Atelier de Formation aux Techniques d'Etude de l'eau dans le système Sol - Plante - Atmosphère

Mbour, 30 Novembre - 10 Décembre 1992

L'EAU DANS LE SOL
SPATIALISATION DES OBSERVATIONS

Pascal BOIVIN

L'eau dans le sol : Spatialisation des observations.

Introduction

Que l'on cherche à quantifier le contenu du sol (variables d'état ou paramètres statiques tels que granulométrie, teneurs en eau, porosité etc..) ou des paramètres dynamiques (tels que conductivité hydraulique, facteur d'échelle etc..), on ne dispose généralement que d'un nombre *limité* d'observations *ponctuelles*.

La notion d'observation ponctuelle dépend elle même du protocole de caractérisation employé. Ainsi une granulométrie sera généralement pratiquée sur 200g de sol sec tamisé et broyé, un conductvité hydraulique pourra être déterminée sur différents supports dont la dimension est variable (Bouma, 1983, Boivin, 1990), mais peut être qualifiée de ponctuelle en regard de la dimension de l'objet étudié (profil de sol, parcelle, unité pédologique...).

A partir des données dont on dispose se présente généralement l'un des problèmes suivants:

-quelle est la valeur de la variable que l'on étudie en un point où elle n'a pas été déterminée?

-quelle est la valeur moyenne de la variable étudiée sur un bloc (en deux ou trois dimensions) pour lequel on dispose d'observations *ponctuelles*?.

Ce sont ces problèmes de *l'interpolation* et de *l'estimation d'une valeur moyenne* à partir d'un ensemble de valeurs ponctuelles que nous allons aborder.

I Problématique

I-1 Les propriétés des sols et leur variabilité

De nombreuses publications sont consacrées à la variabilité des propriétés du sol. Dans la suite de ce texte, seule la variabilité spatiale sera évoquée. On pourra se reporter aux publications de Vauclin (1982) et Gascuel-Odoux (1987) pour une synthèse des principales publications dans ce domaine.

* Notions d'homogénéité et d'uniformité

Un milieu homogène est caractérisé par une loi de distribution unimodale.

Un milieu est uniforme s'il sa fonction de distribution est une combinaison linéaire de DIRAC.

Un sol n'est jamais homogène et uniforme : il impose donc le recours à une description stochastique.

* Notion d'indépendance spatiale des observations

Les observations réalisées sont rarement indépendantes dans l'espace. En d'autres termes, la corrélation entre deux valeurs varie en fonction de la distance qui sépare les observations. Il est donc nécessaire, lors d'une description stochastique, de tenir compte des dépendances spatiales.

Remarque : en cas de dépendance spatiale, on se trouve hors du champ d'application de la statistique "classique" ou fishérienne.

1-2 Méthodes d'estimation et d'interpolation : quel choix ?

Les méthodes d'estimation locale ou globale sont variées en science du sol.

L'une des plus anciennes : la cartographie. Affectation d'un profil type à une unité de sol distribuée en plages cartographiques. Eventuellement une variance est associée (e.g. Geoderma, 1992, Walter, 1990).

Interpolation : il existe une panoblie de méthodes. Exemples :
-polygônes de thyssen;
-pondération inverse de la distance;
-isopondération
etc...

Problèmes : Ces méthodes ont un caractère arbitraire. Elles introduisent généralement un biais. Les interpolateurs ne sont pas tous exacts. Les variances d'estimation ne sont pas connues.

En 1965, Matheron énonce la *théorie des variables aléatoires régionalisées*, qui donne naissance à la *géostatistique*. Cette théorie permet de palier les inconvénients cités ci-avant et s'impose comme la plus performante *dans son domaine d'application*.

II- La géostatistique.

II-1 Cadre conceptuel et hypothèses.

La théorie des variables régionalisées est conçue pour prendre en compte des phénomènes se déployant dans un espace, avec *un caractère aléatoire et un caractère structuré* (c'est à dire une dépendance des valeurs prises entre deux points distincts).

