krigeage tend rapidement vers leur moyenne, c'est-à-dire 0. D'où en pratique un nombre relativement faible, souvent une dizaine, de krigeages à effectuer.

Le modèle bigaussien s'utilise en général assorti d'une anamorphose. Mais d'autres modèles isofactoriels existent, basés également, ou non, sur une loi statistique (gamma, binomiale négative...).

Annexe: Les polynômes d'Hermite

Les polynômes d'Hermite sont étroitement liés à la loi normale, et ils jouent un rôle utile, si ce n'est indispensable, dans les modèles gaussiens. Supposons par exemple que l'on veuille faire un ajustement polynomial d'une anamorphose gaussienne $\Phi(y)$ (opération dont la justification apparaîtra plus loin). Les coefficients de chacun des monômes 1, y, y^2, ... changeront si l'on change le degré du polynôme ajusté. Au contraire, si l'on choisit les polynômes d'Hermite $H_n(y)$ de degré n=0, 1, 2, ... au lieu des monômes, le coefficient f_n de chacun d'eux ne bougera pas:

$$\Phi(y)=f_0 + f_1\ H_1[y] + f_2\ H_2[y] + \ldots = \sum_{n \geq 0} f_n\ H_n[y]$$

On peut ainsi stocker une centaine de coefficients, et choisir ensuite le degré 10, 20 ou 30 de l'approximation polynomiale.

Ces polynômes sont définis à partir de la densité de probabilité gaussienne réduite:

$$g(y)=\frac{1}{\sqrt{2\pi}}\ \exp -\frac{y^2}{2}$$

$$\text{par } H_n(y) = \frac{1}{\sqrt{n!}\ g(y)}\ \frac{d^n\ g(y)}{dy^n}$$

où $\sqrt{(n!)}$ est un facteur de normation. On a par exemple:

$$H_0(y) = 1$$

$$H_1(y) = -y$$

$$H_2(y) = (y^2-1) / \sqrt{2}$$

En dehors de H_0 (constant) ils vérifient, pour Y gaussienne réduite:

$$E[H_n(Y)]=0$$

$$\text{var } H_n(Y)=1$$

$$\text{cov}[H_p(Y(x)),\ H_n(Y(x))] =0 \text{ pour } n \neq p.$$

D'où, à partir de l'expression de $\Phi(y)$:

$$f_n = E[\Phi(Y(x))\ H_n(Y(x))]$$

ce qui permet en pratique le calcul des coefficients de l'anamorphose. On peut aussi montrer:

$$\text{var } Z=\text{var } \Phi(Y)= \sum_{n\geq 1} (f_n)^2$$

Il est possible, de la même façon, de développer en polynômes d'Hermite d'autres fonctions de Y.

Ces polynômes vont être surtout très intéressants dans le cas de couples bigaussiens. En effet si (Y_0,Y_1) est un couple gaussien réduit de corrélation r, on a $E[H_n(Y_0) \mid Y_1]=r^n H_n(Y_1)$, d'où l'on déduit:

$$\text{cov}[H_n(Y_0),\ H_n(Y_1)]=r^n$$

$$\text{cov}[H_n(Y_0),\ H_p(Y_1)]=0 \text{ pour } n \neq p$$

Voici quelques conséquences de ces relations en géostatistique.

Pour une FA Y(x), dont les lois bivariables (Y(x),Y(x+h)) sont (bi−)gaussiennes de corrélation $\varrho(h)$, on a:

$$\text{cov}[H_n(Y(x)), H_p(Y(x+h))]=0 \text{ pour } n \neq p$$

$$\text{cov}[H_n(Y(x)), H_n(Y(x+h))]=[\varrho(h)]^n$$

Les polynômes d'ordre différent sont sans corrélation spatiale, et la corrélation spatiale du polynôme d'ordre n est $[\varrho(h)]^n$, se dégradant rapidement vers un effet de pépite pour n grand. Il s'ensuit que:

$$\text{cov}[Z(x), Z(x+h)]=\text{cov}[\Phi[Y(x)], \Phi[Y(x+h)]]= \sum_{n\geq 1} (f_n)^2\ [\varrho(h)]^n$$

En pratique, l'anamorphose étant connue, une telle relation permet de calculer la structure de la gaussienne à partir de la structure brute, ou inversement. Mais elle montre aussi que la corrélation spatiale de Z est nécessairement moins continue que celle de sa transformée gaussienne. De même, toujours sous des hypothèses bigaussiennes, la corrélation spatiale de la gaussienne est toujours plus continue que celle de toute autre transformée. Noter d'ailleurs que, de façon générale en géostatistique, la structure d'une variable n'a aucune raison d'être conservée quand on transforme cette variable (exception faite du modèle mosaïque, voir exercice 9.3).

