HAL
open science

Comparaison d'approches multivariables pour l'etude de la variabilite spatiale des sols

Goulard Michel, Marc Voltz, Pascal P. Monestiez

▶ **To cite this version:**

Goulard Michel, Marc Voltz, Pascal P. Monestiez. Comparaison d'approches multivariables pour l'etude de la variabilite spatiale des sols. Agronomie, 1987, 7 (9), pp.657-665. hal-02722681

HAL Id: hal-02722681

https://hal.inrae.fr/hal-02722681v1

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Agronomie, 1987, 7 (9), 657-665.

Comparaison d'approches multivariables pour l'étude de la variabilité spatiale des sols

Michel GOULARD, Marc VOLTZ* & Pascal MONESTIEZ**

I.N.R.A., Unité de Biométrie
** Science du sol, Centre de Recherches de Montpellier, 9, place Pierre-Viala, F 34060 Montpellier Cedex*
*** I.N.R.A. Laboratoire de Biométrie, Centre de Recherches d'Avignon, Domaine St-Paul, F 84140 Montfavet*

RÉSUMÉ

Les auteurs effectuent une revue critique des méthodes statistiques et géostatistiques employées pour l'étude de la variabilité spatiale multiparamétrique des propriétés hydrodynamiques du sol.
L'illustration et la dicussion des méthodes fait référence, chaque fois que cela est possible, aux résultats obtenus en science du sol. Deux approches différentes mais complémentaires sont analysées.
La première repose sur des techniques descriptives (analyse en composantes principales, analyse des covariogrammes, analyse des proximités spatiales, analyse krigeante, mise en facteur d'échelle), qui sont examinées suivant leurs aptitudes à faciliter la synthèse et la compréhension des covariations spatiales.
La seconde vise à l'interpolation spatiale d'un paramètre en utilisant des échantillons d'observations d'autres paramètres. Les méthodes utilisées (régression, cokrigeage) sont comparées en fonction de la qualité des interpolations.

Mots clés additionnels : *Géostatistique, facteur d'échelle, régression, cokrigeage, analyse multidimensionnelle.*

SUMMARY

Comparison of multivariate approaches to the study of the spatial variability of soils.

This review focuses attention on statistical and geostatistical methods which are used to study the multivariate spatial variability of the hydrodynamical properties of soil. Whenever possible the discussion of the methods is illustrated by examples found in soil science papers. Two different, but complementary approaches are analysed. The first concerns descriptive procedures (principal component analysis, covariogram analysis, spatial proximity analysis, kriging analysis, scaling factor analysis) which are examined for their ability to synthesize and explain spatial covariations. The second sets out the purpose of spatial interpolation of one parameter from samples of others. The methods usually employed (regression, cokriging) are compared on the basis of their precision.

Additional key words : *Geostatistics, scaling factor, regression, cokriging, multivariate analysis.*

I. INTRODUCTION

L'étude de la variabilité spatiale des paramètres physiques et hydrodynamiques du sol s'est développée de façon considérable depuis une vingtaine d'années. En effet, comme l'ont souligné de nombreux auteurs (NIELSEN *et al.*, 1973 ; PECK, 1983 ; GASCUEL-ODOUX, 1984), la transposition de modèles de transferts hydriques à validité locale vers des échelles plus larges, telles la parcelle, l'unité de sol ou de paysage, passe par ce type d'étude. Nous nous proposons donc de faire une synthèse des méthodes statistiques qui peuvent être utilisées. Nous illustrerons, chaque fois que cela est possible, les méthodes présentées par les résultats de leur application en science du sol.

Les paramètres ou variables d'état du sol sont considérés ici en tant que variables spatiales, appelées aussi variables régionalisées. La première approche consistant à étudier séparément la structure spatiale de chaque variable régionalisée a fait l'objet d'une synthèse (GASCUEL-ODOUX, 1987). Ce type d'approche paraît insuffisant pour la transposition spatiale de modèles multiparamétriques locaux ; dans ce cas l'analyse des covariations spatiales de l'ensemble des paramètres des modèles utilisés doit être envisagée. C'est ainsi que la compréhension des phénomènes de transfert hydrique et de leur variabilité dans un volume pédologique donné passe par l'étude multivariable spatiale des relations teneur en eau-potentiel et conductivité hydraulique-teneur en eau. La deuxième approche citée qui analyse simultanément les variations des variables régionalisées fera l'objet de

cette note. Nous ne parlerons pas ici de la construction d'unités pédologiques par regroupements de points de mesure ou de la validation d'un tel découpage. GIRARD (1983) a réalisé une analyse bibliographique de ce type d'analyses qui concernent des échelles d'approche cartographique inférieures à celles qui nous intéressent. L'accent sera mis sur l'étude des relations entre variables mesurées sur des volumes pédologiques restreints.

Deux objectifs spécifiques apparaissent le plus souvent lors des analyses multiparamétriques de la variabilité du sol :

— Chercher des variables, combinaisons fonctionnelles de paramètres pédologiques, qui synthétisent la variabilité spatiale multiparamétrique du sol et en facilitent la compréhension. Ceci permet notamment de réduire le coût d'une étude de sensibilité des modèles de transfert aux variations conjointes de leurs paramètres, par la diminution du nombre des paramètres envisagés (RUSSO & BRESLER, 1981).

— Chercher des paramètres pédologiques d'accès expérimental aisé qui rendent compte de la variabilité des paramètres hydrodynamiques. L'objectif est ici de limiter le coût expérimental d'une étude des variations spatiales des caractéristiques hydriques du sol, dont la mesure est longue et délicate.

Nous présenterons les principes des méthodes liées à ces objectifs et leurs résultats. Voyons d'abord les méthodes descriptives puis les méthodes d'estimation.

II. MÉTHODES DESCRIPTIVES

Ces méthodes cherchent à décrire synthétiquement un tableau de données résultant de la mesure de plusieurs variables. Elles sont donc plutôt liées au premier objectif présenté en introduction. En rapport avec cet objectif, l'accent sera mis sur la possibilité de trouver des structures spatiales communes aux variables étudiées et des résumés spatiaux. Nous verrons tout d'abord l'analyse en composantes principales (A.C.P.) puis la géostatistique qui permet l'étude des variations spatiales simultanées de plusieurs variables. Nous terminerons par une méthode non linéaire fondée sur le concept de similitude appliquée aux milieux poreux.

A. Analyse en composantes principales

On sait que l'A.C.P. est particulièrement adaptée à la description d'un tableau rectangulaire A de dimension n individus par p variables quantitatives (le terme A_{ij}, situé sur la ligne i et en colonne j du tableau, est la mesure d'une variable j en un point i). Nous allons l'utiliser pour décrire la variation multiparamétrique du sol. Rappelons brièvement le principe de la méthode.

