

Contribution de la géostatistique au traitement de l'information environnementale & géographique.

C. de FOUQUET
ECOLE DES MINES DE PARIS – CENTRE DE GÉOSTATISTIQUE
35, rue Saint-Honoré
77305 Fontainebleau

Résumé

Les méthodes géostatistiques d'estimation sont fondées sur la modélisation de la structure spatiale à partir des données expérimentales. Les paramètres de l'estimation dépendent de la variabilité spatiale. La précision de l'estimation est indiquée par la variance d'estimation.

Les techniques multivariées permettent d'améliorer l'estimation par l'utilisation d'information auxiliaire (carte, images satellitaires par exemple).

1. Cartographie et Estimation

L'établissement d'une carte relative à l'état de l'environnement à partir de données expérimentales, des mesures au sol par exemple, s'effectue habituellement à l'aide d'un algorithme de calcul fixé a priori (Figure 1). La carte demandée est une grille dont les noeuds sont à informer, ou un tracé de courbes d'isovaleur, ou encore un découpage en unités administratives préétablies auxquelles une valeur est affectée. Les méthodes de calcul usuelles utilisent parfois des ajustements par des splines mais se réduisent souvent à des algorithmes frustes : polygones d'influence, inverse des distances ou du carré des distances.

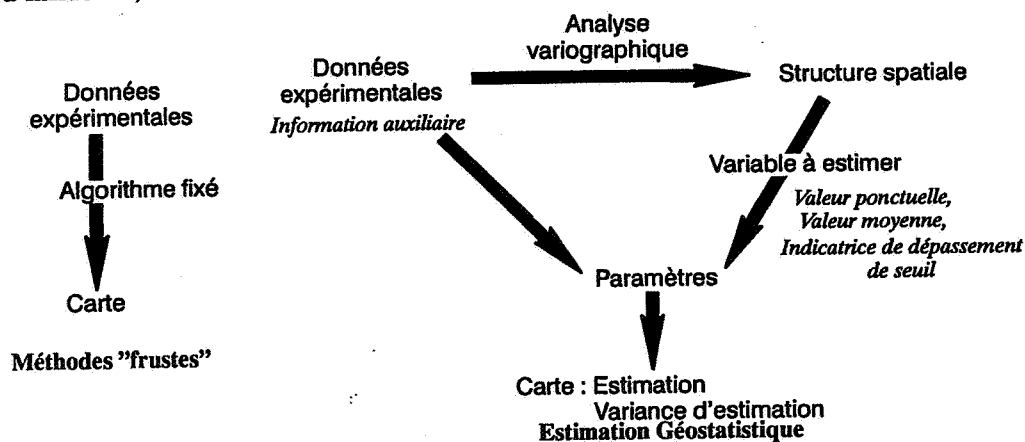


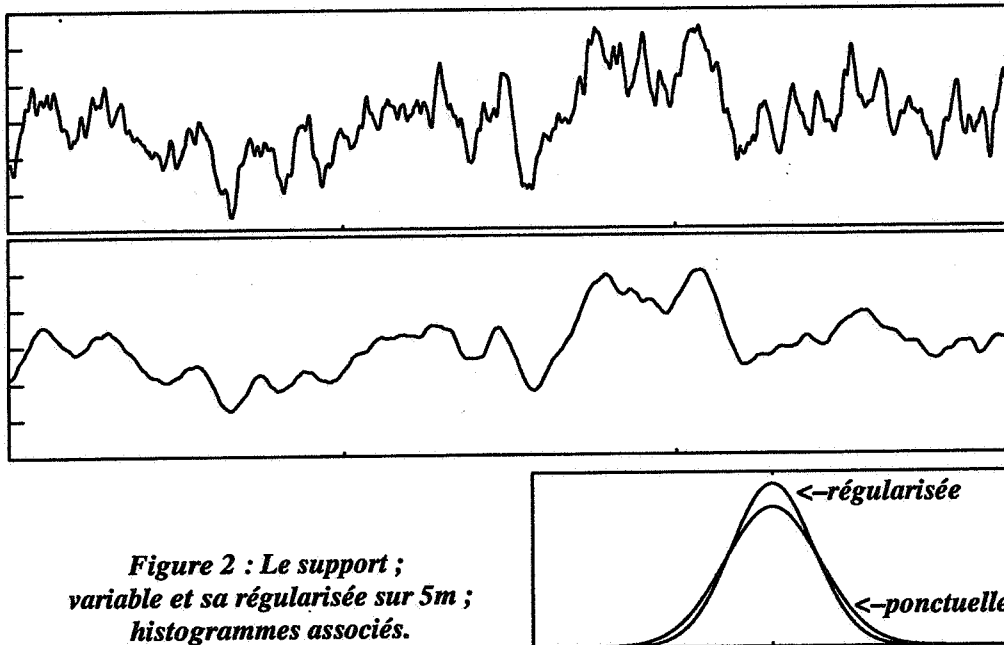
Figure 1 : Etapes schématiques de la cartographie d'une variable environnementale.