Ses principaux développements reposent sur les **hypothèses** suivantes :
Soit Z une Variable Aléatoire Régionalisée (VAR) dont Z(r) est une réalisation constituée de n observations :

Ergodicité :
Les variations spatiales de Z(r) constatées dans une réalisation représentent toutes les variations possibles de l'ensemble. C'est à dire que l'ensemble des valeurs dont on dispose permet de caractériser les propriétés de Z.

Stationnarités :
La densité de probabilité de Z(r) et ses moments associés sont indépendants de l'espace;
La covariance entre les observations Z(r) et Z(r+h) est indépendante de l'espace (du vecteur r) et ne dépend que de la distance h qui sépare deux observations.
La stationnarité de la variance entraîne celle de la variance (le contraire n'est pas vrai). On se limite donc en pratique aux hypothèses d'ergodicité, de stationnarité d'ordre 1 (E(Z(r)) indépendant de r), et de stationnarité des accroissements (l'accroissement Z(r+h)-Z(r) est indépendant de r et ne dépend que de h)(hypothèse intrinsèque).

II-2 Eléments de théorie, formalisme.

a) Le variogramme : description de la structure spatiale.

En faisant l'hypothèse complémentaire :
$E(Z(r+h)-Z(r))=0$
(les acroissements sont en moyenne nuls)

On écrit :
$\gamma(h) = 1/2 \, .E((Z(r+h)-Z(r))^2)$

On peut expliciter cette relation comme représentant la (demi)variance entre deux points de mesure en fonction de la distance qui les sépare.

Cette fonction est appelée *semivariogramme*. Son comportement est utilisé pour étudié la VAR. Ce comportement permet en effet de caractériser la **structure spatiale** de la variable (Journel, 1978; Delhomme, 1978). On s'intéresse en particulier (fig 1) :

1-au comportement à l'origine : en principe, $\gamma(0)=0$. En réalité, on ne dispose pas de couples de mesures séparés par une distance nulle, mais uniquement de couples de mesures séparés par une distance faible h. En extrapolant le variogramme à l'origine, il arrive que ce dernier ne converge pas vers 0. Dans ce cas, on dira qu'il existe un «effet de pépite». Ce dernier peut correspondre :

-à une variance s'exprimant pour des distances inférieures à h;

-à une erreur de mesure ou «bruit de fond» de la mesure.

2-au comportement au voisinage de l'origine : en principe, la variance existant entre deux points de mesure doit augmenter lorsque la distance qui les sépare augmente. Ceci se traduit par une pente du variogramme. Il arrive que le variogramme ait une pente nulle : il est dit «plat» ou «pépitique pur». Ceci signifie que la variable étudiée est purement aléatoire : on se trouve alors dans le domaine d'application de la statistique classique.

3-au comportement aux grandes distances («à l'infini») : le semi variogramme atteint généralement une valeur constante appelée «palier», à partir d'une distance appelée «portée» ou