Le tableau A définit deux espaces :

— l'espace des individus où un point i est représenté par le point de coordonnées (A_{i1}, A_{i2}, ..., A_{ip}) ;

— l'espace des variables où une variable j est représentée par le point de coordonnées (A_{1j}, A_{2j}, ..., A_{nj}).

Le but de la méthode est de trouver une représentation graphique la plus proche possible du nuage de points défini dans l'espace des individus (CAILLIEZ & PAGES, 1976 ; BOUROCHE & SAPORTA, 1980). Pour résoudre ce problème, on cherche les axes, combinaisons linéaires des variables de départ, qui sont de variance maximum sous contrainte de non corrélation entre eux. Les axes trouvés définissent la représentation cherchée, pour le critère de proximité des moindres carrés, mais aussi pour beaucoup d'autres critères (RAO, 1980 ; SABATIER *et al.*, 1984).

Résultats

La méthode appliquée à des variables mesurées en des points spatiaux permet généralement de dégager le fonds régional des mesures, c'est-à-dire la tendance en grand du champ de mesure. Ce travail pourra se faire par simple cartographie des premiers axes.

WEBSTER (1977), dans une étude sur les sols à relief Gilgaï, utilise l'A.C.P. pour étudier la périodicité de paramètres physiques et chimiques. La première composante, qui représente 26 p. 100 de la variation totale du nuage, traduit cette périodicité sur le transect étudié. C'est la seule composante à avoir une structure spatiale, les autres ne sont pas interprétées. De la même façon VOLTZ (1986) analyse par l'A.C.P. la variabilité spatiale des courbes de rétention sur une parcelle. La cartographie du premier axe (fig. 1a), qui représente 60 p. 100 de la variation totale, traduit la tendance générale au niveau du champ : il existe un gradient spatial dans la direction Nord-Est - Sud-Ouest. Ce dernier est perceptible, mais plus faiblement, sur la carte de la deuxième composante (fig. 1b). Dans ces 2 exemples, l'A.C.P. montre qu'une partie de l'information est structurée spatialement ; une composante la résume de façon satisfaisante.

L'A.C.P. effectue une bonne synthèse des déterminants (périodicité, gradients) de la variabilité spatiale d'un ensemble de paramètres. Le calcul de la variance d'une variable s'effectue en utilisant le centrage par rapport à la moyenne empirique. Les variations systématiques seront intégrées dans la variance calculée, ce qui explique la sensibilité de l'A.C.P. aux gradients. Les liaisons plus fines entre variables pourront être masquées, d'autant que le gradient peut se répercuter sur plusieurs axes (VOLTZ, 1986). L'A.C.P. donne alors des résultats comparables à une décomposition du gradient en un ensemble de surfaces polynômiales de degré croissant.

La pratique générale veut que l'analyse soit effectuée sur les données réduites quand les variables ne sont pas homogènes (du point de vue de leurs unités ou de leurs variances). Dans cette situation, la variance résiduelle d'une variable présentant un gradient voit son influence diminuée au profit du gradient. Si l'intérêt principal se porte sur les variations résiduelles (caractéristiques d'une échelle spatiale donnée) il est possible de modéliser les gradients et d'effectuer l'A.C.P. des résidus. Cette approche pose d'autres problèmes : la qualité d'ajustement des gradients peut créer des artefacts dans l'analyse des résidus.

B. Utilisation de la géostatistique

Pour décrire le tableau de données, nous allons utiliser des concepts probabilistes. Les variables mesurées

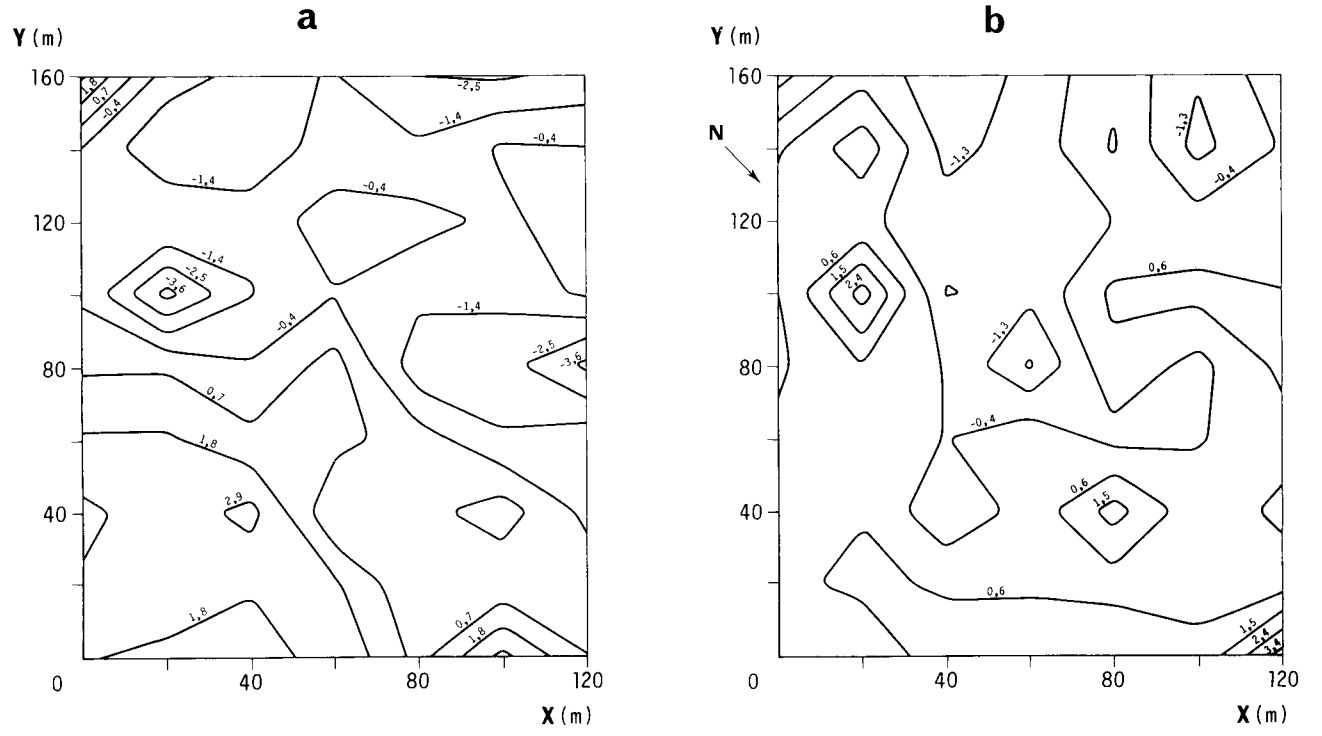

Figure 1

Cartographie par courbes de niveau de 2 composantes d'une A.C.P. sur les courbes de rétention (X, Y coordonnées en m) (d'après Voltz, 1986).
a) 1re composante,
b) 2e composante.

