

Caractérisation d'anomalies géochimiques par la géostatistique multivariable

HANS WACKERNAGEL¹ et CHRISTINE BUTENUTH²

¹*Centre de Géostatistique, ENSMP, 35 rue Saint Honoré, 77305 Fontainebleau, France*

²*Geologisches Institut, RWTH, Wuellnerstrasse 2, 5100 Aachen, F.R.Germany*

(Revised and accepted for publication September 23, 1988)

ABSTRACT

Wackernagel, H. and Butenuth, C., 1989. Caractérisation d'anomalies géochimiques par la géostatistique multivariable [Characterization of geochemical anomalies by multivariate statistics]. In: S.E. Jenness et al. (Editors), *Geochemical Exploration 1987. J. Geochem. Explor.*, 32: 437-444.

The behavior of geochemical elements is different inside and outside a mineral deposit. The correlations which can be measured between geochemical elements depend on the spatial scale at which a data analysis is performed. Multivariate geostatistics conceptualize this observation by defining various spatial categories on the basis of a variogram model, which represent a classification of regional, local and pointwise spatial scales. Factor analysis is then performed using correlation matrices associated to the spatial categories. The scores of the spatial factors for each sample are estimated by cokriging. Distinguishing between pointwise and local anomalies, it can be seen that geostatistics provide concepts for emphasizing local anomalies which are often more difficult to identify than pointwise anomalies.

RESUME

Le comportement d'éléments géochimiques est différent à l'intérieur et à l'extérieur d'un gîte métallifère. Les corrélations mesurées entre des éléments géochimiques dépendent de l'échelle spatiale à laquelle est pratiquée l'analyse de données. La géostatistique multivariable conceptualise cette observation en définissant des catégories spatiales regroupant des échelles spatiales régionales, locales, ponctuelles, sur la base d'un modèle de variogramme. Une analyse factorielle est pratiquée à partir des matrices de corrélation associées à ces catégories spatiales et les scores des facteurs en résultant sont estimés par cokrigage pour chaque échantillon. En distinguant les anomalies ponctuelles et les anomalies locales, on voit que la géostatistique est en mesure de fournir un appareil conceptuel propre à mettre en évidence ces dernières, qui sont souvent plus difficiles à repérer que les premières.

INTRODUCTION

Le statisticien et le géologue ont généralement des attitudes diamétralement opposées face à des échantillons géochimiques.

Le statisticien considère un tout (l'ensemble des échantillons) qu'il modélise globalement. Ce n'est qu'ensuite qu'il en arrive à s'intéresser aux individus (les échantillons) pour essayer de les caractériser au moyen du modèle global. Pour le géologue, surtout s'il travaille à l'échelle du kilomètre, chaque échantillon sort d'un contexte particulier et il éprouve une certaine réticence à les regrouper d'emblée en un seul ensemble.

La géostatistique occupe une position intermédiaire vis-à-vis de ces deux positions par le concept probabiliste qu'elle utilise. Alors que le statisticien considère les échantillons comme des réalisations indépendantes d'une seule variable aléatoire, la géostatistique propose de construire une variable aléatoire pour chaque échantillon. Ces variables aléatoires sont ensuite regroupées en une famille, appelée fonction aléatoire. L'opération intellectuelle qui permet ce regroupement est la formulation d'une hypothèse de stationnarité, fondée sur une analyse spatiale préalable des données. La géostatistique fait par ailleurs la concession au géologue que chaque échantillon appartient à un contexte particulier donné par sa position géographique.

Les méthodes géostatistiques sont tout d'abord des méthodes à caractère local, faisant en général uniquement appel à des propriétés locales de la fonction aléatoire. Nous allons essayer de montrer quelle contribution elles sont susceptibles d'apporter au problème de l'interprétation et de la détection d'anomalies géochimiques.