Autre conséquence: la correction de support. Ecrivant $Z(v)=E(Z(x)|Z(v))$ (voir plus haut), on obtient l'expression de l'anamorphose des blocs par son développement en polynômes d'Hermite:

$$\Phi_v(Y_v)=f_0 + f_1\ r\ H_1[Y_v] + f_2\ r^2\ H_2[Y_v] + \ldots$$

$$= \sum_{n \geq 0} f_n\ r^n\ H_n[Y_v]$$

où le "coefficient de changement de support" r, corrélation du couple gaussien $(Y(x),Y_v)$, est déterminé expérimentalement de façon à respecter la variance des blocs préalablement calculée:

$$\text{var } Z(v)=\text{var } \Phi_v(Y_v)= \sum_{n\geq 1} (f_n)^2\ r^n$$

Enfin les polynômes d'Hermite jouent un rôle central dans le KD gaussien (voir plus haut).

Exercices

▶ *Exercice* 9.1. Krigeage lognormal.

On considère une variable stationnaire lognormale $Z(x)=\exp Y(x)$, dont le logarithme $Y(x)$ est une gaussienne de moyenne 0 et de variance 1. Valeur de $Z(x)$ pour $Y(x)$ égal à $-2, -1, 0, 1, 2$?

Réponse : 0.14, 0.37, 1, 2.72, 7.39.

Moyenne de $Z(x)$? Noter que la moyenne de $Z(x)$ n'est pas simplement l'exponentielle de la valeur moyenne de $Y(x)$.

Réponse : 1.65.

On suppose que le krigeage, en un point x_0 à partir de valeurs dans son voisinage, prend la valeur 0 et est assortie d'une variance de krigeage

de 0.5. Sous une hypothèse multigaussienne, la distribution conditionnelle de $Y(x_0)$ est donc gaussienne, de moyenne 0 et variance 0.5. Que vaut l'espérance conditionnelle de $Z(x_0)$? Remarquer que cet estimateur, appelée parfois krigeage lognormal, n'est pas simplement l'exponentielle du krigeage du logarithme.

Réponse : 1.28.

▶ *Exercice* 9.2. Estimation des réserves par la technique des variables utiles.

Montrer que, dans le modèle gaussien discret, la valeur probable du minerai à coupure z sur un bloc contenant un échantillon à teneur Z(x) s'écrit:

$$\int_y g_{rY(x),\,1-r^2}(t)\ dt$$

avec y et Y(x) vérifiant $z=\Phi_v(y)$ et $Z(x)=\Phi(Y(x))$.

De même le métal probable s'exprime à partir de Y(x) (plus compliqué). On peut donc calculer ces variables en chaque point échantillonné, puis obtenir une estimation des réserves en en faisant le krigeage. Avantage: on peut utiliser alors un krigeage ordinaire, qui évitera l'attraction vers la moyenne dans les zones peu informées, lorsque la stationnarité est mal assurée.

▶ *Exercice* 9.3. Modèle mosaïque.

Le champ est partitionné en compartiments (partition aléatoire stationnaire), et on appellera $\varrho(h)$ la probabilité pour que deux points à distance h appartiennent au même compartiment. A chaque compartiment on attribue une valuation aléatoire, indépendamment des autres compartiments, mais selon une loi de probabilité commune. Montrer que la structure de la FA ainsi construite ne dépend que de $\varrho(h)$:

$$\text{cov}[Y(x),Y(x+h)]=\text{var}Y(x)\ \varrho(h)$$

Montrer que les structures simples ou croisées de toutes fonctions sont proportionnelles à $\varrho(h)$:

$$\text{cov}[f[Y(x)],g[Y(x+h)]]=\text{cov}[f[Y(x)],g[Y(x)]]\ \varrho(h)$$

En déduire que les indicatrices sont en corrélation intrinsèque (cf Chap. 8), que leur cokrigeage coïncide avec leur krigeage séparé, et que le KD coïncide avec le krigeage.

▶ *Exercice* 9.4. Modèle à résidus d'indicatrices orthogonaux.

Soient A_0, A_1 et A_2 des ensembles aléatoires stationnaires indépendants. On considère la FA Y(x) égale à 0 si x appartient à A_0, à 1 si x appartient à A_1 mais non à A_0, à 2 si x appartient à A_2 sans appartenir à A_0 ou A_1, et à 3 si x n'appartient à aucun ensemble. On notera $T_i = P(Y(x) \geq i)$.

Montrer que, si $Y(x) \geq i$, la probabilité pour que $Y(x) \geq j > i$ ne dépend pas du fait que $Y(x+h)$ soit inférieure ou non à i (absence d'effet de bordure sur le domaine défini par $Y(x) \geq i$).

En déduire que le variogramme croisé des indicatrices $I(Y(x) \geq i)$ et $I(Y(x) \geq j)$ est proportionnel au variogramme de $I(Y(x) \geq i)$.

