CHPITRE 3

Chapitre III : Analyse de la structure spatiale de la variable

III.1 Introduction

Le problème classique en géostatistique est la prédiction d’une grandeur physique d’intérêt (variable régionalisée) sur un domaine d’étude à partir d’un ensemble fini d’observations éventuellement espacées irrégulièrement. Les variables régionalisées ont une structure d'auto-corrélation qui dépend du module et de la direction du vecteur séparant deux points de mesure. Mathématiquement, une variable régionalisée est une fonction du point x. Cette fonction est généralement irrégulière et montre deux aspects complémentaires :

- un aspect aléatoire qui explique les irrégularités locales ;

- un aspect structuré qui reflète les tendances du phénomène à grande échelle.

III.2 Analyse de la structure spatiale de la variable

III.2.1 Notion de variogramme

Considérons une propriété du sol notée Z connue en n points de l'espace géographique, chacun de ces points étant repérés par le vecteur $\vec{\mathbf{x}}$ de ses coordonnées géographiques (longitude et latitude). De la sorte, la notation $\mathbf{Z(x_i)}$ représente la valeur observée de la propriété Z au i ème point d'échantillonnage de coordonnées xi.

Deux observations situées l'une près de l'autre devraient, en moyenne, se ressembler davantage (être plus corrélées) que deux observations éloignées. Cette idée peut être exprimée simplement par la comparaison de ces deux valeurs : prenons deux points pour lesquels on connaît des valeurs Z(x1) et Z(x2) de la propriété Z. On cherche à comparer ces deux valeurs. Une façon simple est d'utiliser la variance entre les observations de ces deux sites, notée s². Elle est par définition égale à :

$$s^2 = [Z(x1) - \bar{Z}]^2 + [Z(x2) - \bar{Z}]^2 \quad (1)$$

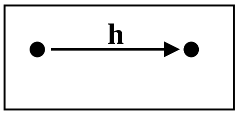
Où : $\bar{\mathbf{Z}}$ est la moyenne entre ces deux observations.

Cette variance s², qui traduit l'importance des écarts à la moyenne, est d'autant plus grande que les observations sont différentes. On peut d'ailleurs développer l'équation (1) pour obtenir une autre expression de cette valeur s² :

$$s^2 = \frac{1}{2}[Z(x1) - Z(x2)]^2 \qquad (2)$$

Cette équation peut être écrite pour tout couple de sites.

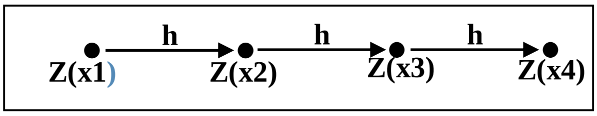
Pour cela, considérons deux sites Z(xi) et Z(xi+h) où xi représente les coordonnées géographiques d'un des sites et h est un vecteur caractérisant la distance entre les sites.



L'équation (2) s'écrit alors :

$$s^2 = \frac{1}{2}[Z(xi) - Z(xi + h)]^2 \qquad (3)$$

Calculons à présent la distance géographique séparant chacun des points d'observation et considérons les m couples de points séparés par une même distance géographique h.



On peut comme précédemment, calculer la variance des observations pour les sites pris deux à deux. La moyenne ŝ² de ces m variances s'écrit en employant (3) :

$$\hat{s}^2 = \frac{1}{2m}\sum_{i=1}^{m}\frac{1}{2}[Z(xi) - Z(xi + h)]^2 \qquad (4)$$

Pour une distance h séparant deux points d'observation, ŝ² rend compte de la ressemblance des observations faites en ces deux points : il sera d'autant plus grand que ces observations sont différentes. ŝ² dans ce cas est qualifiée de "semi-variance".

Cette semi-variance est appelée aussi variogramme, généralement donné par l'expression suivante :

$$\gamma(h) = \frac{1}{2\,m}\sum_{i=1}^{m}[Z(xi + h) - Z(xi)]^2) \qquad (5)$$

De ce fait, on conçoit que deux observations soient en général d'autant plus semblables qu'elles sont proches géographiquement l'une de l'autre.

Le calcul de $\boldsymbol{\gamma}(\mathbf{h})$ pour différentes distances h, va permettre de quantifier cette idée, il permet de suivre l'évolution des écarts entre des observations en fonction de la distance qui les sépare.

III.2.2 Hypothèses de base et définition

L'intérêt de cette notion simple et les conditions de sa généralisation ont été définis par la théorie des variables régionalisées (Matheron, 1965). Cette théorie montre que la généralisation de l'équation (5) suppose deux conditions, regroupées sous le terme d'hypothèse intrinsèque :

- L'espérance mathématique ne dépend pas de x, $E[Z(x)]=m$
  ou
  L'espérance des écarts est zéro, $E[Z(x) - Z(x+h)] = 0$
- La covariance entre $Z(x)$ et $Z(x+h)$ ne dépend que de h $Cov(Z(x), Z(x+h)) = C(h)$; stationnarité du second ordre, $C(h)$ appelé fonction de covariance ou covariogramme.
  ou
  Le variogramme □(h) ne dépend pas de la localisation x, seulement de h (soit en module, soit en module et en direction) ; C à d la différence $[Z(x) - Z(x+h)]$ a une variance finie, qui ne dépend que de la distance h séparant les points.

$$\begin{aligned} \mathbf{Var}\,[\mathbf{Z}(\mathbf{xi}+\mathbf{h})-\mathbf{Z}(\mathbf{xi})] &= \mathbf{2}\,\boldsymbol{\gamma}(\mathbf{h}) \\ &= \mathbf{E}[\mathbf{Z}(\mathbf{xi}+\mathbf{h})-\mathbf{Z}(\mathbf{xi})]^{2} \end{aligned}$$

Cette dernière hypothèse est légèrement moins restrictive que la stationnarité du second ordre est appelée hypothèse intrinsèque.

Conclusion :

Quand ces deux conditions sont vérifiées, la valeur $\hat{s}^2$ définie dans l'équation (4) constitue un estimateur non biaisé de la fonction $\gamma(h)$ définie en (5). Cette fonction $\gamma(h)$ est nommée variogramme.

En étudiant l'évolution du variogramme $\gamma(h)$ en fonction de la distance h séparant des couples d'observation, on va analyser la façon dont se détériore l'information acquise en un point au fur et à mesure que l'on s'éloigne de ce point.