L'idée principale que nous essayerons de mettre en valeur est que la structure de corrélation d'un jeu de données multiéléments n'est pas unique comme le présuppose l'approche globale du statisticien. En effet, la géostatistique permet de distinguer différentes catégories d'échelles spatiales et d'associer à chaque catégorie spatiale une structure de corrélation particulière. En cela elle se rapproche par exemple de l'observation du géologue que les relations entre des éléments chimiques sont différentes à l'intérieur et à l'extérieur d'un amas de substances minérales.

La distinction de plusieurs structures de corrélation permet de caractériser séparément des anomalies de type ponctuel ou local, suivant que ces anomalies sont anormales par rapport à un voisinage local ou que ce voisinage lui-même sort de la norme fixée par le géochimiste.

Nous ferons essentiellement référence ici à des données géochimiques d'exploration tactique, c'est-à-dire prélevées à l'échelle de l'hectomètre sur des régions d'une étendue de plusieurs dizaines de kilomètres carrés. Néanmoins ces réflexions pourront aisément être transposées à des échelles d'observation de taille inférieure.

APPROCHE METHODOLOGIQUE

Décomposition du modèle global

Une approche globale (statistique) d'un jeu de données multiéléments con-

siste à calculer des variances et des covariances entre les éléments, considérés comme des réalisations de variables aléatoires. Ces valeurs expérimentales sont rangées sous la forme d'une matrice de variance-covariance V . S'il s'agit de variables centrées et normées la matrice V est identique à la matrice des corrélations R .

Recherchant une approche plus locale, on peut se poser la question de l'existence d'une décomposition de la matrice globale de variance-covariance V en plusieurs matrices de variances-covariances B_u (avec $u=0, \dots, S$), appelées matrices de corégionalisation, avec:

$$V = B_0 + B_1 + \dots + B_S \quad (1)$$

Dans ce modèle on considère que les relations globales entre les variables décrites par V sont en fait la somme de différentes interactions décrites par les matrices B . L'idée sous-jacente est naturellement d'associer les matrices de corégionalisation B à différentes composantes du signal géochimique. Sans perte de généralité, et afin de simplifier l'exposé, nous allons considérer dans la suite un signal géochimique à trois composantes uniquement:

(a) une composante "ponctuelle", qui est la partie du signal correspondant à des échelles spatiales inférieures à celle de la maille d'échantillonnage. La matrice de corégionalisation B_0 représentera cette catégorie spatiale.

(b) une composante "locale", qui décrit des échelles spatiales de petite taille et dont l'ampleur correspond dans certains cas au diamètre moyen d'une aurole de dispersion géochimique autour d'un amas de substances polymétalliques. Nous associerons la matrice B_1 à cette catégorie locale.

(c) une composante "régionale", qui traduit le comportement du signal géochimique à grande échelle. Une matrice B_2 décrira les relations entre les variables à l'échelle régionale.

Ayant attribué un sens physique au modèle (1), nous allons nous pencher sur la façon de déterminer une telle décomposition, ce qui nous fera déboucher sur un autre modèle, bien plus général.

Distinction de catégories spatiales

Pour procéder à une analyse spatiale multivariable, il nous faut une description du comportement spatial moyen d'une variable ou d'un couple de variables.

On s'intéressera d'abord aux positions des échantillons les uns par rapport aux autres. Construisons pour cela un vecteur h reliant deux échantillons. Ce vecteur situe l'échelle spatiale qui sépare ces échantillons.

La différence des valeurs des variables mesurées sur ces échantillons traduit la dissimilarité des deux échantillons à l'échelle h . On peut de la sorte, en formant des produits de différences de valeurs pour des paires d'échantillons, obtenir une description du comportement spatial moyen d'une variable ou d'un couple de variables.

La fonction $\gamma(h)$, appelée variogramme, est le résultat de cette analyse du comportement spatial moyen en fonction de l'échelle spatiale h .

Le variogramme peut souvent être découpé au moyen de plusieurs fonctions $g(h)$, ayant un palier unité lorsque la distance définie par h est supérieure à une valeur appelée la portée spatiale de $g(h)$. Pour traduire l'importance d'une fonction $g(h)$ dans le variogramme, un coefficient b lui est affecté. Parfois on considère également des fonctions $g(h)$ sans palier.