Isarithmic mapping of the components from retention curves A.C.P. (X, Y coordinates in m) (from Voltz, 1986).
a) first component,
b) second component.

sont considérées comme étant des réalisations de fonctions aléatoires. Notons X^j ces fonctions. $X^j(x)$, qui représente la valeur au point x de X^j, est alors une variable aléatoire. On dispose en certains points x_i d'une mesure de $X^j(x_i)$. Dans un premier temps, ces fonctions seront supposées stationnaires d'ordre 2. Nous aurons donc :

$$E(X^j(x)) = m^j \quad \text{avec} \quad m^j \text{ indépendant de } x$$
$$C^{jj}(h) = E((X^j(x) - m^j)(X^j(x+h) - m^j))$$
$$G^{jj}(h) = 1/2\, E((X^j(x) - X^j(x+h))^2)$$

ne dépendant que de h où h est un vecteur et x + h le vecteur obtenu par translation de x.

La fonction $C^{jj}(h)$ s'appelle fonction de covariance et $G^{jj}(h)$ variogramme. Nous supposerons aussi les fonctions de covariance croisée ainsi que les covariogrammes indépendants de x et uniquement fonction de h.

$$C^{jk}(h) = E((X^j(x) - m^j)(X^k(x+h) - m^k))$$
$$G^{jk}(h) = 1/2\, E(X^j(x) - X^j(x+h)) \times$$
$$\times (X^k(x) - X^k(x+h)).$$

Ces fonctions traduisent la structure spatiale multidimensionnelle, par exemple $C^{jk}(h)$ nous renseigne sur la liaison entre une variable j mesurée en un point x et la variable k mesurée en un point obtenu par translation de x.

Ces hypothèses de stationnarité à l'ordre 2 sont essentielles à l'interférence statistique des fonctions de structures lorsqu'on ne dispose que d'une seule réalisation. Dans le cas où les moyennes ne peuvent être supposées constantes, il sera nécessaire d'ajuster une dérive c'est-à-dire une fonction de x représentant les variations à grande échelle du phénomène et apparaissant comme une tendance de type déterministe à notre échelle d'observation. En général, il s'agit de fonctions polynômiales des coordonnées que l'on ajustera par moindres carrés, les résidus étant alors supposés stationnaires. Il faut cependant être conscient que de telles pratiques induisent des biais importants dans l'estimation des fonctions de structure.

Remarquons par ailleurs que l'égalité

$$G^{jj}(h) = C^{jj}(0) - C^{jj}(h)$$

ne se généralise pas dans le cas multivariable car

$$G^{jk}(h) = C^{jk}(0) - 1/2(C^{jk}(h) + C^{kj}(h)).$$

L'égalité entre $C^{jk}(h)$ et $C^{kj}(h)$ est cependant vérifiée le plus souvent et permet donc de travailler de façon équivalente sur les covariogrammes et sur les covariances croisées.

Jusqu'ici h représentait un vecteur, il est possible dans de nombreuses applications de vérifier l'isotropie du phénomène : h devient alors la distance entre les points x et x + h. Cette hypothèse, qui n'est pas indispensable, permet cependant une description plus aisée car les fonctions précédentes deviennent dans ce cas des fonctions à une seule variable.

Dans un premier temps, nous présentons l'analyse des fonctions que nous avons définies, comme moyen d'étude de la variabilité spatiale multiparamétrique. Nous verrons ensuite 2 analyses de type multivariable

qui, sur le plan descriptif, permettent de poser le problème de façon plus globale.

1. *Variogrammes et covariogrammes expérimentaux*

Dans l'étude des fonctions de covariance, le but est de trouver des structures spatiales communes aux différentes variables. L'étude portera de préférence sur les covariogrammes car elle ne nécessite pas d'estimation de m^j. Le covariogramme $G^{jk}(h)$ peut être estimé par

$$g^{jk}(h) = 1/(2\,n(h)) \sum (X^j(x) - X^j(y)) \times \times (X^k(x) - X^k(y))$$

où la somme porte sur l'ensemble des couples (x, y) qui sont distants approximativement de h, et n(h) est le nombre de ces couples. En pratique, on disposera donc d'un certain nombre de valeurs des covariogrammes pour des distances particulières. De la même façon que dans le cas unidimensionnel (GASCUEL-ODOUX, 1987), les caractéristiques (effet de pépite, portée, palier) des covariogrammes théoriques pourront être étudiées à partir des covariogrammes expérimentaux (JOURNEL & HUIJBREGTS, 1978). Les valeurs des caractéristiques ainsi que leur signification semblent devoir être utilisées avec précaution au niveau des interprétations. Par ailleurs, l'ajustement des modèles structuraux théoriques sera plus délicat que dans le cas unidimensionnel. Si on considère la matrice des covariogrammes d'élément générique $G^{jk}(h)$, elle devra être semi-définie positive pour tout pas h (JOURNEL & HUIJBREGTS, 1978). En pratique, on se restreindra généralement à une classe de modèles particuliers où $G^{jk}(h)$ sera le produit d'un modèle de variogramme (sphérique, linéaire, exponentiel) par un coefficient pouvant être négatif exprimant la liaison entre les variables j et k, ou plus généralement la somme de modèles de ce type-là.

L'information spatiale contenue dans les covariogrammes peut être redondante avec celle contenue dans les variogrammes. L'analyse des rapports $R^{jk}(h) = G^{jk}(h)/(G^{jj}(h)\ G^{kk}(h))^{1/2}$ permettra de préciser la nature spatiale des liaisons entre variables.

Sur la figure 2 en considérant 2 variables, le variogramme $G^{11}(h) = G^{22}(h)$, le covariogramme $G^{12}(h)$ et le rapport $R^{12}(h)$ ont été tracés pour 3 situations.

Pour la première on a pris $G^{11}(h) = G^{22}(h) = g(h)$ et $G^{12}(h) = 0{,}7\ g(h)$, où g est un variogramme. Le rapport $R^{12}(h)$ est alors constant. Ce modèle de corégionalisation dit intrinsèque, indique la présence d'une structure spatiale interne au champ de mesure se répercutant sur l'ensemble des variables mesurées. Ce cas apparaîtra pour des variables proportionnelles entre elles.

La seconde situation est donnée par

$$G^{11}(h) = G^{22}(h) = g_1(h) + g_2(h)$$

et $G^{12}(h) = 0{,}7\ g_2(h)$ avec $g_1(h)$ et $g_2(h)$, variogrammes. g_1 traduit la présence d'une structure strictement locale tandis que g_2 représente une structure de grande portée. Ce modèle indique la persistance d'une liaison entre les variables jusqu'à une grande distance.