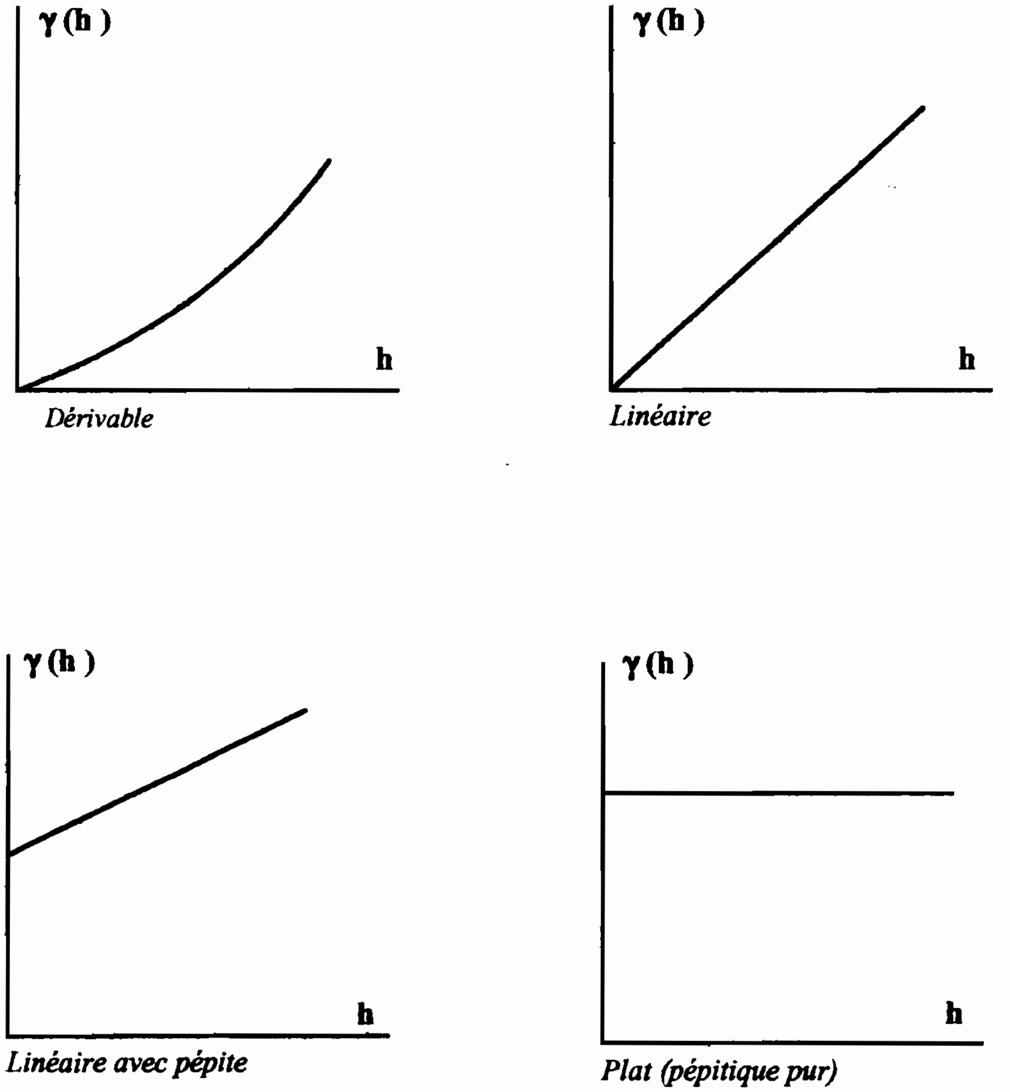

Figure 1 : comportement du variogramme à l'origine (d'après Delhomme, 1978)

«distance d'autocorrélation». Ceci signifie qu'au delà de cette distance, les valeurs de la variable sont statistiquement indépendantes les unes des autres.

Le variogramme peut ne pas être borné. Lorsqu'il se rapproche d'une branche d'hyperbole, ceci indique généralement l'existence d'une dérive (ou tendance de la variable : ses valeurs sont partiellement expliquées par les coordonnées des points d'observation). Dans ce cas, l'hypothèse intrinsèque n'est pas vérifiée. Des transformations de la variable sont proposées (Chiles, 1977), de façon à se ramener au cas sans dérive. Cette opération est toutefois délicate à pratiquer dans le cas des études pédologiques où l'on dispose généralement de peu de points de mesure (moins de cent) (cf. exemple traité en travaux pratiques).

b) Le krigeage : un interpolateur optimal.

Le variogramme peut être modélisé. Quelques modèles, correspondant à des fonctions semi-définies positives, sont proposés (Webster, 1985). En pratique, les modèles les plus couramment utilisés sont les modèles exponentiel et sphérique (fig. 2).