Ces considérations nous permettent de construire un modèle de variogramme dans lequel chaque fonction $g(h)$ définit une catégorie spatiale dont l'importance est donnée par un coefficient b :

$$\gamma(h) = b_0 g_0(h) + b_1 g_1(h) + \dots + b_S g_S(h) \quad (2)$$

Etant donné que le variogramme décrit, soit le comportement d'une variable (variogramme simple), soit le comportement conjoint d'un couple de variables (variogramme croisé), la généralisation multivariable du modèle (2) est immédiate:

$$\Gamma(h) = B_0 g_0(h) + B_1 g_1(h) + \dots + B_S g_S(h) \quad (3)$$

Les matrices de corégionalisation B donnent une description de la structure multivariable au niveau de différentes catégories spatiales (ponctuelles, locales, régionales, ...).

Le modèle de corégionalisation (3) nous livre un outil très général permettant de distinguer plusieurs catégories spatiales au niveau desquelles les variables se comportent différemment. En particulier, on pourra identifier à l'aide des matrices de corégionalisation B des structures de corrélation propres à des catégories ponctuelles, locales, régionales.

Sachant que, lorsque les moyennes et les variances sont strictement stationnaires dans la région considérée, la matrice V est la limite (pour h grand) de la matrice $\Gamma(h)$, le modèle (1) se révèle être un cas limite du modèle (3). Nous avons donc trouvé une décomposition du modèle global (1) qui nous mène hors du contexte particulier de la stationnarité stricte dans lequel ce dernier est valable.

Le modèle de corégionalisation (3) est défini dans le cadre peu contraignant d'une hypothèse de stationnarité intrinsèque (Matheron, 1970). Si, en outre, on décrit le comportement spatial des variables au moyen de covariances généralisées (Matheron, 1973) au lieu des variogrammes utilisés ici, le modèle (3) peut être formulé dans le cadre d'une hypothèse de stationnarité intrinsèque généralisée (Matheron, 1982). L'inférence de catégories spatiales à partir de covariances généralisées n'a cependant pas encore été étudiée dans un cadre multivariable, ce qui explique la formulation un peu moins générale adoptée ici, qui s'avère toutefois suffisamment flexible dans bien des cas.

EXEMPLE D'APPLICATION

Les teneurs des éléments cuivre (Cu), plomb (Pb) et zinc (Zn) ont été mesurées sur 2049 échantillons prélevés au nord de la ville de Brilon (R.F.A.) au cours d'une campagne de prospection tactique.

Nous reproduisons sur la Figure 1 les six variogrammes simples et croisés des trois variables Cu, Pb, Zn.

Le modèle (courbe) ajusté aux valeurs (points) calculées à partir des données expérimentales permet de distinguer trois catégories spatiales:

(a) une catégorie "régionale", qui résume le comportement spatial des variables à des échelles variant entre 130 m et 2300 m,

(b) une catégorie "locale", qui regroupe les échelles spatiales entre 10 m et 130 m,

(c) une catégorie "ponctuelle", qui prend en compte les variations à des échelles inférieures à la maille d'échantillonnage.

Le Tableau 1 montre les coefficients de corrélation correspondant à ces trois catégories spatiales, ainsi que les coefficients usuels de la statistique (coefficients globaux).

Le Cu et le Pb ne sont pas corrélés au niveau ponctuel et il semble y avoir une faible tendance régionale à ce que ces éléments n'aillent pas de pair. On peut penser que dans certaines zones la présence d'un élément est plus forte, alors que dans d'autres zones la présence de l'autre élément est plus accentuée.