L'estimation géostatistique nécessite une étape intermédiaire de reconnaissance de la structure spatiale de la variable. L'inférence de cette structure se fait par l'examen de la variabilité des valeurs expérimentales en fonction de la localisation des données. Une fois synthétisée sous forme adaptée –le variogramme dans le cas d'une estimation linéaire– cette structure spatiale détermine les paramètres de l'estimateur utilisé pour la cartographie, en réponse à un problème donné. L'information contenue dans les données expérimentales intervient doublement : par la valeur numérique aux points expérimentaux, comme pour les méthodes usuelles ; via les paramètres de l'algorithme de calcul, fixés par la structure spatiale. La résolution effective d'un

krigeage suppose que soit précisée la variable à estimer: valeur ponctuelle ou valeur moyenne sur un certain support, indicatrice de dépassement d'un seuil fixé, par exemple. La carte des valeurs estimées pour ces différentes variables sont analogues mais non identiques. La variance de krigeage indique la précision de l'estimation.

2. Le support

Les caractéristiques statistiques et la structure spatiale d'une variable z du point x (concentration ou teneur, nombre surfacique ou volumique d'unités par exemple) dépendent du support, c'est-à-dire de la géométrie du volume ou de la surface élémentaire, sur lequel la variable est considérée (Figure 2) La concentration moyenne z_v calculée sur un segment ou sur un volume, un "bloc", est égale à la moyenne de la concentration ponctuelle $z(x)$ des points du bloc : $z_v = \frac{1}{V} \int_V z(x) dx$. L'histogramme de la concentration moyenne de z sur les blocs v présente moins de valeurs extrêmes, fortes ou faibles, et est plus resserré autour de la moyenne m , que l'histogramme de la concentration "ponctuelle" $z(x)$. La variance de la concentration moyenne par bloc est inférieure à celle des échantillons ponctuels $z(x)$.



**Figure 2 : Le support ;
variable et sa régularisée sur 5m ;
histogrammes associés.**

L'histogramme indique la proportion des valeurs supérieures à un seuil, correspondant par exemple à une norme. Cette proportion dépend du support; ainsi, même dans le cas d'une information exhaustive, la délimitation d'une zone polluée diffère selon que l'on considère les valeurs "ponctuelles" représentées par les échantillons ou les valeurs moyennes sur des blocs v , où v est le volume pertinent en fonction des objectifs et de la méthode d'assainissement employée (Figure 3). Schématiquement, dans le cas d'un polluant atmosphérique, la concentration ponctuelle (concentration instantanée, concentration horaire moyenne) doit rester inférieure à un premier seuil élevé s_1 , afin d'éviter des réactions immédiates, voire l'asphyxie des populations; pour limiter les effets cumulatifs à long terme, la concentration mensuelle (annuelle ?) doit rester

inférieure à un seuil s_2 inférieur à s_1 –le seuil s_2 pouvant être franchi par la concentration ponctuelle.