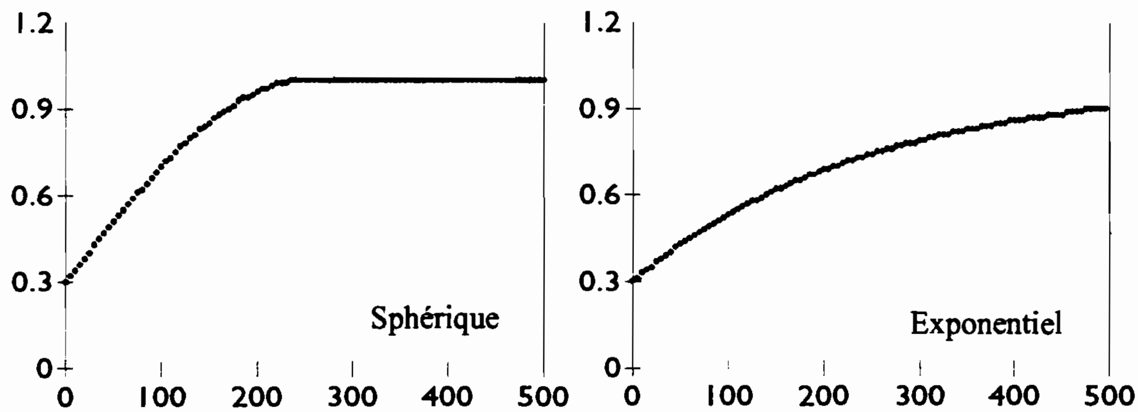

Figure 2 : Modèle sphérique et modèle exponentiel

Le modèle retenu pourra alors être utilisé comme interpolateur linéaire pour l'estimation ponctuelle, zonale ou globale de la variable. Cette opération est appelée krigeage. On trouvera dans Journel (1978) une description du système de krigeage. Elle permet de réaliser des interpolations (estimations sur des blocs, sur des points, calcul de cartes) dont on démontre en particulier qu'elles ont les propriétés suivantes:

-interpolation «non biaisée» : l'erreur moyenne est (théoriquement) nulle ou (en pratique et statistiquement) proche de 0;

-interpolation optimale : la moyenne quadratique de l'erreur est minimisée.

En plus d'une carte calculée par interpolation, cette technique offre l'avantage de fournir une carte de variance d'estimation : en tout point interpolé sont connues la valeur estimée de la variable (qui est sa valeur la plus probable en ce point), et la variance d'estimation. Ces caractéristiques en font un outil privilégié d'interpolation.

II-3 Calculs - Application pratique

Une étude géostistique comporte plusieurs étapes. La première étape est de disposer d'un échantillonnage. On montre expérimentalement (Gascuel et Boivin in Geoderma, 1992) ou par simulation (Webster et Oliver, 1992) qu'environ 150 points d'obervation permettent pour un site donné de réaliser une étude géostatistique dans de bonnes conditions.

Outre le nombre de valeurs dont on dispose, la répartition des données dans l'espace (schéma d'échantillonnage) influe beaucoup sur la qualité de l'interprétation qui pourra être faite. Nous reviendrons sur ce point.

Une fois les données disponibles, les calculs du variogramme et de krigeage peuvent être réalisés. Ils supposent de disposer d'un ordinateur et d'un logiciel de géostatistique. Ce point sera également discuté plus loin.

a) Variogrammes

Technique de calcul

Pour calculer un variogramme, il faut disposer d'un ensemble d'observations localisées. Chaque point du variogramme correspond à une valeur moyenne calculée sur un ensemble de couples de points. Ainsi pour une distance h, seront recherchés tous les couples de points ayant cette interdistance, de façon à estimer la valeur du variogramme à ce pas. Selon que les points sont répartis sur un maillage régulier ou non, on sera amené à regrouper les couples de points par classes d'interdistances. On estime qu'il faut environ cinquante couples pour estimer un variogramme à un pas donné, sans que ce chiffre repose sur une base théorique.

Interprétation

Il est recommandé d'examiner au préalable la fonction de répartition des valeurs. L'existence de valeurs aberrantes (cf exemple donné en TP), une fonction de répartition multimodale, risquent de fausser l'interprétation du variogramme. En particulier, la présence de 1 ou 2 valeurs erronées et aberrantes dans un lot de 100 valeurs peut suffire à fausser totalement l'allure du variogramme.