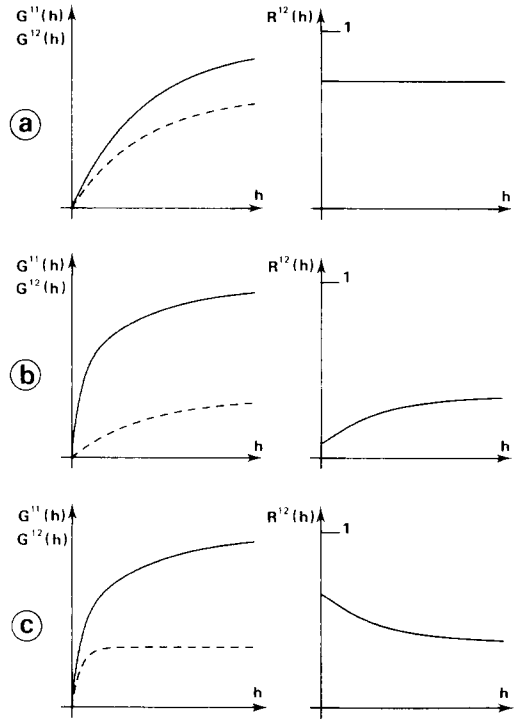

Figure 2

Exemples de modèles de corégionalisation pour 2 variables. Sur la partie gauche sont tracés le variogramme commun aux 2 variables (——) et le covariogramme associé (----).
Sur la partie droite est tracé le rapport entre le covariogramme et le variogramme des 2 variables.
a) modèle intrinsèque,
b) modèle de liaison global,
c) modèle de liaison locale.

Examples of coregionalisation models for two variables. On the left is the variogram of the two variables (——), and the associated covariogram (----).
On the right side is the ratio between the covariogram and the variogram.
a) intrinsic model,
b) model of global correlation,
c) model of local correlation.

Enfin, pour la dernière situation, nous avons choisi comme précédemment

$$G^{11}(h) = G^{22}(h) = g_1(h) + g_2(h)$$

mais $G^{12}(h) = 0{,}7\ g_1(h)$ avec les mêmes fonctions g_1 et g_2. Ce modèle traduit la présence d'une liaison entre les variables uniquement à petite distance.

Ces exemples théoriques simples montrent l'utilité des rapports $R^{jk}(h)$ qui permettent de distinguer des situations qualitativement différentes. Ceci serait plus délicat en utilisant uniquement les variogrammes et les covariogrammes, surtout qu'en pratique l'étude se fera à partir des estimateurs expérimentaux.

Résultats

L'utilisation descriptive des covariogrammes n'a encore séduit que peu de monde en sciences du sol (BRESLER *et al.*, 1984). Les problèmes relatifs à l'estimation des covariogrammes et les difficultés d'interprétation concrète expliquent ce peu d'attrait. Le plus souvent, cette étape est un simple préliminaire au cokrigeage (Mc BRATNEY & WEBSTER, 1983 ; VAUCLIN *et al.*, 1983). Il faut noter que dans toutes les

études mentionnées, aucune précaution n'est prise au niveau de l'ajustement du modèle.

2. *L'analyse des proximités spatiales*

Les idées de cette analyse sont contenues dans (LEBART, 1969 et MONESTIEZ, 1978). Elles ont été reprises dans un cadre plus géostatistique par ROYER (1984). Nous avons vu précédemment la nature de l'information contenue dans le covariogramme. Dès que le nombre des variables est important l'étude des covariogrammes devient pléthorique, d'où l'idée, pour réduire cette étude, de travailler dans l'espace des variables (cf. A.C.P.). On recherche les combinaisons linéaires des variables initiales allant de la plus structurée spatialement à la plus erratique pour un pas h donné. Pour cela, on choisit un pas h. La quantification de la structure spatiale d'une variable est donnée par le rapport entre son variogramme au pas h et sa variance. Si ce rapport est proche de 1, la variable est très erratique au pas considéré, s'il est petit devant 1, la variable est très régulière, représentant alors une composante structurée du champ. La solution mathématique du problème est obtenue par diagonalisation de V^{-1} G où V est la matrice de variance-covariance et G la matrice des variogrammes-covariogrammes au pas considéré. Les résultats de cette analyse donnent une description des variables en terme de variation locale et de variation régionale. L'analyse se fera sur des estimations des matrices V et G. La présence de dérives, c'est-à-dire des moyennes non constantes, est ici également un handicap certain.

Résultats

ROYER (1984) traite 2 ensembles de données géochimiques à titre d'exemples d'application de la méthode. Dans les 2 cas, l'analyse de proximité spatiale permet de dégager le fond régional des zones considérées, en terme de formation géologique et de caractéristiques géochimiques. Une application en science du sol n'a pas encore été menée malgré son adéquation à certains problèmes rencontrés, tel que la caractérisation des sources d'hétérogénéité physique d'une parcelle.

3. *Analyse krigeante*

L'analyse des proximités spatiales permet de séparer les variables en 2 groupes, suivant leur comportement erratique ou structuré. Cette analye s'effectue à un pas donné et n'intègre donc pas la continuité spatiale des liaisons à différentes distances. Ceci sera fait par l'analyse krigeante, proposée par MATHERON (1982). L'idée est issue de l'analyse factorielle en facteurs communs et spécifiques. Chaque variable est supposée se décomposer suivant des facteurs, indépendants entre eux, structurés spatialement. C'est-à-dire :

$$X^j(x) = \sum_k \sum_l a_k^{jl}\, Y_k^l(x) \quad \text{avec}$$

$$\begin{aligned} E((Y_k^l(x) - Y_k^l(x+h))(Y_j^i(x) - Y_j^i(x+h)) &= \\ &= 2\, g_k(h) \quad \text{si} \quad j = k \quad \text{et} \quad l = i \\ &= 0 \qquad\qquad \text{sinon} \end{aligned}$$

où $g_k(h)$ est un variogramme.

Cette décomposition de chaque variable est équivalente à la décomposition de la matrice des covariogrammes

$$G(h) = S_1\, g_1(h) + \cdots + S_r\, g_r(h)$$

avec S_k matrice d'élément générique

$$S_k^{ij} = \sum_l a_k^{il}\, a_k^{jl}\,.$$

La méthode nécessite alors 3 étapes :

— choix des structures g_k et ajustement d'un modèle pour G(h) du type exposé précédemment avec S_k semi-définie positive pour k – 1, ..., r ;

— décomposition de chacune des matrices sous la forme AA′, où A′ est la transposée de A. Chacune des matrices S_j est assimilable à une matrice de variance-covariance entre les variables. Le schéma général de l'A.C.P. (ESCOUFIER, 1987 ; GAILLIEZ & PAGES, 1976) peut être utilisé pour obtenir cette décomposition. Les liaisons les plus importantes entre les variables mesurées pour une structure donnée seront ainsi dégagées ;

— estimation aux points de mesure des composantes structurées Y_k^i par utilisation particulière du cokrigeage en utilisant les valeurs des variables mesurées. Ceci permettra de repérer les points qui sont liés plus ou moins à chacune des structures.