Vérifier la relation suivante:

$$E\left[\frac{I(Y(x) \geq i+1)}{T_{i+1}} \,|\, I(Y(x) \geq i)\right] = \frac{I(Y(x) \geq i)}{T_i}$$

Montrer que les résidus correspondants:

$$\frac{I(Y(x) \geq 1)}{T_1} - 1,$$

$$\frac{I(Y(x) \geq 2)}{T_2} - \frac{I(Y(x) \geq 1)}{T_1},$$

$$\frac{I(Y(x) \geq 3)}{T_3} - \frac{I(Y(x) \geq 2)}{T_2}$$

sont sans corrélation spatiale. Comment effectuer alors le cokrigeage des indicatrices, et donc le KD?

10 LES SIMULATIONS

Qu'est−ce−qu'une simulation? Pourquoi des simulations?

La cartographie d'une variable régionalisée, telle qu'on peut la faire par exemple par krigeage à partir de points de données, n'est pas LA réalité. La recherche d'une précision optimale conduit inévitablement à un lissage: quantité de détails présents en réalité ne peuvent être reproduits à partir des seuls points de données et sont gommés de la cartographie. Paradoxalement même, d'une certaine façon, plus la variabilité spatiale de la réalité est prononcée, moins les détails sont accessibles à la cartographie, et plus le lissage est important (dans le cas pépitique pur d'une variabilité extrême, l'estimation hors des points de données est constante et égale à la moyenne des données).

Parfois il importe, non d'obtenir la meilleure précision, mais de reproduire la variabilité spatiale: ainsi pour tester des scénarios d'exploitation minière, ou plus généralement pour évaluer une quantité complexe et sinon inaccessible. On a alors recours aux simulations. De même que la réalité était considérée comme une réalisation du modèle de FA, de même chaque simulation est une réalisation, sur un certain domaine, de ce même modèle. En outre nous appellerons simulation conditionnelle une simulation restituant, aux points de données, les valeurs qui y sont connues. (Fig. 17.)

Il existe nombre de méthodes et variantes pour simuler des FA. Certaines nécessitent un maillage, une discrétisation régulière de l'espace (voir exercices). Nous préférons présenter ici des méthodes s'en affranchissant.

Jetons aléatoires, dilution de points poissonniens

La méthode des jetons consiste en ceci. On génère, à 2 ou 3D, des points Poissonniens selon une densité (nombre moyen de points par unité de surface/volume) θ donnée. Entendons par là, pour simplifier, des points aléatoires, indépendants, et uniformes sur le domaine (exercice 10.3). En chacun de ces points, on implante un "jeton", par exemple disque à 2D ou sphère à 3D, de covariogramme géométrique K(h) (chapitre 2). La valeur en un point quelconque est alors le nombre (aléatoire) de jetons recouvrant ce point. Hors des bordures (voir exercice 10.4), ce nombre a pour moyenne $\theta K(0)$ et obéit à la covariance $\theta K(h)$. En prenant des sphères à 3D, on peut ainsi simuler

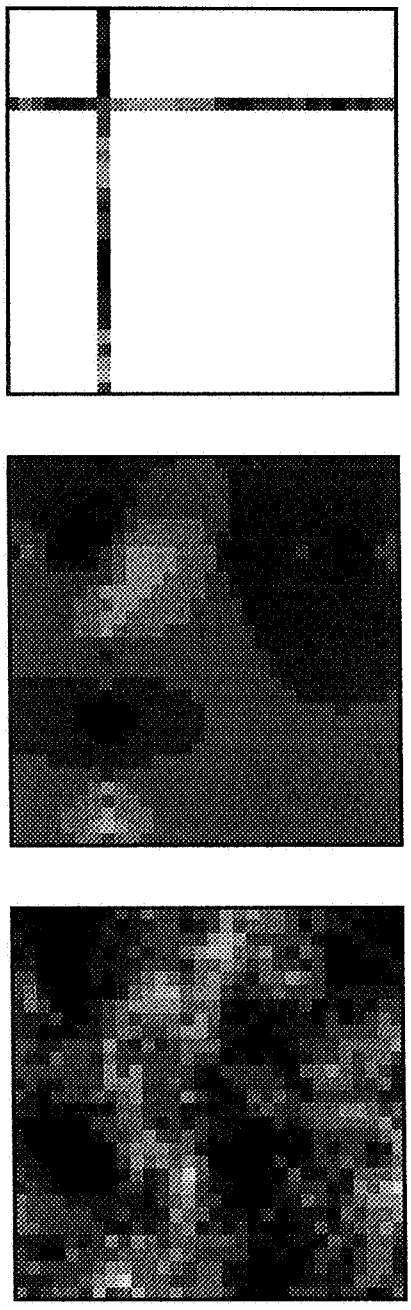

Fig. 17: Krigeage (au milieu) et une simulation conditionnelle (en bas), à partir des mêmes données (en haut).

une FA stationnaire de covariance sphérique (d'où ce nom, cf exercice 2.5).