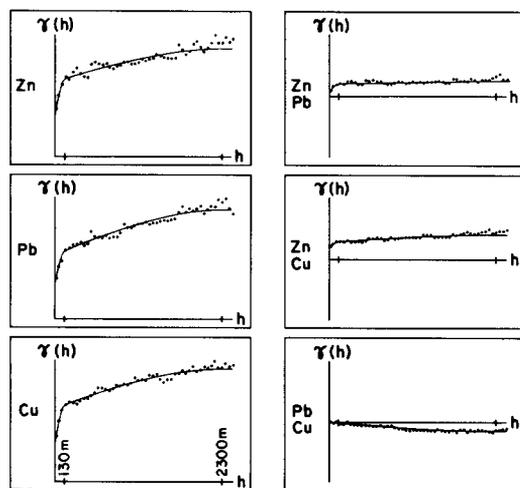


Fig. 1. Données géochimiques de Brilon: variogrammes simples et croisés; h : distance, en m; $\gamma(h)$: valeur du variogramme.

Fig. 1. Brilon geochemical data: single and cross-variograms; h : distance, in m; $\gamma(h)$: value of the variogram.

TABLEAU 1

Coefficients de corrélation statistiques et géostatistiques.

TABLE 1

Statistical and geostatistical correlation coefficients.

	Corrélation statistique	Corrélation ponctuelle (0-10 m)	Corrélation locale (10-130 m)	Corrélation regionale (130-2300 m)
Cu-Pb	-0.08	0.0	-0.04	-0.36
Cu-Zn	0.42	0.57	0.31	0.42
Pb-Zn	0.35	0.23	0.46	0.11

La corrélation entre le Cu et le Zn est la plus forte dans la catégorie ponctuelle. Au vu des données, cela s'explique par le fait que le Zn adopte généralement une valeur très faible, sauf pour quelques échantillons isolés pour lesquels le Cu prend également des valeurs plus fortes que la moyenne. La corrélation des deux éléments, plus faible dans la catégorie locale, s'accroît à nouveau à une échelle régionale parce que les échantillons isolés au niveau local se trouvent réunis à grande échelle (U. Siewers, pers. commun., 1986).

Le couple de variables Pb et Zn est surtout corrélé à une échelle locale, ce qui pourrait donner à penser qu'il s'agit d'un comportement opposé à celui du couple Cu et Zn. Les valeurs fortes de Pb ne sont pas situées aux mêmes endroits que celles du Zn, mais cependant dans le voisinage de celles-ci.

Cette interprétation brosse le tableau d'anomalies locales en Pb et Zn pour lesquelles le Cu est associé ponctuellement au Zn, mais pas au Pb.

ESTIMATION DES SCORES DE FACTEURS SPATIAUX

On pratique souvent une analyse factorielle sur la base de la matrice de variance-covariance V ou sur la matrice de corrélation R . La même opération est possible à partir des matrices de corégionalisation B , et l'on obtient alors des ensembles de facteurs pour chaque catégorie spatiale.

L'intérêt d'une telle démarche est que l'on peut distinguer des facteurs ponctuels, locaux ou régionaux, ce qui permet une analyse bien plus fine que l'analyse de données conventionnelle qui ne dispose que de facteurs globaux.

Les scores de différents facteurs spatiaux à l'endroit d'une anomalie potentielle sont obtenus par cokrigeage, c'est-à-dire en minimisant la variance d'une estimation effectuée à l'aide de données situées dans le voisinage. A l'aide de ces scores le géochimiste pourra distinguer différents types d'anomalies, tels que des anomalies ponctuelles et des anomalies locales.

Nous avons essayé de représenter schématiquement, sur les Figures 2 et 3,

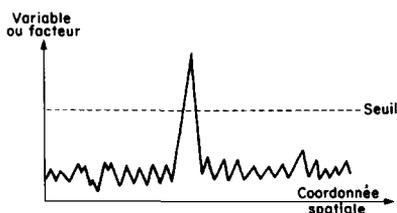


Fig. 2. Schéma d'une anomalie ponctuelle.

Fig. 2. Example of a pointwise anomaly.

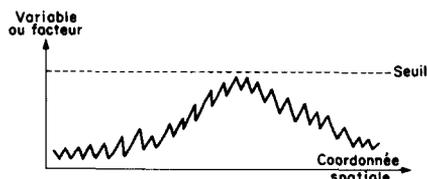


Fig. 3. Schéma d'une anomalie locale.

