

Application de deux méthodes statistiques multivariées à la cartographie des sols.

Douaoui A.

Institut d'Agronomie, Centre universitaire BP 151, Chlef

Douaoui A., 1995 – Application de deux méthodes statistiques multivariées à la cartographie des sols. Ann. Agron. I.N.A., Vol. 16, N° 1 et 2, pp.113 – 126

Résumé : La cartographie classique des sols qui consiste à délimiter des unités pédologiques homogènes d'une zone donnée, reposait sur le choix de profils jugés représentatifs de l'unité cartographique. Plusieurs critiques ont été formulées à l'égard de cette approche tels que le seuil d'impureté des sols complexes, la variabilité spatiale des propriétés du sol et les critères de classification qui sont préétablis. Dans le but de pallier à ces critiques et de représenter le plus fidèlement possible l'organisation spatiale du sol et son hétérogénéité, deux méthodes statistiques multivariées ont été retenues. La première méthode est l'analyse en composante principale (ACP). C'est une méthode essentiellement descriptive ; son objectif est de présenter, sous forme graphique, le maximum de l'information contenue dans un tableau de données. La deuxième méthode, prenant en compte les corrélations spatiales entre les observations, est une combinaison entre la géostatistique et l'ACP. Elle permet de cartographier les valeurs des coordonnées des individus sur les différents axes pris en considération.

Mots clés : cartographie des sols , ACP, géostatistique.

Application of two multivariate statistical methods in soil survey

Abstract : Classical soil survey which draw the homogeneous pedological map unit boundaries of a given area is mainly based on the profile type representing the cartographic map unit. Several critical appreciations were made towards this approach and the percentage of impurity of complex map unit. Also the spacial variability and classification criterium were established before mapping. From these critical appreciations and trying to map accurately the spacial organization of soils and its heterogeneous characteristic two multivariate statistical methods were

experimented. The first method is the ACP which is mainly descriptive, and graphic with maximum information presented on tables. The second method is a combination of ACP and geostatistic which study the spacial correlation between data observed. This last method permits to draw the coordinated values of individuals on different axis.

Key words : soil survey, ACP, geostatistic

INTRODUCTION

L'impossibilité d'analyser ou même d'observer le sol en chaque point de l'espace pose un problème de précision dans la cartographie, surtout que les propriétés physico-chimiques varient d'un point à un autre. En plus le fonctionnement du sol ne dépend pas uniquement d'une seule propriété, mais de plusieurs en même temps.

Différentes approches statistiques permettent d'apprécier cette hétérogénéité spatiale et constituent un outil précieux dans la connaissance des sols.

Dans cette étude, nous utilisons l'analyse en composantes principales (ACP) et la géostatistique combinée à l'ACP. Ces deux méthodes sont des approches statistiques multivariées permettant une cartographie des sols en groupes homogènes.

Ce travail comporte donc trois parties . La première partie consiste à établir une carte pédologique classique. La deuxième partie est une application de l'ACP pour déterminer les groupes de sols homogènes. La dernière partie est une cartographie multiparamétrique par l'utilisation combinée de la géostatistique et de l'ACP.

La parcelle d'investigation fait partie de la station expérimentale de l'institut d'agronomie de Chlef.

1- Cartographie pédologique

Le principe de la démarche cartographique qui consiste à délimiter des unités pédologiques en fonction d'une classification donnée (CPCS, soil taxonomy ...) comprend un certain nombre d'étapes. Il s'agit en premier lieu de définir les critères de cartographie pour ensuite délimiter les zones homogènes. une fois ces limites établies, on choisit les profils à décrire et prélever des échantillons de terre de chaque horizon

pour les analyser au laboratoire. Après la synthèse de tous les résultats obtenus, on dressera la carte des sols.

Ainsi, l'étude cartographique de la parcelle d'étude montre l'existence de deux unités pédologiques : des sols calcimagnésiques bruns calcaires et des sols peu évolués d'apport alluvial brunifiés (fig. 1).

La transition entre les deux unités étant progressive, avec une largeur de 20 à 25 m, est donc difficile à tracer.