Les variantes sont multiples. La taille ou la forme des jetons peut être par exemple aléatoire. Par ailleurs, compter le nombre de jetons recouvrant chaque point revient à implanter, en chaque point Poissonnien, un jeton valué à 1 et à faire la somme de ces valuations. Plutôt que d'implanter, en chaque point Poissonnien, une fonction égale à 1 sur le jeton, et à 0 sinon, on peut implanter une fonction f(x), s'annulant de préférence au-delà d'une certaine distance. Cette variante, connue sous le vocable de dilution de points Poissonniens, revient en effet à diluer par la fonction f(x) une masse ponctuelle qui serait implantée en chaque point Poissonnien. La FA obtenue par sommation a une covariance proportionnelle au covariogramme transitif de la fonction f(x) (Cf. Chap. 2):

$$g(h)=\int f(x)\, f(x+h)dx$$

Les méthodes spectrales

Ces méthodes sont basées sur la décomposition d'un processus stationnaire en processus sinusoïdaux, et de sa covariance en somme de cosinus. Considérons par exemple à 1D le processus sinusoïdal de fréquence u:

$$A\cos(2\pi ux+\phi)$$

Si A est une variable aléatoire de variance σ^2, et ϕ une phase uniforme entre 0 et 2π, ce processus a une covariance stationnaire égale à:

$$C(h)=\sigma^2\cos(2\pi uh)$$

La somme de tels processus indépendants:

$$\sum A_i\cos(2\pi u_i x+\phi_i)$$

a alors comme covariance:

$$C(h)=\sum \sigma_i^2\cos(2\pi u_i h),$$

et l'ensemble des fréquences u_i et de leurs "énergies" σ_i^2 en constitue le "spectre". Plus généralement:

$$\int_0^\infty dA(u)\cos(2\pi ux+\phi(u))$$

aura comme covariance:

$$C(h)=\int_0^\infty d\sigma^2(u)\cos(2\pi uh)$$

(avec $\int_0^\infty d\sigma^2(u) = C(0)$ supposé fini).

Il s'agit là en fait d'un résultat général (théorème de Bochner). Toute covariance stationnaire en effet, et pas seulement à 1D, étant de type positif, se décompose en somme de cosinus:

$$C(h)=\int \cos 2\pi<uh>\ \chi(du)$$

où χ est une mesure positive sommable ($\int\chi(du)=C(0)$ fini), et inversement.

(Dans cette expression, si C(h) est défini à 3D par exemple, u est un vecteur fréquence à 3 composantes, et <uh> représente le produit scalaire habituel: $u_1h_1+u_2h_2+u_3h_3$.)

Ceci permet théoriquement de simuler la FA correspondante comme une somme de processus sinusoïdaux indépendants. Il s'agit cependant d'une somme, en général infinie, sur l'espace des fréquences, et les hautes fréquences sont d'autant plus sollicitées que le comportement aux petites distances de la variable et de sa covariance est irrégulier. Il devient alors problématique de simuler, en pratique par une somme finie, une FA non dérivable (cas courant des covariances linéaires à l'origine), ce qui réduit l'intérêt de cette méthode directe.

Les bandes tournantes

Pour contourner la difficulté, G.Matheron a proposé de considérer les fréquences u=rs, d'abord selon leur direction s=u/|u|, puis selon leur module r=|u|. On a alors:

$$C(h)=\int F(ds)\int_0^\infty \cos 2\pi r<sh>G_s(dr)$$

où F(ds)est la loi du vecteur direction s sur la sphère unité, et $G_s(dr)$ celle du module r le long de cette direction s. Mais (toujours le

théorème de Bochner) $\int_0^\infty \cos 2\pi r<sh>G_s(dr)$ est elle−même une covariance stationnaire à 1D, disons $C_s(<sh>)$, d'où la décomposition générale:

$$C(h)=\int F(ds)\, C_s(<sh>)$$

Ainsi, en simulant le long d'une droite de direction s, et pour des directions s suivant la loi F(ds), une FA unidimensionnelle Ys de covariance C_s, puis en effectuant un épandage $Y_s(<sh>)$ de chacune de ces simulations dans l'espace (Fig. 18), on obtient par:

$$Z(x)=\int F(ds)\, Y_s(<sh>)$$

une simulation de covariance stationnaire C(h).

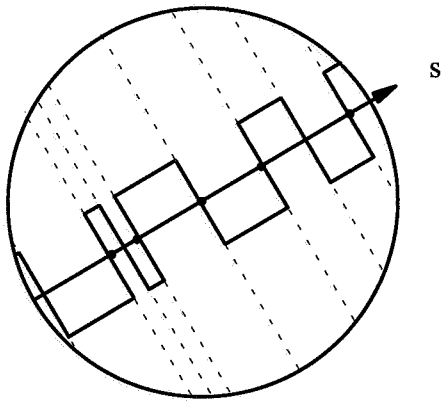

Fig. 18: Simulation 1D (ici alternance de valeurs +1 et −1) et son épandage dans l'espace.