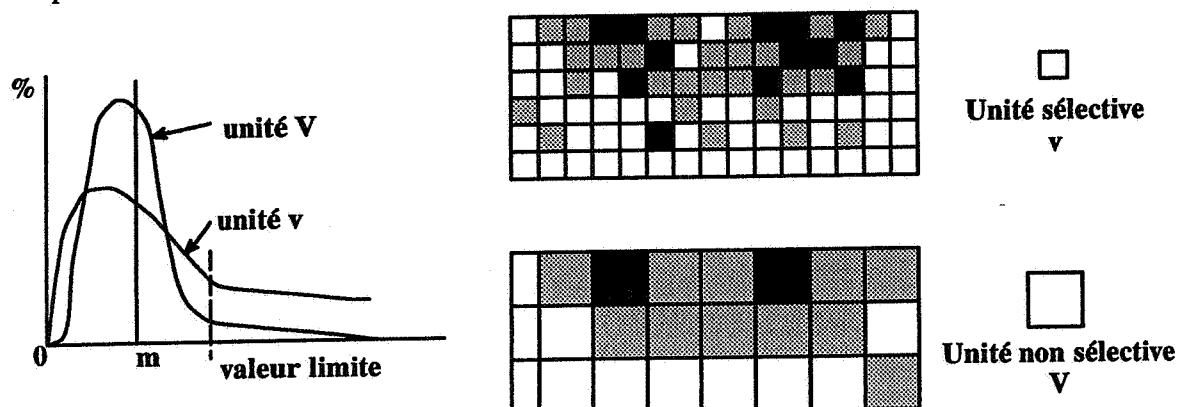


Figure 3 : Volume pollué en fonction du support

L'établissement d'une carte de risque de dépassement d'une valeur limite dans le cas où le support pertinent diffère du support des mesures expérimentales fait appel à des techniques élaborées de changement de support et met en oeuvre des méthodes d'estimation non linéaire. On prendra garde que, avec ou sans changement de support, cartographier le risque par comparaison des valeurs estimées linéairement (krigeage ou estimateurs frustes) au seuil limite conduirait à des biais qui peuvent être importants.

Dans la suite, on se limite aux interpolateurs linéaires obtenus par combinaison linéaire des valeurs aux points expérimentaux.

3. La structure spatiale : le variogramme et ses applications; le krigage.

a. Le variogramme

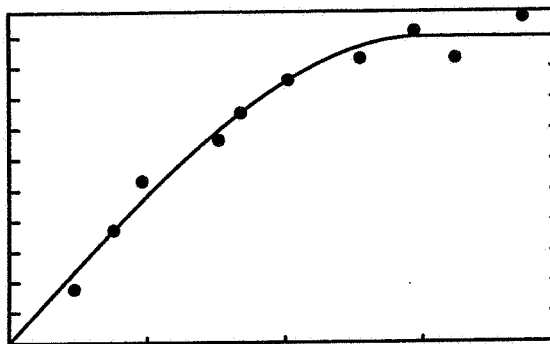
Le variogramme expérimental de la variable z est calculé à partir des données expérimentales x_α par la relation

$$\gamma_{\text{exp}}(h) = \frac{1}{2N(h)} \sum_{\alpha} (z(x_\alpha + h) - z(x_\alpha))^2$$

où $N(h)$ désigne le nombre de couples de points expérimentaux distants de h . Le variogramme expérimental mesure l'écart quadratique moyen en fonction de la distance : à une faible variabilité de z correspond une faible valeur du variogramme; à une plus forte variabilité de z avec la distance correspond une valeur plus forte du variogramme. Le variogramme expérimental décrit l'augmentation de la variabilité avec la distance entre deux valeurs. Les variogrammes expérimentaux, calculés pour différentes orientations de h (Nord-Sud, Est-Ouest et les bissectrices par exemple) permettent de déceler les anisotropies, c'est-à-dire la différence de structure spatiale selon la direction –par exemple, direction de plus forte ou plus faible variabilité suivant le relief.

Pour effectuer les calculs d'estimation il est nécessaire d'ajuster le variogramme expérimental par une fonction variogramme $\gamma(h)$ (Figure 4). La régularité du variogramme en $h=0$ décrit la régularité de la variable. Le comportement du variogramme aux grandes distances décrit la stationnarité ou l'absence de stationnarité de la variable sur le champ.

Figure 4 : Variogramme expérimental et son ajustement



Le modèle de variogramme ajusté synthétise les caractéristiques de la structure spatiale nécessaires pour l'estimation.

b. le krigeage

Le krigeage z^* de z en un point x ou sur un volume v est une combinaison linéaire des valeurs expérimentales $z(x_\alpha)$:

$$z_v^* = \sum_{\alpha} l_{\alpha} z(x_{\alpha})$$

Les poids de krigeage l_{α} , relatifs au bloc v à estimer, sont calculés à l'aide du variogramme et déterminés par les conditions suivantes :

- absence de biais : en moyenne, l'erreur d'estimation est nulle.
- estimation optimale, au sens où la variance de l'erreur d'estimation est minimale.

Parmi la classe des estimateurs linéaires, le krigeage assure la meilleure précision. La précision est donnée par la variance de krigeage.

