

CHANGEMENT DE MAILLE ET CHANGEMENT DE SUPPORT

WACKERNAGEL H.

RESUME

En parlant d'échelle, il est important de distinguer entre la maille et le support de l'information, ce dernier étant un élément de volume ou de surface, parfois assimilé à un point, sur lequel une information est définie. Une modification de la maille ou une modification du support ont des implications différentes, ce qui sera illustré ici dans le contexte simple de la géostatistique linéaire. Le problème pourra également être multivariable, à savoir que l'on dispose de plusieurs sources d'information sur des mailles ou des supports différents.

1 MAILLE ET ECHELLES

La maille d'échantillonnage définit les distances entre des paires d'échantillons. Nous désignons par h un vecteur décrivant à la fois la distance entre deux points dans l'espace et l'orientation de l'axe passant par la paire de points. Les différentes échelles seront décrites par l'ensemble de paires de points de données réalisées par la maille d'échantillonnage. La plus petite échelle à laquelle l'échantillonnage nous permette d'accéder sera, dans le cas d'une maille régulière, le pas de cette maille. La plus grande échelle sera égale à la moitié de la largeur du domaine échantillonné: la moitié seulement, car c'est au plus jusqu'à cette distance que l'on peut former des couples faisant intervenir tous les points de mesure; au-delà, les points échantillonnés au centre du domaine n'interviennent plus.

2 VARIABILITE EN FONCTION DE L'ECHELLE

Nous allons mesurer la variabilité à différentes échelles tout simplement en calculant une mesure de dissemblance entre deux données z_1 et z_2 situées en deux points x_1 et x_2 . Cette dissemblance entre deux valeurs, désignée par γ , vaudra.

$$\gamma = \frac{(z_1 - z_2)^2}{2}$$

On fait dépendre la dissemblance γ de la distance et de l'orientation d'une paire de points, décrits par le vecteur $h = x_2 - x_1$, indifféremment de la position de la paire dans le domaine étudié.

$$\gamma(h) = \frac{(z(x_1+h) - z(x_1))^2}{2}$$

En formant la moyenne des dissemblances γ entre valeurs pour toutes les nh paires de points reliées par un vecteur h donné pour une maille donnée, on obtient la notion de variogramme expérimental.

$$\gamma^*(h) = \frac{1}{2n_h} \sum_{\alpha=1}^n (z(x_\alpha+h) - z(x_\alpha))^2$$

Généralement la dissemblance des valeurs augmente en moyenne en fonction de l'éloignement spatial des points de mesure et atteint fréquemment un palier de variation à grande échelle. Lorsque la pente du variogramme change abruptement on peut penser à des paliers intermédiaires.

Le comportement aux très petites échelles, près de l'origine du variogramme, est d'une importance capitale, car il est un indicateur du degré de continuité de la variable régionalisée, à savoir: différentiable, continue mais non différentiable, ou carrément discontinue. Dans ce dernier cas on aura affaire à une variable régionalisée donnant lieu à un effet de pépite, symptôme de valeurs changeant abruptement à très petite échelle, comme les teneurs de l'or lorsqu'il y a des pépites.

3 AJUSTEMENT D'UNE FONCTION DE COVARIANCE

Lorsque le variogramme présente un palier aux grandes échelles, on ajustera une fonction de covariance $C(h)$ en se basant sur la formule.

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

où b caractérise un palier de variation.