Dans le cas isotrope et à 3D, s décrit uniformément la sphère unité, et C_s ne dépend plus de s, soit C_1. On montre alors le résultat:

$$C(h) = \frac{1}{h}\int_0^h C_1(u)\, du$$

Inversement:

$$C_1(h) = \frac{d}{dh}[hC(h)]$$

donne la covariance qu'il convient de simuler sur chaque droite pour en déduire C(h) à 3D (les formules sont plus compliquées pour passer de 1 à 2D).

Pratiquement on utilise un grand nombre N (couramment quelques centaines) de directions s_j de droites tirées uniformément, et le long de ces droites, on construit des simulations Y_{sj} de covariance C_1. La FA définie par:

$$Z(x) = \frac{1}{\sqrt{N}} \sum Y_{sj}(<s_j .x>)$$

obéit alors à la covariance C(h).

Exemples:

1) La covariance exponentielle:

$$C(h) = \exp(-|h|/a)$$

à 3D sera obtenue par simulations 1D de la covariance:

$$C_1(h) = (1 - |h|/a) \exp(-|h|/a)$$

Ch. Lantuéjoul a montré qu'on pouvait obtenir une telle simulation 1D en valuant respectivement à +1 et −1 la première et la seconde moitié de chacun des intervalles formés par les points consécutifs d'un processus de Poisson sur la droite.

2) Simuler à 3D la covariance sphérique de portée a:

$$C(h) = 1 - (3/2)(|h|/a) + (1/2)(|h|/a)^3 \text{ pour } |h| < a,$$

s'obtiendra en simulant pour chaque (direction de) droite la covariance:

$$C(h) = 1 - 3(|h|/a) + 2(|h|/a)^3 \text{ pour } |h| < a,$$

Celle−ci peut être obtenue en diluant des points Poissonniens sur la droite par la fonction f(x) égale à x pour $|x| < a/2$ et à 0 sinon (cf. exercice 2.4).

3) On remarque que tout monôme $|h|^\alpha$ de la covariance de départ donne, dans la covariance 1D associée, un monôme de même degré α, résultat qui demeure valable en non stationnaire pour des variogrammes ou des covariances généralisées.

Simuler un variogramme linéaire revient alors à simuler un variogramme linéaire sur chaque droite. Ceci peut s'obtenir en

considérant, pour chacune des droites, une valuation s'incrémentant aléatoirement de +1 ou −1 au passage de chacun des points d'un processus de Poisson.

FA gaussienne et conditionnement d'une simulation

Lorsqu'on fait une simulation, on veut, dans le cas stationnaire, non seulement reproduire une covariance donnée, mais également une distribution, un histogramme. Or les techniques de simulations obtenues par sommation à partir d'un grand nombre de droites ou de jetons génèrent des distributions (asymptotiquement) gaussiennes. C'est pourquoi on simule le plus souvent, non pas directement la variable brute, mais sa transformée gaussienne (avec sa propre structure), avant de repasser en brut.

Par ailleurs les FA multi−gaussiennes (où théoriquement toutes les combinaisons linéaires sont gaussiennes, Cf Chap. 9) se prêtent bien au conditionnement. En effet, on a en tout point x:

$$Y(x) = Y(x)^K + [Y(x) - Y(x)^K]$$

où $Y(x)^K$ représente le krigeage de Y(x) à partir des données (de la transformée gaussienne). Dans le cas d'une FA multi−gaussienne, le résidu en tout point $[Y(x) - Y(x)^K]$ est indépendant des valeurs aux points de données. L'idée est alors de substituer à ce résidu indépendant mais inconnu un résidu simulé ayant exactement la même structure.

Pour ce faire, on fabrique une simulation non conditionnelle de la variable, soit $Y_S(x)$, sur le domaine considéré. Puis on calcule en tout point x, le résidu de son krigeage à partir des valeurs prises par Y_S aux points de données:

$$[Y_S(x) - Y_S(x)^K]$$

La recombinaison:

$$Y_{SC}(x) = Y(x)^K + [Y_S(x) - Y_S(x)^K]$$

donne alors une autre simulation de la FA, mais qui est maintenant conditionnelle: en un point de donnée on retrouve bien la valeur connue:

$$Y_{SC}(x_i) = Y(x_i) + [Y_S(x_i) - Y_S(x_i)] = Y(x_i)$$

Autres modèles

Il existe une immense variété de modèles qu'il est possible de simuler, qu'il s'agisse de fonctions, d'ensembles ou de partitions aléatoires. Souvent il sera nécessaire de construire un modèle ad hoc qui pourra représenter une réalité donnée. Voici deux modèles élémentaires, effectivement utilisés dans la modélisation des ressources naturelles.