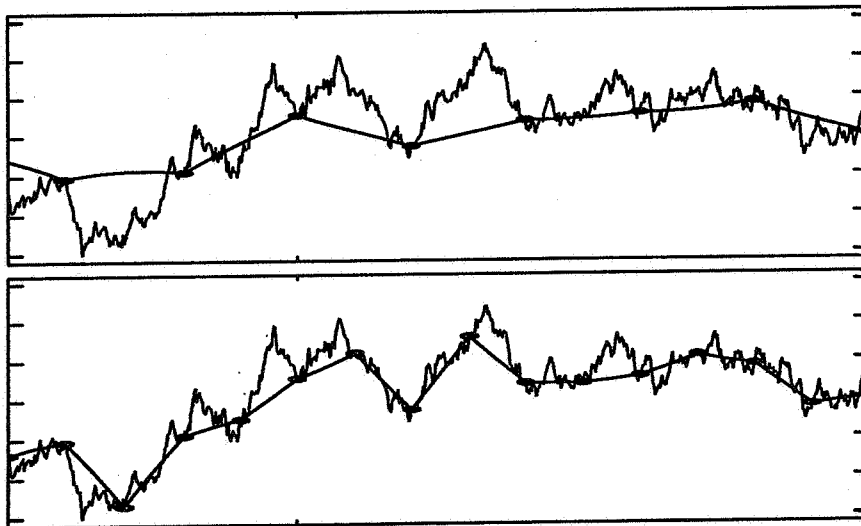


Figure 5 : Krigeage ponctuel : influence de la densité d'information

Les poids de krigeage s'obtiennent par simple résolution d'un système linéaire, dont les coefficients sont les valeurs du variogramme entre les points de données. Les poids dépendent de la valeur à estimer, valeur ponctuelle en x ou valeur moyenne sur un bloc v , par le second membre, égal à la valeur moyenne du variogramme entre les points de données et le point x ou le bloc v à estimer :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \gamma(x_{\alpha} - x_{\beta}) + \mu = \gamma(x_{\alpha}, v) & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases}$$

Le krigeage est un interpolateur exact : le krigeage ponctuel redonne les valeurs expérimentales aux points de données (Figure 5).

c. Optimisation d'une maille de reconnaissance

La variance de krigeage ne dépend pas de la valeur prise par la variable sur les données. Après un premier échantillonnage permettant l'inférence du variogramme, l'examen de la variance de krigeage pour différentes configurations d'échantillonnage permet de retenir le meilleur schéma en fonction de la variable à estimer et du budget disponible pour l'acquisition des données.

d. Information auxiliaire

L'estimation et la cartographie peuvent être améliorées par l'utilisation d'une information auxiliaire. Le cokrigeage est une estimation linéaire de z par une combinaison linéaire des données expérimentales $z(x_{\alpha})$ de la variable d'intérêt et d'une ou plusieurs variables auxiliaires $v_i(x_{\beta})$:

$$z_v^* = \sum_{\alpha} l_{\alpha} z(x_{\alpha}) + \sum_{\beta_i} \lambda_{\beta_i} v_i(x_{\beta_i})$$

Le cokrigeage nécessite l'inférence et la modélisation de la famille des variogrammes simples et croisés de toutes les variables. Lorsque cette inférence n'est pas possible, par exemple si les variables ne sont pas connues deux-à-deux en un nombre suffisant de points communs, une technique comme la dérive externe permet d'utiliser une ou plusieurs variables auxiliaires comme "fond de carte" de z^* .

L'information auxiliaire consiste en une variable échantillonnée aux mêmes points ou en d'autres points que la variable d'intérêt z ; il peut également s'agir d'une image satellitaire ou d'une carte d'une variable liée à z , par exemple la carte topographique. Des méthodes de type régression sont utilisables pour convertir une information de nature qualitative en information quantitative auxiliaire.

La géostatistique participe ainsi au traitement de l'information environnementale

- par l'utilisation d'une information auxiliaire, fournie par exemple par un système d'information géographique,
- par le calcul d'estimations à partir des données expérimentales, visualisables sous forme de cartes, utilisables par les SIG.

4. Un exemple en inventaire forestier

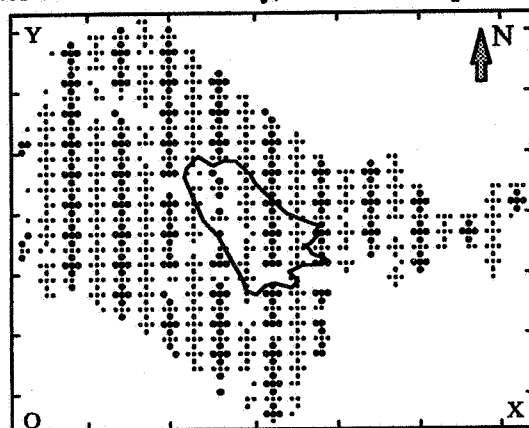
L'étude méthodologique de l'inventaire forestier de Zürichberg illustre l'amélioration de l'estimation apportée une information auxiliaire.