Dans la suite, le modèle $C(h)$ sera soit un effet de pépite,

$$C_{pep}(h) = \begin{cases} b & \text{lorsque } |h| = 0 \\ 0 & \text{lorsque } |h| > 0 \end{cases}$$

soit un modèle sphérique.

$$C_{sph}(h) = \begin{cases} b(1 - \frac{3|h|}{2a} + \frac{1|h|^3}{2a^3}) & \text{lorsque } 0 \leq |h| \leq a \\ 0 & \text{lorsque } |h| > a \end{cases}$$

Le paramètre a indique la portée de la covariance: la covariance s'éteint lorsque la portée est atteinte. Le paramètre b indique la valeur maximale de la covariance: la covariance sphérique diminue de manière continue à partir de la valeur maximale b , située à l'origine, jusqu'à disparaître complètement lorsque la portée est atteinte. Le modèle de l'effet de pépite pourrait être considéré comme un cas particulier d'un sphérique avec une portée a infiniment petite. Il y a cependant une différence notable entre les deux modèles:

$C_{pep}(h)$ décrit un phénomène discontinu où les valeurs changent brutalement d'un point à un autre, tandis que $C_{sph}(h)$ décrit un phénomène continu, mais non différentiable, qui serait en quelque sorte rugueux au toucher.

4 MODELES EMBOITES

On peut souvent distinguer plusieurs paliers de variation qui sont en rapport étroit avec la morphologie de la variable régionalisée [4].

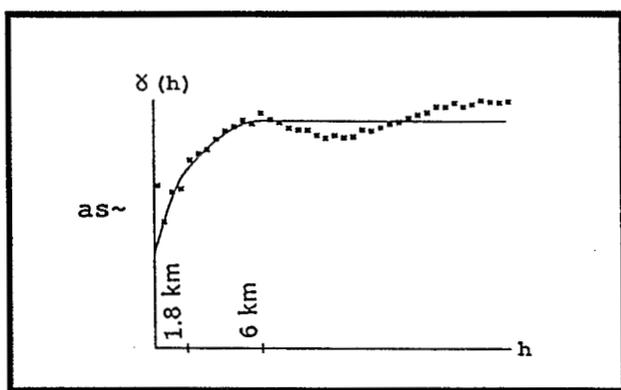


Figure 1: Variogramme expérimental ajusté avec deux modèles sphériques de portées 1.8 et 6 km.

Définissons d'abord une fonction de covariance normée à un palier unitaire.

$$c(h) = \frac{C(h)}{b}$$

c'est-à-dire

$$C(h) = b c(h)$$

On numérotera les paliers de variation b_u observés sur le variogramme avec un index $u = 0 \sim 5$. On peut ainsi construire un *variogramme gigogne*.

$$g(h) = b \cdot \sum_{u=0}^s b_u c_u(h), \text{ où } \sum_{u=0}^s b_u = 1$$

avec une covariance pépitique $C_0(h)$ et plusieurs covariances sphériques $c_u(h)$ de portées a_u différentes. Un exemple d'ajustement d'un modèle de variogramme gigogne par deux modèles sphériques (de portées 1.8 et 6 km) est donné sur la Figure 1.

De même peut-on alors décomposer la variable régionalisée en une somme de composantes spatiales indépendantes agissant à différentes échelles. i.e. atteignant leurs différents paliers de variation b_u à des échelles différentes.

$$z(x) = z_0(x) + \dots + z_u(x) + \dots + z_s(x) + m(x)$$

Il est clair qu'une composante à petite échelle ne pourra être identifiée que si la maille d'échantillonnage est suffisamment fine. De même, une composante à grande échelle n'apparaîtra sur le variogramme que si le rayon du domaine échantillonné est assez important.

On supposera, dans ce modèle, les différentes composantes $z_u(x)$ de moyenne nulle, de sorte que la moyenne de $z(x)$ sera contenue dans la dérive $m(x)$. Cette dérive, comme son nom l'indique, pourra varier lentement d'un bout de la région à l'autre: la fonction $m(x)$ sera néanmoins suffisamment plane pour être assimilée à une constante aux petites et moyennes échelles, et elle sera ainsi systématiquement filtrée par le variogramme expérimental.

5 TRANSFERT D'INFORMATION

Armés d'un modèle caractérisant une variabilité spatiale complexe, identifiée sur un variogramme expérimental, nous allons nous intéresser au problème suivant: ayant une maille d'échantillonnage avec une disposition très irrégulière des échantillons dans l'espace, comment transférer cette information sur un réseau de points

régulier? Surtout, quelles déformations subira l'information contenue dans les données lors du passage de la maille d'échantillonnage à la maille d'estimation?