Le seuillage d'une FA

Il s'agit simplement de seuiller une simulation d'une FA, à loi stationnaire gaussienne par exemple. L'application d'un seuil unique conduit à simuler un ensemble aléatoire et son complémentaire, autrement dit à partitionner l'espace en deux faciès. L'application de seuils multiples donne une partition en plusieurs faciès séquentiels (pour passer d'un faciès à un autre, on passe par les faciès intermédiaires). Ce modèle est couramment utilisé dans l'industrie pétrolière pour modéliser certains dépôts fluvio−deltaïques (faciès s'étageant du grès à l'argile en passant par du grès argileux et de l'argile gréseuse).

Les seuils dont déterminés de façon à respecter les proportions des différents faciès. Très souvent cependant, celles−ci montrent des variations systématiques en vertical, que l'on peut bien mettre en évidence expérimentalement en traçant les courbes de proportions des faciès selon la verticale. Pour respecter cette donnée, le seuillage devra lui aussi varier en vertical.

Le schéma booléen

Dans la technique des jetons aléatoires, on implantait en des points Poissonniens des objets, les jetons, et on faisait la somme des valuations. Le schéma booléen lui, est l'ensemble aléatoire obtenu par réunion de tels objets. Il existe quantité de variantes (dépendance entre objet et point d'implantation, non−stationnarité du processus de points Poissonniens, contraintes exercées de façon à ce que les objets ne puissent se recouvrir...).

Exercices

▶ *Exercice* 10.1. Variogramme linéaire sur la droite discrétisée.

Soient Y_1, Y_2, ... des VA indépendantes d'espérance nulle et de variance unité. On pose $Z(0)=0$ et:

$$Z(i) = Z(i-1)+Y_i$$

Montrer que les accroissements de Z sont d'espérance nulle et de variance linéaire (voir aussi exercice 4.2).

▶ *Exercice* 10.2. Covariance exponentielle sur la droite discrétisée.

$Y_1, Y_2, \ldots$ sont des VA gaussiennes, indépendantes, d'espérance nulle et de variance unité. On pose $Z(0)=0$ et:

$$Z(i) = rZ(i-1) + \sqrt{1 - r^2}\, Y_i$$

avec $0<r<1$.

Montrer que les $Z(i)$ sont multi-gaussiens, d'espérance nulle, de variance unité, et de covariance stationnaire:

$$C(h)=\mathrm{cov}[Z(i), Z(i+h)]=r^{|h|}=\exp(-|h| \mathrm{Log}\,(1/r))$$

Que se passe-t-il si $-1<r<0$?

Réponse : La covariance est de signe alterné:

$$(-1)^{|h|} \exp(-|h| \mathrm{Log}\,|1/r|).$$

▶ *Exercice* 10.3. Points au hasard (processus de Poisson).

L'espace (à 1, 2 ou 3D) est divisé en parcelles (segments, carrés ou cubes) élémentaires dx_i infiniment petites. Dans chacune des parcelles, on implante indépendamment un point avec une probabilité θdx_i (aucun point sinon). Le nombre de points $N(V)$ ainsi générés dans un domaine V est une variable aléatoire. Montrer qu'elle est égale en espérance à $\theta|V|$. En déduire que θ représente la densité de points par unité de longueur, surface ou volume (prendre $|V|=1$).

$N(V)$ suit ce que l'on appelle une distribution de Poisson de paramètre $\theta|V|$, et on montre qu'à nombre de points $N(V)=n$ fixé, ces points se distribuent indépendamment et uniformément sur V. Montrer que si V_1 et V_2 sont deux domaines disjoints, $N(V_1)$ et $N(V_2)$ sont deux variables de Poisson indépendantes de paramètres $\theta|V_1|$ et $\theta|V_2|$, et que leur somme suit une loi de Poisson de paramètre égal à la somme des paramètres.

Montrer que la probabilité pour que $N(V)$ soit nul est égale à $\exp-\theta|V|$. A 1D, en déduire que les intervalles séparant deux points de Poisson consécutifs obéissent à une loi exponentielle $P(L \geq h)=\exp-\theta h$ (procédé pratique de simulation à 1D).

A 2D, considérons un rectangle de côtés L_1 et L_2, et un processus de Poisson de paramètre θ_1 sur L_1. Chacun de ces points est considéré

comme l'abscisse d'un point du rectangle, dont l'ordonnée est tirée aléatoirement sur L_2. Montrer que les points ainsi générés forment un processus de Poisson de paramètre $\theta_2 = \theta_1/L_2$ sur le rectangle (procédé pratique de simulation à 2D ou plus.)

▶ *Exercice* 10.4. Effets de bord.

En chacun des points de Poisson d'un certain domaine, on implante un jeton. Montrer que le nombre de jetons recouvrant un point du domaine est en moyenne plus faible sur les bordures. En déduire que, pour éviter cela, les points doivent être générés sur un domaine plus grand que le champ d'intérêt.

▶ *Exercice* 10.5. Moyennes mobiles.