La variable z étudiée est le nombre de tiges à l'hectare. Un échantillonnage au sol très dense au sol est complété, à maille quatre fois plus dense, par une campagne de photographies aériennes (figure 6). L'information des photographies est codée selon 3 variables qualitatives :

- l'âge du peuplement : 3 à 6 suivant la référence en usage dans la profession

- le degré de mélange : 1 ou 2 selon que les conifères ou les feuillus sont majoritaires
- le degré de couverture : 1 peu dense ou 2 dense.

Par une méthode de régression, les 3 variables qualitatives sont converties en une variable quantitative auxiliaire y . La valeur moyenne de z sur des carrés de 1 ha est estimée par cokrigage des données de z et de y , ces dernières quatre fois plus nombreuses.



- Comptage au sol
- • Photographie aérienne
- ~ Limite du "petit domaine"

Figure 6 : Echantillonnage de la forêt de Zürichberg

Un "petit domaine" de 17 ha a été échantillonné de façon exhaustive, ce qui permet la comparaison des prévisions au résultat exact. Seuls les résultats moyens sur l'ensemble du petit domaine sont présentés ici.

Le cokrigage est comparé au krigeage par les seules données z du comptage au sol. Le tableau 1 montre que l'information auxiliaire améliore l'estimation. Les résultats ont permis de proposer une réduction au quart de la densité d'échantillonnage au sol. La comparaison des erreurs d'estimation à la variance de krigeage a montré la validité de celle-ci.

	Valeur vraie	Krigeage comptage au sol seul	Cokrigeage	Cokrigeage réduction du comptage au sol (1/4 des placettes)	Estimateur classique *
Valeur	280.23	294.35	281.67	294.97	257.14
Ecart-type		40.56	26.91	15.10	48.59

* Estimateur classique : cf. Mandallaz, 1991. A unified approach to sampling theory for forest inventory based on infinite population and superposition models. Chair of forest inventory and planning, E.T.H., Zürich.

Tableau 1 : Cokrigeage et comparaison aux valeurs vraies sur le petit domaine.

5. Conclusion : Géostatistique & Systèmes d'Information Géographique

A partir des données expérimentales, il existe non pas une mais plusieurs cartes de l'état de l'environnement, selon le problème posé.

La validité de l'estimation géostatistique dépend de la qualité de l'analyse variographique, étape interactive d'interrogation des données. Automatiser cette étape ne paraît pas souhaitable. L'interactivité toujours plus développée des logiciels permet au contraire de l'inclure parmi les traitements courants de l'information "géoréférencée" – au prix d'une formation préalable des utilisateurs, comme pour toute technique. La géostatistique, fondée sur l'examen et la modélisation de la structure spatiale des variables étudiées, s'inclut ainsi naturellement dans la panoplie des outils pour les SIG.

Références Bibliographiques

Sur la géostatistique

Matheron G. La Théorie des Variables Régionalisées et ses Applications. *Les Cahiers du Centre de Morphologie Mathématique*, fasc. 5, ENSMP édit., 1971, 212 p.

Chauvet P. Traitement des données à support spatial : la géostatistique et ses usages. ENSMP, 1992

Sur les applications au traitement de données environnementales

Gohin F. Analyse géostatistique des champs thermiques de surface de la mer, Thèse de Docteur Ingénieur, ENSMP, 1987

Wackernagel H., Hudson G. Kriging mit externer Drift am Beispiel von Temperaturmessungen aus Schottland. In G.J. Peschel Greifswald, *Anwendung geostatistischer Verfahren.*, Sven-von-Loga Verlag, 1992, Vol. 3, pp. 15-24. (Beiträge zur Mathematischen Geologie und Geoinformatik).

de Fouquet, C., Mandallaz, D. Using geostatistics for forest inventory with air cover : an example. In A. Soares (Ed.), *Geostat Tróia '92*, Kluwer Academic Publ., Dordrecht, Holland, 1993, Vol. 2, pp. 875-886. (Quantitative Geology and Geostatistics 5).