La méthode examinée ici sera le krigeage ordinaire, et nous allons essayer d'éclaircir son fonctionnement en le comparant au filtrage de composantes spatiales.

6 KRIGEAGE ORDINAIRE

Le krigeage ordinaire est une opération qui est répétée en chaque noeud x_0 d'une grille régulière recouvrant le domaine étudié. Pour un ensemble de n points de données x_i , d'un voisinage centré autour d'un point x_0 de la maille d'estimation, on peut construire, en minimisant la variance d'estimation, le système de *krigeage ordinaire* suivant,

$$\begin{pmatrix} C(x_1-x_1) & \dots & C(x_1-x_n) & 1 \\ \dots & \dots & \dots & \dots \\ C(x_n-x_1) & \dots & C(x_n-x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(x_1-x_0) \\ \dots \\ C(x_n-x_0) \\ 1 \end{pmatrix}$$

où les λ_α sont des pondérateurs à affecter aux points de données et où μ est un paramètre de Lagrange qui intervient pour des raisons algébriques. Le membre gauche du système contient les covariances entre les points de données, tandis que le membre droit contient les covariances entre chaque point de donnée et le point d'estimation x_0 .

Ce système une fois résolu permet de transférer au point x_0 l'information des points de données, par le calcul d'une valeur estimée z^* ,

$$z^* = \sum_{\alpha=1}^n \lambda_\alpha z(x_\alpha)$$

En effectuant la multiplication matricielle, on peut réécrire le système sous la forme.

$$\sum_{\beta=1}^n \lambda_\beta C(x_\alpha-x_\beta) + \mu = C(x_\alpha-x_0)$$

La condition que les poids soient de somme unité exprime le fait que dans le cas extrême où les données sont égales à une constante dans un voisinage donné, la valeur estimée reproduira cette constante.

Nous nous bornerons dans la suite à écrire l'équation générale du krigeage ordinaire.

$$\sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + \mu = \sum_{u=0}^s C_u(x_{\alpha} - x_0)$$

en sous-entendant le système et la condition sur les poids, représentée par le paramètre μ . Dans le membre droit on a écrit de manière explicite un modèle de covariance gégogne.

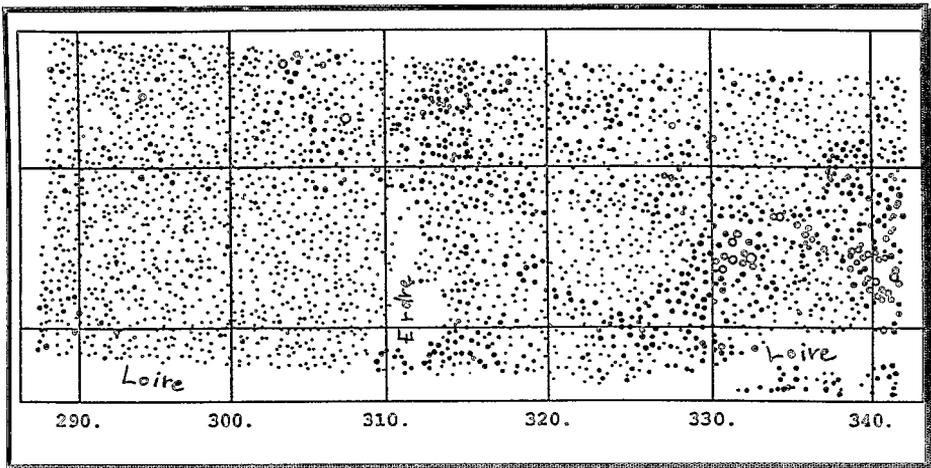


Figure 2: Carte de la position des échantillons représentant les valeurs (en log) de l'Arsenic par des cercles proportionnels (unité: km)

7 FILTRAGE DE COMPOSANTES SPATIALES

On filtre une composante spatiale en retirant les covariances correspondantes $C_u(x_{\alpha} - x_0)$ du membre droit de l'équation (2) du krigeage ordinaire. Ainsi peut-on par exemple éliminer l'effet de pépite en retirant les termes correspondant à $C_0(h)$ du membre droit de (2) et l'on obtient alors une estimation de la variable régionalisée amputée de la composante pépitique $z_0(x)$.