Résultats

WACKERNAGEL (1985, 1986) a utilisé l'analyse krigeante sur plusieurs ensembles de données issues de prospections géochimiques ou pédologiques. Les exemples traités ne montrent pas d'avantages déterminants de l'analyse krigeante par rapport aux méthodes précédentes. Ceci est à relier à la difficulté d'aboutir à un choix objectif des structures g_k dont dépendent les résultats.

4. *Conclusions*

Les concepts de base des méthodes liées à la géostatistique paraissent très intéressants d'un point de vue théorique. Le premier frein qu'est leur mise en œuvre relativement lourde est accentué par les problèmes d'estimation évoqués précédemment (respect des contraintes sur les modèles ajustés, validité des hypothèses). L'interprétation qui en résulte, reste, en dehors de quelques cas d'école, sommaire et peu concluante. Bien qu'il semble nécessaire de passer par cette description pour la transposition spatiale de modèles locaux de transfert hydrique, le manque d'expérience concluante à cet effet rebute les utilisateurs potentiels.

C. Le facteur d'échelle : une analyse non linéaire

Les méthodes précédentes cherchaient des liaisons linéaires entre les variables. Si elle est la plus simple, cette recherche n'est pas obligatoirement adéquate. Dans le cadre non linéaire, il est nécessaire de proposer la forme analytique du modèle à utiliser : c'est la démarche d'un certain nombre d'auteurs (WARRICK *et al.*, 1977 ; SIMMONS *et al.*, 1979 ; RUSSO & BRESLER, 1980 ; VAUCLIN, 1982 ; DOUPEUX, 1980) qui utilisent

pour cela le concept de milieux poreux semblables introduit par MILLER & MILLER (1955).

Deux milieux poreux sont qualifiés de semblables si leur géométrie interne diffère au plus d'un facteur d'échelle ; c'est-à-dire, qu'à l'échelle microscopique tous les détails géométriques d'un des milieux peuvent être multipliés par une constante pour obtenir les détails géométriques du deuxième (KLUTE & WILKINSON, 1958). D'après ce concept, en faisant l'hypothèse d'invariance des coefficients de tension superficielle et de viscosité cinématique, on peut écrire l'équation suivante :

$$W_r(\theta) = (a_i)^p W_i(\theta) \qquad (1)$$

pour W propriété hydrodynamique donnée
avec a facteur d'échelle
r indice du site de référence
i indice d'un site de mesure quelconque
p exposant dépendant de la propriété étudiée (VAUCLIN, 1982) :
— – 1 pour le potentiel matriciel,
— 2 pour la conductivité hydraulique,
— 1/2 pour la sorptivité,
— 1 pour la diffusion capillaire.

Il apparaît que, pour un sol, obéissant à ce concept de similitude, l'analyse de la variabilité spatiale de l'ensemble des propriétés hydrodynamiques peut se ramener à celle d'un seul paramètre : le facteur d'échelle a. Ce dernier prend pour chaque site une valeur unique, indépendante des propriétés envisagées. Pour 2 propriétés W et X, respectivement d'exposant p et k, on obtient

$$W_i(\theta) = W_r(\theta)(X_i(\theta)/X_r(\theta))^{p/k} . \qquad (2)$$

Si le milieu poreux ou site de référence est entièrement caractérisé du point de vue de ses propriétés hydrodynamiques, cela permet de définir des relations pour l'ensemble des propriétés sur tous les sites. Par la seule mesure de la variabilité spatiale d'une propriété et la caractérisation complète d'un site, on décrit totalement la variabilité des propriétés sur une surface donnée.

Résultats

Bien que séduisant au niveau théorique, le concept de similitude est d'une application nettement moins attrayante. Un certain nombre d'auteurs a examiné son domaine de validité. KLUTE & WILKINSON (1958), WILKINSON & KLUTE (1959) et ELRICK *et al.* (1959) ont montré sur des modèles de laboratoire que la mise en facteur d'échelle était possible dans le cas de sables et limons grossiers non structurés, mais que la présence de colloïdes mettait en défaut la théorie. Ces contraintes amènent à penser que peu de types de sols pourront satisfaire à l'hypothèse de similitude. Quelques auteurs (WARRICK *et al.*, 1977 ; RUSSO & BRESLER, 1980) ont testé cette méthode *in situ*, en aménageant la relation (1), de façon à éviter le problème de l'invariance de la teneur en eau à saturation. Pour cela, ils supposent que la relation (1) est également vérifiée quand les propriétés sont exprimées en fonction du taux de saturation :

$$S = \theta/\theta_{sat} .$$

Ces auteurs ont examiné plus particulièrement l'unicité du facteur a_i. Ils ont estimé la valeur des a_i pour des sites ponctuels d'une surface, successivement à partir de la courbe de rétention et de la relation conductivité hydraulique-taux de saturation. L'estimation des a_i se fait, après modélisation de la propriété étudiée X(S) par une fonction $f(S, t_1, ..., t_k)$ (où $t_1, ..., t_k$ sont les paramètres du modèle), par minimisation simultanée en $a_i, t_1, ..., t_k$ de

$$\sum_i \sum_j (f(S, t_1, ..., t_k) - (a_i)^p X_i(S_j))^2$$

où $X_i(S_j)$ est l'observation de X(S) au point i du champ et au taux de saturation S_j. Il n'apparaît pas systématiquement d'égalité entre les a_i, comme le montre les résultats de AHUJA *et al.* (1984) (fig. 3) et de RAO *et al.* (1983). Le concept du facteur d'échelle ne peut donc être généralisé à tout sol et son emploi demande une vérification de l'hypothèse de similitude. Dans le cas de non-similitude la relation (2) entre propriétés hydrodynamiques n'est plus valable. La variabilité spatiale de l'ensemble des propriétés ne peut donc plus être estimée à partir de l'étude spatiale d'une seule propriété.

D. Conclusions

Mis à part le thème du facteur d'échelle, peu de travaux ont porté sur la variabilité spatiale multiparamétrique des propriétés hydriques du sol. Trois raisons peuvent expliquer ce constat :

— L'une est que le nombre de mesures nécessaires à ce type d'étude est considérable et constitue un frein sérieux.

— L'autre est lié au type de résultats escomptés. Comme le signale GASCUEL-ODOUX (1987), au sujet des structures spatiales des propriétés hydrodynamiques, leur généralisation n'est guère possible, intégrant des conditions expérimentales très diverses (méthodes de mesure, contexte pédologique, taille de

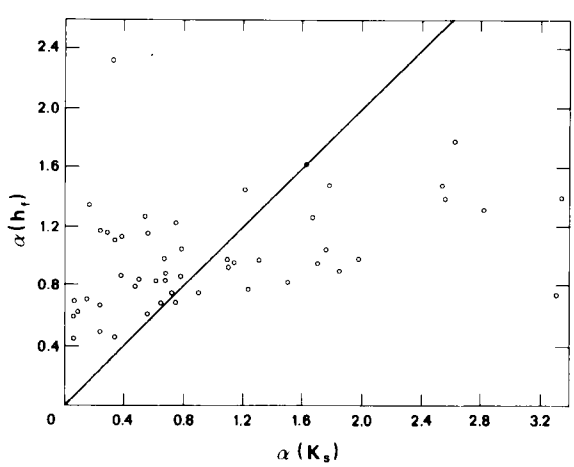

Figure 3

Exemple de remise en cause de l'unicité du facteur d'échelle α pour 2 variables (d'après Ahuja et al., *1984) :*
h_f : potentiel matriciel du front d'infiltration,
K_s : conductivité hydraulique à saturation.