Rien de plus naturel, pour faire apparaître une structure spatiale à 1D, que de se donner des variables indépendantes à maille régulière le long de la droite, et d'en faire une moyenne mobile. Soient donc Y(0), Y(b), Y(2b), ... des VA indépendantes de même espérance et même variance, implantées aux points 0, b, 2b... On définit, sur la droite ainsi discrétisée, la moyenne mobile:

$$Z(ib)=f(0)Y(ib)+f(b)Y((i+1)b)+f(2b)Y((i+2)b)+...$$
$$+f(kb)Y((i+k)b)$$

où f est une fonction nulle en dehors de [0, kb]. Montrer que Z(ib) est stationnaire. Espérance et covariance?

Que se passe-t-il, pour une discrétisation b très fine, lorsque les Y(ib) sont chacun égaux à 1 avec une probabilité θb, à 0 sinon?

Réponse : D'après l'exercice 10.3, on est proche d'une dilution de points poissonniens de densité θ, méthode qui apparait donc comme une généralisation au cas continu des moyennes mobiles.

11 CONCLUSION

Nous avons pu constater, dans la progression des chapitres, le confort de plus en plus grand que procure le maniement des modèles probabilistes (FA) pour représenter les phénomènes réels étudiés. Mais si l'on revient aux chapitres 2, 3 et 4, on s'aperçoit que ce cadre probabiliste n'est, au fond, pas absolument indispensable. Un de ses grands mérites est sa maniabilité.

Ce n'est pas le seul avantage. Les modèles de FA permettent de donner un sens, en les intégrant dans un contexte plus large, à des statistiques expérimentales. Prenons l'exemple d'une statistique appelée parfois variogramme relatif, calculée pour une variable positive en prenant la moyenne, pour chaque classe de distance h, des valeurs:

$$\left[\frac{z(x+h)-z(x)}{z(x+h)+z(x)}\right]^2$$

A première vue, cette statistique semble assez parlante. Mais, pour aller plus loin, à quoi relier cette statistique isolée? Quelle est l'allure d'un phénomène régi par un variogramme relatif donné? Quel type d'information nous procure, au fond, ce variogramme relatif? Une telle statistique gagnerait un intérêt considérable si l'on pouvait lui trouver un cadre interprétatif pas trop restrictif.

Laissons-là ce point, et remarquons que très souvent en géostatistique, on travaille sur un phénomène (par exemple un gisement minier) considéré comme une réalisation unique d'une FA. La stationnarité (invariance par translation), qu'il s'agisse de la FA ou de ses incréments, apparait alors comme une hypothèse nécessaire à l'inférence du modèle par des statistiques expérimentales.

Si les modèles de FA peuvent offrir une grande commodité, ils peuvent aussi se révéler pernicieux. On a vu, chapitre 7, le piège que peut receler le modèle, pourtant si simple d'apparence, constitué d'une dérive et d'un résidu stationnaire. Même s'il est possible de raisonner, en théorie, sur un modèle donné, c'est son caractère opératoire qui lui confère un intérêt pratique.

La géostatistique n'est pas une discipline isolée. Déjà dans les années 50, ont été développées, dans des domaines d'application non miniers, des techniques voisines (Matern en génie forestier, Gandin et son analyse objective en météorologie), toutes ces techniques faisant

évidemment partie des statistiques spatiales. Par l'usage qu'elle fait des modèles probabilistes de fonctions aléatoires, la géostatistique peut être vue comme une partie intégrante des probabilités. Mais à côté de ses développements propres, l'originalité de la géostatistique réside dans la mise à disposition de tels outils probabilistes à l'ingénieur, au praticien de la mine, du pétrole, ou d'un autre domaine, et ceci en particulier grâce à des concepts clés (par exemple: la variance d'estimation, la variance de dispersion, le support et les lois de changement de support), adaptés à la compréhension physique du phénomène. Appliqué à l'information très fragmentaire que constituent les données, ceci se traduit par un savoir–faire, avec tours de main et approximations adéquates.

La géostatistique ne se résume pas à un ensemble de méthodes et de modèles répertoriés. Dans tel gisement minier, le géostatisticien devra manier habilement différents ingrédients pour rendre compte d'une relation particulière entre teneur et géologie. Dans d'autres cas, une recherche plus approfondie sera nécessaire, par exemple pour modéliser par des ensembles ou partitions aléatoires les différents faciès de gisements pétroliers, ou encore pour tenir compte d'équations physiques dans l'écoulement spatio–temporel d'une nappe d'eau. La géostatistique apparait comme une discipline toujours en évolution, tant sur le plan pratique, que sur le plan méthodologique, comme le montre l'intérêt croissant pour les ensembles aléatoires et pour les modèles physiques.

12 REFERENCES ET BIBLIOGRAPHIE

Armstrong M. (Ed.), *Geostatistics*. Proc. 3rd Intern. Geostatistics Congress, Avignon, 5–9 Sept. 1988, Kluwer Acad. Publ., Dordrecht, Holland, 1989. (Quantitative Geology and Geostatistics 4).