A titre d'exemple, nous présentons deux filtrages réalisés sur des données du BRGM prélevés dans le cadre de l'inventaire géochimique national dans la région délimitée par les cartes géologiques au 1/50000 "Nort-sur-Erdre" et "Ancenis". La Figure 2 représente les valeurs prises par l'arsenic (en log) dans cette région d'une

dimension de 56 x 23 km². On remarquera des valeurs fortes isolées, sur la gauche, qui disparaîtront lors du filtrage, et des groupes de valeurs élevées, au centre et sur la droite, qui, sans être nécessairement constitués de valeurs extrêmes, seront mis en évidence par les filtrages.

Le variogramme expérimental omnidirectionnel de ces données, avec un ajustement par un modèle gigogne composé d'un effet de pépite et de deux sphériques de portées 1.8 et 6 km, a déjà été montré sur la Figure 1.

Le filtrage de la Figure 3 élimine la variation associée à l'effet de pépite et à la composante de courte portée, mettant en évidence les structures à grande échelle, souvent liées à la géologie. Pour extraire ces structures, il est nécessaire d'utiliser un voisinage de krigeage suffisamment grand: ici on a utilisé les 50 points les plus proches [5].

La Figure 4 représente le filtrage de l'effet de pépite et de la composante de longue portée, mettant bien en évidence les structures à caractère plus local, qu'on pourrait qualifier d'anomalies locales, en opposition à des anomalies ponctuelles définies sur des échantillons isolés.

Les trois représentations prises ensemble permettent, associées à la carte géologique, de délimiter dans une prospection les zones les plus favorables, en effectuant un tri entre les anomalies ponctuelles (souvent décevantes) et les anomalies locales [6].

8 KRIGEAGE ORDINAIRE ET FILTRAGE

Prenons une variable régionalisée pour laquelle on aurait identifié trois composantes de variation non corrélées,

$$z(x) = z_0(x) + z_1(x) + z_2(x) + m(x)$$

A cette situation correspond le modèle de covariance.

$$C(h) = C_0(h) + C_1(h) + C_2(h)$$

avec $C_0(h)$ un effet de pépite et $C_1(h)$, $C_2(h)$ des modèles sphériques de portées a_1 et a_2 , numérotés de sorte que $a_1 < a_2$.

Supposons que la maille d'échantillonnage soit très irrégulière entraînant une répartition des données extrêmement inégale dans l'espace. Afin d'établir une carte, on va construire un réseau de points, aussi fin que l'on voudra, et l'on exécutera en chaque point l'opération de krigeage ordinaire.