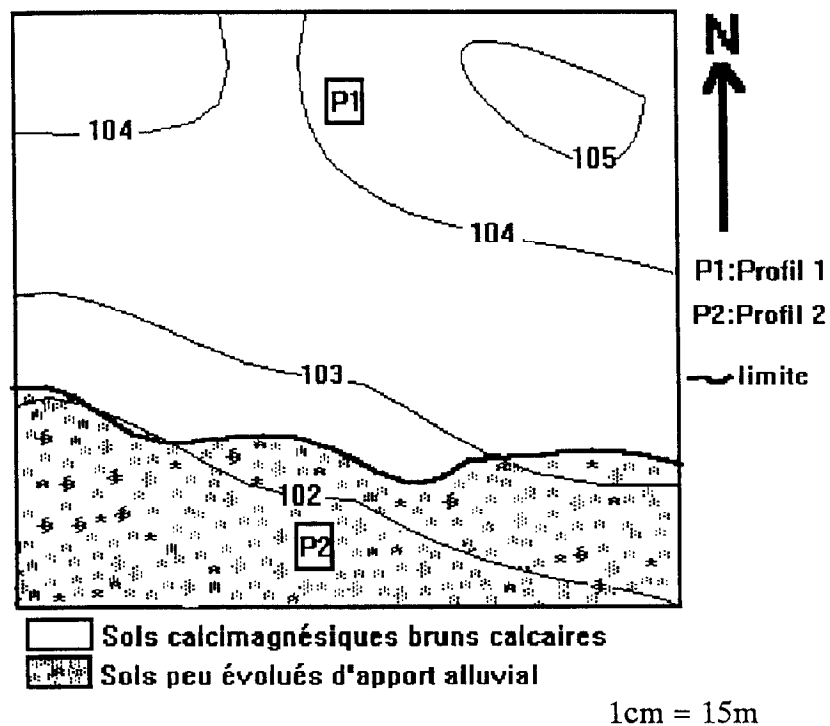


Figure 1. carte pédologique

2. Méthodes statistiques

2.1. échantillonnage

Le dispositif de prélèvement a été conçu pour répondre particulièrement aux soucis de la géostatistique. Il est caractérisé par une augmentation de l'intervalle entre

les sondages de 5 m, 10m et 20 m. Pour chaque sondage, on prélève deux échantillons ; un échantillon de l'horizon de surface et un second échantillon de profondeur à partir de 40 cm. L'ensemble des sondages est au nombre de 90.

Les variables étudiées sont : les argiles (A), limons fins (LF), limons grossiers (LG), limons totaux (LT), sables fins (SF), sables grossiers (SG) et sables totaux (ST), le calcaire total (CaCO₃), le pH et la salinité (CE). Ils ont été déterminés selon les méthodes classiques préconisées par l'AISS.

Les indices 1 et 2 correspondent respectivement au premier et au second horizon. Le nombre total de variables étudiées est de 20.

2. 2. Analyse en composantes principales

2.2.1. Aperçu bibliographique

Différentes approches permettent d'établir des relations entre des variables de positions dans l'espace, considérées en général comme individus statistiques ; elles cherchent à décrire l'hétérogénéité spatiale du sol par l'étude simultanée de plusieurs variables (Voltz, 1986 ; Goulard & al., 1987 ; Walter, 1990).

Dans notre cas, on utilise l'analyse en composantes principales qui est une méthode statistique essentiellement descriptive ; son objectif est de présenter, sous forme graphique, le maximum de l'information contenue dans un tableau de données (Philippeau, 1992).

Ce tableau de données est constitué par des colonnes correspondant à des variables (ici, ce sont les variables granulométriques et chimiques) et des lignes correspondant aux individus (ici, ce sont les échantillons)

Ces deux ensembles (individus et variables) n'ont pas de rôles symétriques comme c'est le cas des correspondances. En réalité, ils ne sont pas dans le même espace (Fenelon, 1981).

Il s'agit donc de trouver des axes (composantes principales) qui sont des combinaisons linéaires de variables de départ à variance maximum non corrélés entre eux (Goulard & al., 1987).

Les résultats de l'ACP étant influencés par les unités différentes des variables (Philippeau, 1992), nous avons donc donné le même poids (même importance) aux différentes variables en travaillant sur des variables centrées et réduites, c'est à dire en les divisant par leur écart type respectif.

Plusieurs auteurs ont mis cette méthode au service de la pédologie. En plus de ceux déjà cités ci-dessus, on retrouve notamment, Girard (1983), Oliver (1984), Arrouays (1987) et Girard & King (1988).

2.2.2. Résultat et discussion

Il s'agit en premier lieu d'observer les valeurs propres des composantes principales. Ces valeurs représentent la variance des individus sur l'axe correspondant, c'est à dire la part de la variation expliquée par chaque axe. Dans un souci d'interprétation, ces variation sont exprimées en pourcentage de la somme des valeurs propres totales (tab.I).

Dans cet exemple, la première composante, égale à 5.42, explique 27,09 % de la variance initiale. La deuxième composante, égale à 2,86 , explique 14.34 % de la variance initiale (tab.I).

Tableau I. Les valeurs propres

Numéro	Valeur propre	pour-cent.	pour-cent. cumule
1	5.4186	27.09	27.09
2	2.8682	14.34	41.43
3	2.1824	10.91	52.35
4	1.8722	9.36	61.71
5	1.5271	7.64	69.34
6	1.2691	6.35	75.69
7	1.0100	5.05	80.74
8	0.8293	4.15	84.88
9	0.6376	3.19	88.07
10	0.5568	2.78	90.86

Les différents axes n'étant pas corrélés entre eux, la part de l'information expliquée par deux composantes principales est donc égale à la somme des pourcentages exprimés par ces deux axes.