Dans un premier temps, la validité des hypothèses est discutée au vu du variogramme. Ainsi un variogramme non convergent, ou décroissant à partir d'une distance, ou dont le palier est inférieur à la variance globale, seront autant d'indicateurs de l'existence d'une dérive.

Dans le cas où une dérive est identifiée, il est possible :

i)-de calculer des variogrammes directionnels (calcul où l'on ne prend en compte que des bipoints orientés), en cherchant à identifier des directions préférentielles sans dérive;

ii)-de filtrer cette dérive par régression de la variable étudiée sur les coordonnées des observations, puis de n'étudier que les résidus (variable - dérive);

iii)-de mettre en oeuvre des méthodes plus sophistiquées de prise en compte de la dérive (accroissements d'ordre k, Journel, 1978), en pratique ces techniques ne sont généralement pas applicables avec des lots de données limités à 100 ou 200 valeurs.

Le cas i) sera ensuite traité de façon ordinaire, en supposant que la structure spatiale de la variable est identifiée pour la direction sans dérive, et que la stationnarité est *localement* vérifiée.

Le cas ii) revient à supposer que la variable recueillie est la somme d'une dérive et d'une VAR, et que cette dérive est estimée correctement par régression de la variable sur les coordonnées des points.

Lorsque cette étape est franchie, on ajuste généralement un modèle au variogramme expérimental, de façon à déterminer ses principales caractéristiques : pépite, palier, portée. La recherche du modèle peut être visuelle ou utiliser des protocoles d'ajustement non linéaire. Certains logiciels proposent les deux possibilités.

Le variogramme renseigne sur la structure spatiale de la variable étudiée. Il permet en particulier de savoir à partir de quelle distance deux valeurs sont autocorrélées, et quelle est l'importance de l'autocorrélation. Le palier donne une estimation non biaisée de la variance. L'effet de pépite est un majorant de l'erreur de mesure.

Ces différentes interprétations sont importantes, par exemple pour décider d'un échantillonnage (au cours du temps, pour estimer une valeur moyenne...), pour interpréter des essais agronomiques en tenant compte des corrélations spatiales (cf. Vauclin et al, 1990, Van Es et Van Es in Geoderma, 1992).

Remarque : on limite généralement l'interprétation d'un variogramme à la demi-longueur du site étudié. Au delà de cette distance, il est en effet fréquent qu'un nombre limité de valeurs particulières pèse très fortement sur l'allure du variogramme.

b) Krigeage : un apperçu.

La mise en oeuvre d'un krigeage suppose de disposer d'un ensemble de valeurs localisées et d'un modèle de variogramme représentant la structure spatiale de la variable étudiée. Les propriétés du krigeage ne seront vérifiées que si les hypothèses inhérentes à la démarche sont elles-mêmes vérifiées, c'est à dire si l'on se trouve bien dans le *domaine d'application* de cette technique.

En théorie, l'estimation d'une valeur ponctuelle peut se faire en tenant compte de toutes les valeurs connues de la variable à estimer. En pratique, si la variable manifeste une structure dans l'espace (c'est à dire si l'on ne se trouve pas dans le domaine d'application de la statistique classique), les poids pris par les points les plus proches pour l'estimation d'une valeur représentent la totalité de l'estimation, si l'on prend en considération la précision de la mesure. On se limite donc à tenir compte pour une estimation ponctuelle d'un *voisinage glissant* de quelques voisins connus (souvent moins de 10).

On prend en outre, au niveau du logiciel, les dispositions suivantes :

-rayon maximum de recherche pour la prise en compte d'un point d'appuis à l'estimation (voisinage);

-nombre minimum de valeurs voisines connues rencontrées dans ce voisinage, pour qu'une estimation soit calculée;

-nombre maximum de voisins à prendre en compte.

-distance entre deux valeurs estimées (sur quelle maille doit se faire l'estimation);

-polygône d'estimation;

Sur cette base, une carte des valeurs estimées est produite (un maillage de valeurs ponctuelles ou calculées sur des blocs). Une carte des variances d'estimation est également calculée.