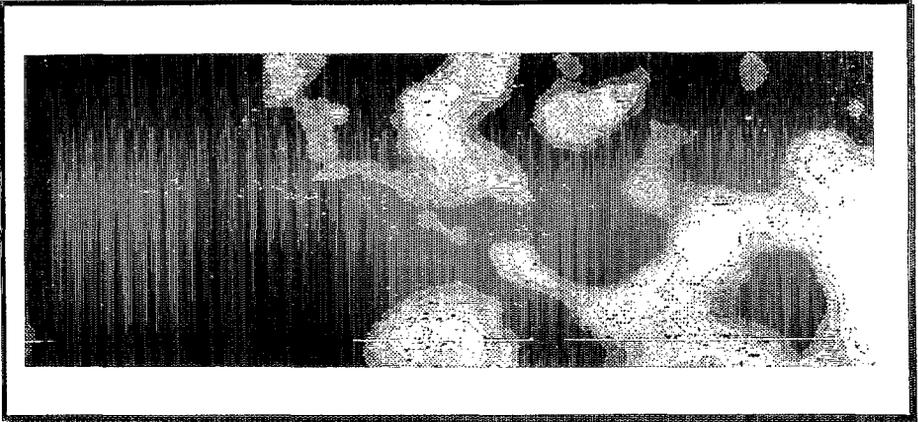


Figure 3: Filtrage des composantes $z_0(x)$ et $z_1(x)$ (voisinage de krigeage: les 50 échantillons les plus proches).

Vu la grande irrégularité dans la disposition des échantillons, différents cas de figure peuvent se présenter, que nous allons examiner séparément dans la suite.

8.1 x_0 est éloigné $> a_2$ des données

Ce cas de figure peut se produire si x_0 se trouve en plein milieu d'un grand trou dans la maille d'échantillonnage. Le krigeage ordinaire est alors équivalent à un *krigeage de la moyenne* dont l'équation générale s'écrit.

$$\sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha} - x_{\beta}) + \mu = 0$$

Le membre droit est nul puisque toutes les composantes s'annulent pour un écartement supérieur à a_2 . L'information transférée sera une estimation de la dérive $m(x)$ dans le voisinage centre sur x_0 .

$$m^*(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha} z(x_{\alpha})$$

On notera bien la différence entre un krigeage de la moyenne et le calcul de la moyenne arithmétique des données dans le voisinage de krigeage: cette dernière ne tient pas compte de la disposition spatiale des échantillons.

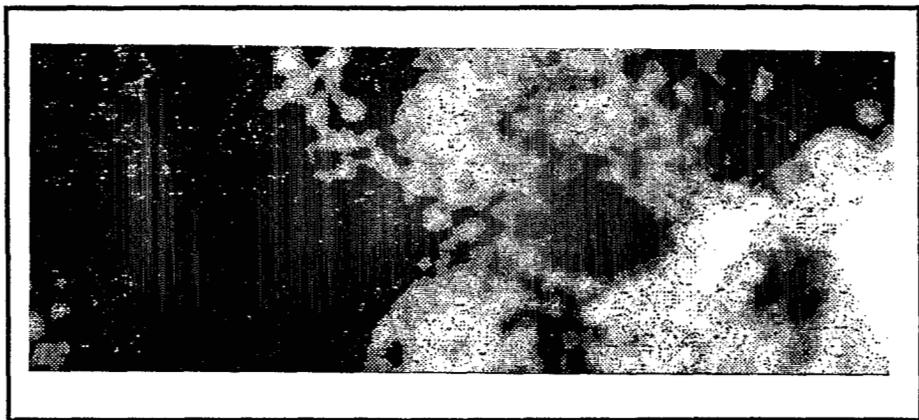


Figure 4: Filtrage des composantes $z_0(x)$ et $Z_2(X)$ (voisinage de krigeage: les 50 échantillons les plus proches).

8.2 x_0 est écarté $< a_2$ et $> a_1$ du point x_α le plus proche

Avec cette disposition spatiale le système de krigeage ordinaire est équivalent à un filtrage des composantes $z_0(x)$ et $z_1(x)$, l'équation générale s'écrivant.

$$\sum_{\beta=1}^n \lambda_\beta C(x_\alpha - x_\beta)^{\mu} = C_2(x_\alpha - x_0)$$

8.3 x_0 est écarté $< a_2$ et > 0 du point x_α le plus proche

Dans cette situation le krigeage ordinaire transférera de l'information non seulement sur la composante de longue portée $z_2(x)$, mais aussi sur la composante $z_1(x)$, qui varie plus rapidement dans l'espace et qui suscitera une description plus détaillée de la variable régionalisée dans les zones de la carte où les données abondent. On devine que le krigeage ordinaire est maintenant équivalent au filtrage.

$$\sum_{\beta=1}^n \lambda_\beta C(x_\alpha - x_\beta)^{\mu} = C_1(x_\alpha - x_0) + C_2(x_\alpha - x_0)$$

8.4 x_0 coïncide avec un point de donnée

Dans ce dernier cas, le krigeage ordinaire ne filtrera rien du tout et restituera fidèlement la donnée mesurée au point $x_\alpha = x_0$: c'est un interpolateur exact !