Example of departure from a 1 : 1 relationship in the scaling factor α for two variables (from Ahuja et al., *1984) :*
h_f : wetting front pressure head,
K_s : saturated hydraulic conductivity.

la surface d'étude). Cette conclusion est *a fortiori* valable dans le cas d'études multiparamétriques.

— La dernière est d'ordre méthodologique. On a vu en effet que l'A.C.P., si elle est d'un abord très facile, n'utilise pas l'information spatiale. D'un autre côté, les méthodes géostatistiques mieux adaptées posent souvent des problèmes d'identification et ne sont employées que comme étapes préalables aux méthodes d'estimation que nous allons voir.

III. MÉTHODES D'INTERPOLATION MULTIPARAMÉTRIQUE

Le but de ces méthodes est de permettre l'estimation de la valeur d'une variable Z en un point non mesuré en utilisant l'information apportée par d'autres variables mesurées sur le champ spatial. L'approche multiparamétrique cherche à améliorer la qualité de l'estimation en exploitant les liaisons pouvant exister entre Z et d'autres variables, dont la variabilité spatiale est mieux connue, par exemple, par un échantillonnage plus dense. Deux types de méthodes peuvent être utilisés : la régression linéaire qui utilise peu le fait spatial et le cokrigeage, généralisation du krigeage au cas multivariable.

A. Régression linéaire

On dispose ici d'un échantillon d'apprentissage, composé de points du champ sur lesquels ont été mesurées simultanément la variable d'intérêt Z et d'autres variables X^j, j = 1, p, qui vont servir à prédire ou approcher Z. La régression permettra dans un premier temps de faire la description de l'échantillon d'apprentissage et ensuite la prédiction proprement dite. La méthode de régression linéaire cherche à approcher Z au point x par Z*(x) combinaison linéaire des variables X^j mesurées au point x :

$$Z^*(x) = \sum_j a_j X^j(x) + b$$

Z*(x) est cherché de telle façon qu'il soit sans biais, et minimise l'erreur quadratique moyenne, soit :

$$E(Z^*(x) - Z(x)) = 0$$
$$E((Z^*(x) - Z(x))^2) \text{ minimum .}$$

A partir des estimations des moyennes et de la matrice de variance-covariance, on obtient donc une estimation des coefficients cherchés. On notera cependant que les points d'observation étant liés spatialement, il serait nécessaire que la méthode d'estimation, des moyennes et de la matrice de variance-covariance, en tienne compte. En pratique, c'est la procédure classique (utilisation des estimateurs empiriques de la moyenne et de la matrice de variance-covariance) qui est utilisée, car la prise en compte de ces problèmes demanderait l'étude géostatistique préalable. La qualité globale de l'approximation de la variable sur l'échantillon d'apprentissage est donnée par le coefficient de détermination R2. L'approximation est d'autant meilleure que R2 est proche de 1, mais, pour que cette qualité reste valable hors de l'échantillon, il faut s'assurer de l'indépendance spatiale des résidus. Dans le cas de la figure 4, il apparaît que cette condition n'est pas respectée.

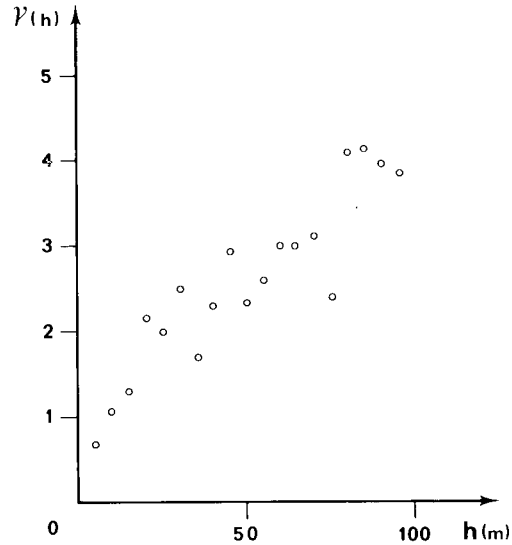

Figure 4

Mise en évidence par le variogramme expérimental γ (γ fonction de la distance h) de corrélation spatiale pour le résidu d'une régression entre variables régionalisées. (Teneur pondérale en eau à − 200 kPa en fonction de la fraction d'argile.)

Evidence of spatial dependence in residuals of a regression between regionalized variables by variogram γ (γ function of distance h) analysis. (Example of a regression between water content at − 200 kPa and clay content.)

L'estimation en tous points de Z se fait directement en utilisant la régression si les variables X^j sont connues exhaustivement sur le champ. Le plus souvent en pratique, ce n'est pas le cas, l'utilisation de la régression doit être associée à une interpolation univariable. Pour ce faire, soit les X^j sont interpolées séparément avant prédiction (cas le plus général), soit, lorsqu'on dispose de suffisamment de points où sont connues simultanément les X^j en plus de l'échantillon d'apprentissage, on procède à l'interpolation de Z*.

Résultats

Les principales variables retenues pour expliquer les caractères hydrodynamiques du sol sont les différentes fractions granulométriques classiques (argile, limon, sable), la teneur en matière organique, la densité apparente ainsi que des indicateurs de la salinité. Le coefficient de détermination des régressions pour ces paramètres (NIELSEN *et al.*, 1973 ; BYERS & STEPHENS, 1983) est généralement inférieur à 0,4. Seuls VAUCLIN *et al.* (1983) et BRESLER *et al.* (1984) ont déterminé des liaisons plus fortes, respectivement dans les cas de la prédiction de la teneur en eau à pF 2,5 par la teneur en sable (R2 = 0,69) et de la prédiction de la conductivité hyraulique par un modèle linéaire incluant la conductivité électrique et la teneur en sable (R2 = 0,60). Les liaisons constatées dans les études portant sur des volumes pédologiques restreints sont bien inférieures à celles mises en évidence pour des échantillons présentant une grande amplitude texturale (MASON *et al.*, 1957 ; HALL *et al.*, 1977 ; GUPTA & LARSON, 1979 ; RAWLS *et al.*, 1982 et 1983). Ce résultat indique que l'erreur d'approximation des paramètres hydrodynamiques par des paramètres d'accès aisé devient trop importante quand on abaisse

le domaine d'échantillonnage au niveau de la parcelle ou de l'unité de sol. Ainsi AHUJA *et al.* (1985) ont constaté l'inefficacité à l'échelle d'un bassin versant (1,6 ha) de modèles linéaires obtenus par RAWLS *et al.* (1982 et 1983) sur un échantillon de 2 543 horizons de 500 types de sol.