Armstrong M. & Dowd P.A. (Eds.). *Geostatistical simulations.* Dordrecht: Kluwer Academic Publ., 1994. 255 p. (Quantitative Geology and Geostatistics 7).

Chauvet P. Processing data with a spatial support : geostatistics and its methods. *Les Cahiers de Géostatistique,* fasc. 4, Ecole des Mines de Paris, 1993. 57 p.

Chauvet P. Aide–mémoire de géostatistique linéaire. *Les Cahiers de Géostatistique*, fasc. 2, Ecole des Mines de Paris, 1994, 210 p.

Cressie N., *Statistics for Spatial Data*, Wiley, New York, 1991, 900 p.

David M. (1977), *Geostatistical Ore Reserve Estimation,* Elsevier, Amsterdam.

David M. (1988), *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation,* Elsevier, Amsterdam. 232 p.

De Fouquet C. Simulation conditionnelle de fonctions aléatoires : cas gaussien stationnaire et schéma linéaire (cours de simulation), Centre de Géostatistique, Fontainebleau, Ecole des Mines de Paris, 1993, 91 p.

Delfiner P., Matheron G. Les fonctions aléatoires intrinsèques d'ordre k, Centre de Géostatistique, Fontainebleau, Ecole des Mines de Paris, 1980, 36 p.

Gandin L. S. (1965), *Analyse Objective*, Organisation météorologique Mondiale, Moscou.

Guarascio M. et al. (Eds.), *Advanced Geostatistics in the Mining Industry*, Proceedings, NATO ASI: Advanced Geostatistics in the Mining Industry (Rome, Oct. 1975), D. Reidel Publ. Co. 1976.

Isaaks E.H. and Srivastava R.M., *An Introduction to Applied Geostatistics,* Oxford University Press, 1989.

Journel A. and Huijbregts C. (1978), *Mining Geostatistics* (Academic Press, London).

Krige D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand, *Journal of the Chemical, Metallurgical and Mining Society of South Africa,* December 1951, Vol. 52, no. 6, pp. 119–139, Johannesburg.

Matern B. (1960) *Spatial Variations*, réédité par Springer–Verlag 1986.

Matheron G. *Les Variables Régionalisées et leur Estimation.* Paris : Masson, 1965, 306p.

Matheron G. La Théorie des Variables Régionalisées et ses Applications. *Les Cahiers du Centre de Morphologie Mathématique*, fasc. 5, Ecole des Mines de Paris, 1970, 212 p.

Matheron G. Estimer et Choisir. *Les Cahiers du Centre de Morphologie Mathématique*, fasc. 7, Ecole des Mines de Paris, 1978, 175 p.

Matheron G. *Estimating and Choosing – An essay on probability in practice.* Translated from the French "Estimer et Choisir" by A.M. Hasofer, Berlin: Springer–Verlag, 1989, 141 p.

Matheron G., Armstrong M. (Eds.), *Geostatistical Case Studies*, D. Reidel Publ. Co., Dordrecht, Holland, 1987, 248 p. (Quantitative Geology and Geostatistics 2).

Rivoirard J., *Introduction to disjunctive kriging and nonlinear geostatistics.* Oxford University Press, 1994. 181 p. Spatial Information Systems.

Sandjivy L., Etude géostatistique de données recueillies sur du mil au Niger, Centre de Géostatistique, Fontainebleau, Ecole des Mines de Paris, 1980.

Sichel H. S. (1949). Mine valuation and maximum likelihood, Master's thesis, University of the Witwatersrand, Johannesburg.

Sichel H. S. (1966). The estimation of means and associated confidence limits for small samples from lognormal populations, in *Symposium on Mathematical Statistics and Computer Applications in Ore Valuation* (South African Institute of Mining and Metallurgy, Johannesburg), 106–22.

Soares A. (Ed.), *Geostat Tróia '92*, Proc. 3rd Intern. Geostatistics Congress, Kluwer Academic Publ., 1993.

Verly G., David M., Journel A.G., and Maréchal A. (Eds). *Geostatistics for Natural Resources Characterization.* Proc. 2nd Intern. Geostatistics Congress, Tahoe, Sept. 1983, NATO ASI Series, Reidel, Dordrecht, Holland, 1984.

Wackernagel H., *Multivariate Geostatistics.* Springer–Verlag, 1995, 256 p.

de Wijs H.J. Statistics of ore distribution, Part I. Frequency distribution of assay values, *Geologie en Mijnbouw*, 13e Jaargang, No 11, Nov. 1951, pp. 365–375

de Wijs H.J. Statistics of ore distribution, Part II. Theory of binomial distribution applied to sampling and engineering problems, *Geologie en Mijnbouw*, Jan. 1953, pp. 12–24.