Cette fidélité peut cependant être indésirable lorsqu'il n'y a que quelques points de coïncidence purement fortuits, entre une maille d'échantillonnage irrégulière et une maille d'estimation régulière. Pour éviter des sauts abrupts des valeurs estimées en ces quelques points, certains programmes de krigeage à vocation cartographique (tel Bluepack) proposeront de filtrer systématiquement la composante pépitique $z_0(x)$.

9 DONNEES MULTIVARIABLES

La généralisation multivariable du modèle de variogramme (1) est immédiate, en écriture matricielle,

$$\Gamma(h) = B \cdot \sum_{u=0}^s B_u c_u(h) \text{ avec } B = \sum_{u=0}^s B_u$$

où les matrices de corégionalisation B_u sont semi-définies positives.

Le *cokrigeage* ordinaire d'une variable $z_{i_0}(x)$, basé sur ce modèle de corégionalisation, a l'équation générale,

$$\sum_{j=1}^N \sum_{\beta=1}^n \lambda_j^\beta C_{ij}^\beta(x_\alpha - x_\beta) + \mu_i = \sum_{u=0}^s b_{i_0}^u c_u(x_\alpha - x_0)$$

avec des pondérateurs de somme unité pour la variable à estimer et de somme nulle pour chacune des variables auxiliaires.

Dans le membre droit du système on voit entrer en scène les coefficients de corégionalisation b_{ij}^u dont l'influence dépendra, comme précédemment, de l'écart entre x_0 et les données du voisinage. L'élément nouveau est le jeu subtil des corrélations qui peuvent être fort différentes aux diverses échelles spatiales.

10 SUPPORT

On ne s'intéressera souvent pas seulement à des valeurs définies sur des points, mais également à des valeurs moyennes sur des éléments de surface (comme les "pixels" de la télédétection) ou de volume V . Un volume $V(x)$ centré au point x s'appelle le *support* de la variable régionalisée $Z_V(x)$, qui peut être comprise comme la régularisée d'une variable $z(x)$ de support ponctuel,

$$z_V(x) = \frac{1}{|V|} \int_{x' \in V} z(x') dx'$$

Le passage d'une variable ponctuelle à une variable régularisée a des effets sur les divers paramètres géostatistiques liés au variogramme et à la loi spatiale, qui sont loin d'être triviaux. Le variogramme, par exemple, subit une transformation non-linéaire: l'effet de pépite s'atténue, le comportement linéaire à l'origine du modèle sphérique se transforme en un comportement parabolique, la portée est en général modifiée et le palier diminué.

11 KRIGEAGE DE BLOCS

Le krigeage ordinaire, à partir d'informations ponctuelles, d'une valeur définie sur un volume V_0 centre sur un noeud x_0 d'une grille d'estimation, possède une équation générale du type,

$$\sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha}, x_{\beta}) + \mu = \sum_{u=0}^s \bar{C}_u(x_{\alpha}, V_0)$$

où les $\bar{C}_u(x, V)$ sont des covariances moyennes point-bloc.

Numériquement, le krigeage d'un bloc peut être rapproché d'un filtrage, puisque le résultat va être plus lisse que le krigeage ordinaire ponctuel correspondant, mais la comparaison s'arrête là. Plus le support augmente, plus les covariances point-bloc sont diluées, et le krigeage ordinaire d'un bloc de très grande taille devient, à la limite, identique à un krigeage de la moyenne [7].