Fig. 3. Example of a local anomaly.

les deux situations correspondant à des deux types d'anomalies. Sur les abscisses, nous avons reporté les coordonnées d'échantillons le long d'une traversée. Les ordonnées des graphes représentent, sans les spécifier, les valeurs d'une variable ou d'un facteur, selon que le caractère anormal d'un échantillon est repéré par un seuillage pratiqué sur une variable ou sur un facteur.

La Figure 2 montre le cas de l'anomalie ponctuelle, que nous définissons comme étant une anomalie isolée dans un fond régional non perturbé.

La Figure 3 illustre le cas de l'anomalie locale, c'est-à-dire une anomalie pour laquelle le fond régional est lui-même sensiblement au-dessus de la moyenne dans le voisinage de l'échantillon anormal. Ce type d'anomalie peut se situer dans le rayon d'influence d'une auréole de dispersion géochimique autour d'un amas. L'extension de cette auréole sera suffisamment grande pour perturber significativement plusieurs échantillons dans un voisinage local.

L'anomalie locale, qui pourrait par exemple être située juste en dessous du seuil, peut facilement passer inaperçue, puisqu'elle n'a rien d'extraordinaire par rapport à son voisinage immédiat. Nous avançons que la méthode géostatistique permet, en opposant des facteurs locaux et régionaux, de mettre en évidence ce genre d'anomalie. La condition à laquelle cela peut marcher est que la catégorie locale utilisée est, de fait, à peu près de la taille de l'auréole géochimique entourant l'anomalie locale.

CONCLUSION

Le but de cet article est d'attirer l'attention du géochimiste sur un ensemble de méthodes géostatistiques multivariées pour l'analyse et l'interprétation de données d'exploration géochimique. Les deux idées qui prévalent sont que d'une part, la structure de corrélation globale d'un jeu de données peut être subdivisée en plusieurs structures de corrélations régionales, locales, ponctuelles, qui permettent d'analyser les relations entre des éléments géochi-

miques en fonction de différentes catégories spatiales. D'autre part, l'estimation de facteurs spatiaux à l'endroit de chaque échantillon devrait aider à mettre en évidence des anomalies locales, qui sont en général plus difficiles à repérer que des anomalies ponctuelles.

On trouvera la formulation mathématique du modèle géostatistique mono-variable dans Sandjivy (1983) et multivariable dans Matheron (1982), Wackernagel (1988).

REMERCIEMENTS

Messieurs F. Bender et W. Stahl, Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover, ont eu l'amabilité de mettre à disposition des données rassemblées pendant le projet "Rhenoharzynikum".

BIBLIOGRAPHIE

- Granier, Cl., 1973. Introduction à la Prospection Géochimique des Gîtes Métallifères. Masson, Paris, 143 pp.
- Matheron, G., 1970. La Théorie des Variables Régionalisées et ses Applications. Les Cahiers du Centre de Géostatistique, 5, 211 pp.
- Matheron, G., 1973. The intrinsic random functions and their applications. *Adv. Appl. Prob.*, 5: 439-468.
- Matheron, G., 1982. Pour une analyse krigeante de données régionalisées. Centre de Géostatistique, note N-732, 22 pp.
- Sandjivy, L., 1983. Analyse krigeante de données géochimiques. *Sci. Terre, Sér. Inf.*, 18: 141-172.
- Wackernagel, H., 1988. Geostatistical techniques for interpreting multivariate spatial information. In: C.F. Chung et al. (Editors), *Quantitative Analysis of Mineral and Energy Resources*. Reidel, Dordrecht, pp. 393-409.
- Wackernagel, H., Webster, R. et Oliver, M.A., 1988. A geostatistical method for segmenting multivariate sequences of soil data. In: H.H. Bock (Editor), *Classification and Related Methods of Data Analysis*. North-Holland, Amsterdam, pp. 641-650.