B. Cokrigeage

On se replace ici dans les conditions exposées au paragraphe (II-B) au sujet de la stationnarité des variables et de la symétrie des fonctions de covariance. Les moyennes des variables sont supposées constantes. On se propose d'estimer la variable Z en un point non mesuré x par $Z^*(x)$, combinaison linéaire des mesures de Z en d'autres points du champ et des mesures d'autres variables $X^1, X^2, \ldots, X^P$ en divers points du champ :

$$Z^*(x) = \sum_i a_i Z(x_i) + \sum_j \sum_i a_i^j X^j(x_i^j)$$

où x_i désigne un point d'observation de Z et x_i^j un point d'observation de X^j.

$Z^*(x)$ sera choisi sans biais et minimisant l'erreur quadratique moyenne.

A partir des estimations des fonctions de covariance et des covariances croisées, on obtient une estimation des coefficients. On devra donc passer par une étude descriptive des covariances et ajuster des modèles structuraux. Les problèmes évoqués au paragraphe (II-B-1) sont évidemment à prendre en compte. La variance d'erreur de prédiction s'exprime comme dans le krigeage par une combinaison linéaire, fonction des coefficients et des covariances des différentes variables avec Z(x), et ceci peut être utilisé pour l'échantillonnage (Mc BRATNEY & WEBSTER, 1983). Il est possible, comme pour le krigeage, d'obtenir par cokrigeage, des estimations d'intégrale sur un domaine. L'interpolation sera généralement faite par voisinage glissant, ne nécessitant que localement l'hypothèse de stationnarité. En pratique, on utilise rarement plus de 3 variables, y compris la variable à interpoler ; au-delà, les problèmes d'estimation des covariogrammes rendent les gains de précision d'interpolation non-significatifs.

Résultats

Les résultats en science du sol sont relativement peu nombreux, puisque seuls Mc BRATNEY & WEBSTER (1983) ainsi que VAUCLIN *et al.* (1983) ont testé la méthode de cokrigeage. Ces auteurs en étudient l'intérêt en le comparant au krigeage avec le critère des variances théoriques d'estimation. Ainsi Mc BRATNEY & WEBSTER (1983) indiquent sur un exemple une variance de l'erreur d'interpolation par cokrigeage qui est de 20 p. 100 à 50 p. 100 inférieure à celle du krigeage, si l'on adjoint à l'échantillonnage de la variable d'intérêt, un échantillonnage de 2 à 25 fois plus dense de variables secondaires, corrélées à la précédente. VAUCLIN *et al.* (1983) constatent toutefois sur d'autres données des gains de variance inférieurs. Notons que l'avantage économique du cokrigeage vis-à-vis du krigeage dépend de la nature et de l'existence des liaisons entre variables principales et variables secondaires, mais aussi de leur différence de coût expérimental. Son utilisation ne sera « rentable » que s'il permet de diminuer pour une précision d'estimation donnée l'effort d'échantillonnage global par comparaison à celui nécessité par le krigeage.

C. Discussion

De façon théorique, on sait que le cokrigeage est meilleur, au sens de l'erreur quadratique moyenne, que la régression puisqu'il utilise plus d'information. Au niveau des applications le cokrigeage est une méthode plus générale permettant de traiter des échantillonnages délicats pour la régression (non coïncidence des points de mesure pour les différentes variables). Dans les cas où les méthodes s'appliquent bien, les comparaisons tendent à favoriser le cokrigeage (Mc BRATNEY & WEBSTER, 1983 ; VAUCLIN *et al.*, 1983) cependant, ces résultats doivent être modulés car la comparaison se fait sur des variances d'erreur de prédiction théoriques (les hypothèses faites pour chacune des méthodes ne sont en général pas directement vérifiables). Aller plus loin dans le choix des méthodes nécessiterait un plus grand nombre d'études, réunissant des conditions d'expériences diverses, et l'emploi systématique de procédures de validation croisée sur des échantillons tests.

La variabilité spatiale multiparamétrique est importante pour améliorer la cartographie d'une variable en s'aidant d'autres paramètres mesurés sur le champ. Ce besoin sera notamment ressenti lors de la cartographie des paramètres hydrodynamiques du sol, qui sont d'accès peu aisé sur le plan expérimental. La recherche de liaisons spatiales entre ces paramètres et des variables d'accès plus aisé a été entreprise par un certain nombre d'auteurs qui ont privilégié l'étude de liaisons non linéaires fondée sur le concept de facteur d'échelle. L'analyse des liaisons linéaires a été réduite du fait des problèmes méthodologiques déjà soulignés ; de plus, les corrélations déterminées sur des volumes pédologiques restreints se sont révélées faibles. Aussi, peu de travaux ont été motivés dans le sens d'une cartographie des paramètres hydrodynamiques à partir de variables d'accès aisé.

Quant aux problèmes méthodologiques, cette revue bibliographique nous a montré plusieurs points.

— Tant au niveau de la description que de l'estimation, les méthodes géostatistiques se révèlent conceptuellement très adaptées mais leur application pratique n'est efficace que dans le cas d'un petit nombre de variables régionalisées stationnaires.

— Pour un nombre important de variables, il apparaît que les méthodes multidimensionnelles d'analyse des données sont encore les seules utilisables d'un point de vue concret. Elles peuvent être utilisées pour dégager les variables qui seront soumises ensuite à l'étude géostatistique.

— Les questions qui restent posées sont d'une part la quantification de la qualité des méthodes classiques appliquées aux variables régionalisées et d'autre part l'application des méthodes géostatistiques dans un cadre plus général.

Reçu le 27 novembre 1986.
Accepté le 2 juin 1987.

RÉFÉRENCES BIBLIOGRAPHIQUES

Ahuja L. R., Naney J. W., Nielsen D. R., 1984. Scaling soil water properties and infiltration modeling. *Soil Sci. Soc. Am. J.,* 48, 970-973.

Ahuja L. R., Naney J. W., Williams R. D., 1985. Estimating soil water characteristics from sampler properties or limited data. *Soil Sci. Soc. Am. J.,* 49, 1100-1105.

Bouroche J. M., Saporta G., 1980. *L'analyse des données.* P.U.F., collection que-sais-je ?

Bresler E. S., Dagan G., Wagenet R. J., Laufer A., 1984. Statistical analysis of salinity and texture effects on spatial variability of soil hydraulic conductivity. *Soil Sci. Soc. Am. J.,* 48, 16-25.

Byers E., Stephens D. B., 1983. Statistical and stochastic analysis of hydraulic conductivity and particle-size in a fluvial sand. *Soil Sci. Soc. Am. J.,* 47, 1072-1081.