12 VARIABLES SUR SUPPORTS DIFFERENTS

Dans certaines applications, on dispose de plusieurs sources d'informations définies sur des supports différents (par exemple, des

mesures de réflectance de télédétection et des données au sol pédologiques, géochimiques ou biologiques).

Une manière plus générale d'écrire (8) est de l'exprimer sous la forme d'un produit de convolution stochastique de valeurs ponctuelles d'une fonction aléatoire stationnaire d'ordre deux $Z_t(x)$ par une fonction de pondération $p(x)$,

$$Z_{p_t}(x) = \int_{-\infty}^{+\infty} Z_t(x+y)p_t(y)dy = Z^* \check{p}_t$$

où $\check{p}_t = p(-x)$ est la transposée de la fonction de pondération.

La covariance croisée régularisée s'écrit ensuite,

$$C_{ij}^-(h) = C_{ij} * \check{p}_i * p_j$$

Les variogrammes croisés régularisés peuvent être posés de la même manière,

$$\overline{\gamma}_{ij}(h) = \gamma_{ij} * \check{p}_i * p_j - A_{ij}$$

avec des constantes A_{ij} garantissant que $\overline{\gamma}_{ij}(0) = 0$.

Par le choix de fonctions de pondération appropriées, on peut construire des modèles de corégionalisation adaptés à la physique des variables considérées [1][2][8][9].

CONCLUSION

Nous nous sommes attachés, dans la première partie de ce bref exposé, à montrer de quelle manière le krigeage ordinaire tient compte des différentes composantes identifiées dans une analyse de la variabilité spatiale.

Nous n'avons fait qu'effleurer le problème du changement de support dans un contexte de géostatistique linéaire. Dans maintes situations, en particulier lorsqu'on s'intéresse à la proportion d'une région (éventuellement subdivisée en des unités d'une taille différente du support des échantillons) pour laquelle la variable régionalisée dépasse un certain seuil critique, il est plus avantageux

Le domaine des méthodes applicables à des données multivariées mesurées sur des supports différents est, quant à lui, encore peu exploré.

REFERENCES

- [1] MATHERON G (1965) *Les Variables Régionalisées et leur Estimation*. Masson, Paris.
- [2] MATHERON G (1970) *La Théorie des Variables Régionalisées et ses Applications*. Fascicule 5, Les Cahiers du Centre de Morphologie Mathématique, Ecole des Mines de Paris, Fontainebleau.
- [3] MATHERON G (1989) *Estimating and Choosing*. Springer-Verlag, Berlin .
- [4] SERRA J (1968) *Les structures gigognes: morphologie mathématique et interprétation métallogénique*. Mineral. Deposita, 3, 135-154.
- [5] GALLI A, GERDIL-NEUILLET F & DADOU C (1984) *Factorial kriging analysis: a substitute to spectral analysis of magnetic data*. In: Verly G et al (eds) *Geostatistics for Natural Resources Characterization*, 543-557, NATO ASI Series C-122, Reidel, Dordrecht.
- [6] WACKERNAGEL H & BUTENUTH C (1989) *Caractérisation d'anomalies géochimiques par la géostatistique multivariée*. J. Geochem. Explor., 32, 437-444.
- [7] RIVOIRARD J (1984) *Le Comportement du Poids de Krigeage*. Thèse de Docteur-Ingénieur, Ecole des Mines de Paris.
- [8] HUIJBREGTS C (1971) *Reconstitution du variogramme ponctuel à partir d'un variogramme régularisé*. N-244, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- [9] SEGURET S & CHAUVET P (1988) *Pour une méthodologie de déconvolution de variogrammes*. N-51/88/G, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- [10] LANTUEJOUL C (1990) *Cours de Sélectivité*. C-140, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.
- [11] RIVOIRARD J (1990) *Introduction au Krigeage Disjonctif et à la Géostatistique Non-Linéaire*. C-139, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.