III - Un exemple d'application

Présentation d'une étude de variabilité et d'une spatialisation de mesures ponctuelles de conductivité hydraulique à saturation en travaux pratiques.

IV Autres aspects

IV-1 La géostatistique multivariée.

Une généralisation de la géostatistique au cas de l'étude plusieurs variables a été proposée (voir par ex. H. Wackernagel, in Geoderma, 1992). Elle repose sur la généralisation de l'écriture du variogramme : le covariogramme s'écrit :

$\gamma_{12}(h)= E((Z1(x+h)-Z1(x))*(Z2(x+h)-Z2(x)))$
où Z1 et Z2 sont des VAR.

Le système de cokrigeage permet d'estimer la valeur de la variable Z à partir des valeurs voisines connues de Z **et** des valeurs connues d'autres variables corégionalisées.

Ce type d'étude est assez lourd numériquement. Peu de logiciels permettent de réaliser un cokrigeage. L'une des contraintes est de disposer d'un lot de valeurs connues des différentes variables **aux mêmes points et simultanément**.

Le domaine d'application est généralement le suivant :

-aide à l'estimation d'une variable peu échantillonnée (coûteuse) à partir d'autres variables;

-changement de variable (transfert d'échelle).

IV-2 Logiciels

L'un des standards est le logiciel BLUEPACK du centre de Géostatistique de l'Ecole des Mines de Fontainebleau. Il est cher et ne fonctionne pas sur micro-ordinateur.

De nombreux logiciels sur PC ont vu le jour ces dernières années. Nous en citerons 2 :

GEOEAS (logiciel américain, du domaine public)

GEOSTAT- PC (logiciel en Français, commercialisé par l'ORSTOM).

IV-3 Formation

La pratique de la géostatistique suppose une formation approfondie. Nous recommandons en particulier les stages de géostatistique appliquée de l'Ecole des Mines à Fontainebleau (France).

Bibliographie

BOIVIN P. 1990, Caractérisation physique des sols sulfatés acides de la vallée de Katoure (basse Casamance, Sénégal) : variabilité spatiale et relation avec les caractéristiques pédologiques. Thèse Université Paris VI, Ed. ORSTOM Paris, Série Etudes et Thèses, 226 pages.

BOUMA, J., 1983, Use of soil survey data to select measurement techniques for hydraulic conductivity. Agr. Water man., 6 (2/3), 177-190.

CHILES, J.P., 1977,Géostatistique des phénomènes non stationnaires, Thèse université Nancy1.

DELHOMME, J.P., 1978, Application de la théorie des variables aléatoires régionalisées dans les sciences de l'eau, Bull. BRGM, III, N4, pp.341-375

GASCUEL-ODOUX, C., 1987, Variabilité spatiale des propriétés hydriques du sol, cas d'une seule variable: revue bibliographique, Agronomie, 7 (1), 61-71

GEODERMA, 1992, Proceedings of the first AISS Congress "Pedometrics", Wageningen, the Netherlands.

JOURNEL, A.G., 1977, Géostatistique minière Multig. ENSMP, Centre de géostatistique, 2 tomes, 733p.

MATHERON, G. 1965, Les variables régionalisées et leur estimation Editions MASSON

VAUCLIN, M., 1982, Méthodes d'étude de la variabilité spatiale des propriétés d'un sol, Colloque S.H.F.-I.N.R.A., Avignon, Juin 1982, pp 9-45

VAUCLIN, M., MUNOZ-PARDO, J., RUELLE, P. et VACHAUD, G., 1990, Analyse statistique et géostatistique de la variabilité spatiale d'une parcelle agronomique, Multig. IMG

WALTER, C., 1990, Estimation des propriétés du sol et quantification de leurs propriétés à moyenne échelle, Thèse de Docteur de l'Université Paris 6, 172 pages.

WEBSTER, R., 1985, Quantitative spatial analysis of soil in the field. In : B.A. Stewart (editor), Advances in Soil Science. 3. Springer. New York, 1-70.

WEBSTER, R. and OLIVER, M., 1992, Sample adequately to estimate variograms of soil properties Journal of Soil Science, 43:177-192