Cailliez F., Pages J. P., 1976. *Introduction à l'analyse des données.* S.M.A.S.H., 9 rue Duban, 75016 Paris.

Doupeux C., 1980. *Essai de description unidimensionnelle de la variabilité spatiale des propriétés hydrodynamiques d'un sol* in situ. Thèse 3e cycle, Paris-Sud, Orsay, 113 p.

Elrick D. A., Scandrett S. H., Miller F. E., 1959. Tests of capillary flow scaling. *Soil Sci. Soc. Am. Proc.,* 23, 329-332.

Escoufier Y., 1987. The duality diagram : a means for better practical applications. In *"Developments in numerical ecology"* Legendre P. & Legendre L. (eds.) ; Nato advanced study institute series G (ecological sciences). Springer Verlag, Berlin.

Gascuel-Odoux C., 1984. *Applications de la géostatistique à l'étude de la variabilité spatiale des propriétés hydriques du sol.* Thèse docteur-ingénieur ENSM Paris et ENSA-INRA Rennes, 235 p.

Gascuel-Odoux C., 1987. Variabilité spatiale des propriétés hydriques du sol, méthodes et résultats, cas d'une seule variable : revue bibliographique. *Agronomie,* 7 (1), 61-71.

Girard M. C., 1983. *Recherche d'une modélisation en vue d'une représentation spatiale de la couverture pédologique.* Thèse Doct. d'état-*Sols,* 12, 430 p.

Gupta S. C., Larson W. E., 1979. Estimating soil water retention characteristics from particle size distribution, organic matter content and bulk density. *Water Resour. Res.,* 15, 1633-1635.

Hall D. G. M., Reeve M. J., Thomasson A. J., Wright V. F., 1977. *Water retention, porosity and density of field soils.* Soil Survey Technical monograph 9, Harpenden (G.B.), 76 p.

Journel A. G., Huijbregts Ch. J., 1978. *Mining geostatistics.* Academic press. New York, 600 p.

Klute A., Wilkinson G. E., 1958. Some tests of the similar media concept of capillary flow. I. Reduced capillary conductivity and moisture characteristic data. *Soil Sci. Soc. Am. Proc.,* 22, 278-280.

Lebart L., 1969. *Analyse statistique de la contiguïté.* Publication I.S.U.P., 18, 81-112.

Mason D. D., Lutz J. F., Petersen R. G., 1957. Hydraulic conductivity as related to certain soil properties in a number of great soil groups. Sampling errors involved. *Soil Sci. Proc.,* 21, 554-560.

Matheron G., 1982. *Pour une analyse krigeante des données régionalisées.* Centre de Géostatistique et de Morphologie Mathématique, Fontainebleau, rapport N-732.

Mc Bratney A. B., Webster R., 1983. How many observations are needed for regional estimation of soil properties ? *Soil Sci.,* 135, 177-183.

Miller E. E., Miller R. D., 1955. Theory of capillary flow. I. Practical implications. II. Experimental information. *Soil Sci. Soc. Am. Proc.,* 19, I : 267-271, II : 271-275.

Monestiez P., 1978. *Présentation de 2 méthodes utilisant la notion de contiguïté pour l'analyse des données géographiques.* Thèse 3e cycle, Univ. Pierre et Marie Curie, Paris, 64 p.

Nielsen D. R., Biggar J. W., Erh T., 1973. Spatial variability of field measured soil water properties. *Hilgardia,* 42, 215-259.

Peck A. J., 1983. Field variability of soil physical properties. *Adv. Irrigation,* 2, 189-200.

Rao C. R., 1980. Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In *Multivariate Analysis V.* Ed. by P. R. Krishnaiah. North-Holland Publishing Company : 3-22.

Rao P. S. C., Jessup R. E., Hornsby A. C., Cassel D. K., Pollans W. A., 1983. Scaling soil microhydrologic properties of Lakeland and Konawa soils using similar media concepts. *Agr. Water Manage,* 6 (2/3), 277-290.

Rawls W. J., Brakensiek D. L., Saxton K. E., 1982. Estimation of soil water properties. *Trans. ASAE,* 25, 1316-1320, 1328.

Rawls W. J., Brakensiek D. L., Soni B., 1983. Agricultural management effects on soil water processes. Part I : Soil water retention and Green and Ampt infiltration parameters. *Trans. ASAE,* 26, 1747-1752.

Royer J. J., 1984. Proximity analysis : a method for multivariate geodata processing. Application to geochemical processing. Sciences de la terre, série informatique géologique, 20. Proc. Int. Coll. *"Computers in earth sciences for natural resources characterization",* 9-13 avril, Nancy.

Russo D., Bresler E., 1980. Scaling soil hydraulic properties of a heterogeneous field. *Soil Sci Soc. Am. J.,* 44, 681-684.

Russo D., Bresler E., 1981. Effect of field variability of soil hydraulic properties on solutions of unsatured water and salt flows. *Soil Sc. Soc. Am. J.,* 45, 675-681.

Sabatier R., Jan Y., Escoufier Y., 1984. Approximations d'applications linéaires et analyse en composantes principales. In *Data Analysis and Informatics III.* Ed. by Diday *et al.* Elsevier Science Publishers BV (North-Holland) : 569-580.

Simmons C. S., Nielsen D. R., Biggar J. W., 1979. Scaling of field measured soil water properties. 1. Methodology ; 2. Hydraulic conductivity and flux. *Hilgardia,* 47, 77-174.

Vauclin M., 1982. *Méthodes d'études de la variabilité spatiale des propriétés d'un sol.* Ed. I.N.R.A., Publ. : les colloques de l'I.N.R.A., 15, 9-43.

Vauclin M., Viera S. R., Vachaud G., Nielsen D. R., 1983. The use of cokriging with limited field soil observations. *Soil Sci. Soc. Am. J.,* 47, 175-184.

Voltz M., 1986. *Variabilité spatiale des propriétés physiques du sol en milieu alluvial.* Thèse docteur ingénieur, ENSA Montpellier, 220 p.

Wackernagel H., 1985. *L'inférence d'un modèle linéaire en géostatistique.* Thèse ENSM, Paris, 100 p.

Wackernagel H., 1986. Geostatistical techniques for interpreting multivariate spatial information. *NATO-ASI* 22 juin-4 juillet 1986, Italie.

Warrick A. W., Mullen G. J., Nielsen D. R., 1977. Scaling field-measured soil hydraulic properties using similar media concept. *Water Resour. Res.,* 13, 355-362.

Webster R., 1977. Spectral analysis of gilgai soil. *Aust. J. Soil Res.,* 15, 191-204.

Wilkinson G. E., Klute A., 1959. Some tests of the similar media concept of capillary flow. II. Flow system data. *Soil Sci. Soc. Am. Proc.,* 23, 434-437.