Dans ce cas, le plan 1 formé par les deux premiers axes explique 41.43 % de l'information totale et le plan 2 formé par le premier et le troisième axes, explique 38 % de l'information totale. on a donc atteint au total un cumul de 52,35 % (tab. I).

2.2.2.1. Etude des variables

A partir du tableau II qui donne la corrélation entre les variables et les axes ainsi que les carrés respectifs, il en ressort que :

- ce sont principalement les variables granulométriques qui contribuent le plus à la formation des principaux axes, en particulier les limons, pour les variables chimiques, c'est le calcaire total qui a la plus grande part,

- les limons, en général, sont corrélés négativement aux argiles, c'est à dire lorsque ces derniers augmentent la fraction limoneuse diminue, la réciproque est vérifiée,

- les sables ne sont que très faiblement corrélés aux fractions granulométriques, mis à part les SGH2 et les STH2 qui semblent augmenter avec les argiles et diminuer avec les limons,

- le CaCO_3H_2 s'oppose aux différentes fractions sableuses du même horizon.

2. 2. 2. 2. Etude des individus

La figure 2, représentant le plan formé entre les deux premières composantes principales, montre que les individus se répartissent comme suite :

- tous les individus appartenant au sol peu évolué se positionnent à gauche de l'axe 2.

- les individus ne faisant pas partie du sol peu évolué et qui se situent à gauche du second axe, ont une position un peu particulière, puisqu'ils se situent, généralement, le long de cet axe et à sa proximité.

Tableau II. corrélations des variables aux axes

variables	AXE 1	AXE 2	AXE3	AXE 4
AH1	0.72	-0.073	-0.027	-0.10
LGH1	-0.25	0.23	0.82	0.11
LTH1	-0.87	-0.22	0.024	-0.03
SFH1	0.5	0.47	0.12	0.47
SGH1	0.44	0.42	-0.24	-0.24
STH1	0.54	0.6	-0.06	0.16
AH2	0.63	0.24	-0.037	-0.34
LFH2	-0.58	0.4	-0.29	0.14
LGH2	-0.44	-0.28	0.44	0.036
LTH2	-0.87	0.10	0.12	0.14
SFH2	0.29	-0.48	0.2	0.55
SGH2	0.62	-0.24	-0.26	-0.19
STH2	0.61	-0.55	-0.057	0.27
CaCO3H1	-0.59	0.74	0.026	-0.3
CaCO3H2	-0.2	0.73	0.025	-0.3
pHH1	0.36	-0.19	0.48	-0.49
pHH2	0.2	-0.33	0.41	-0.6
CEH1	-0.33	0.29	-0.034	0.0006
CEH2	-0.11	0.39	0.15	-0.3

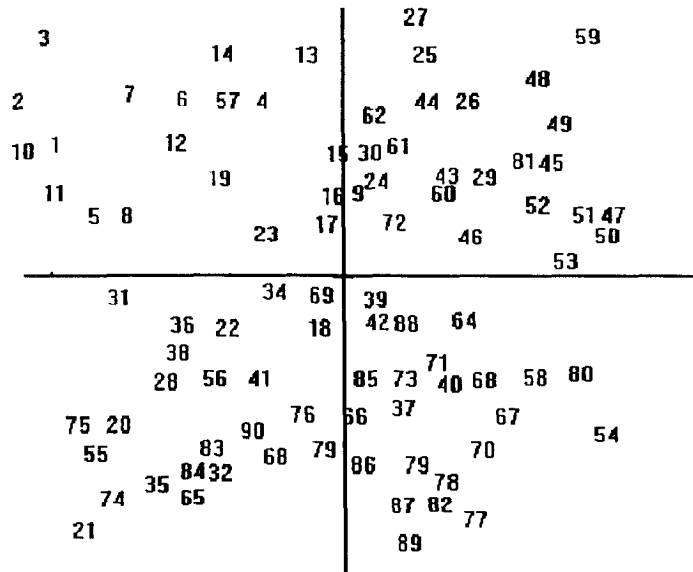


Figure 2 . plan principal des individu

- les quatres compartiments A, B, C, D (fig. 2) représentent quatre groupes différents se répartissant dans la parcelle d'étude en zones différentes, ainsi chaque compartiment correspond à une zone pédologique homogène continue (fig. 3). Les zones A et B regroupent les sols peu évolués et une petite partie des sols calcimagnésiques alors que les zones C et D font partie des sols calcimagnésiques.

- le problème du seuil d'impureté ne s'est pas posé, dans la mesure où le traçage de la limite entre les zones à été fait de façon plus simple.

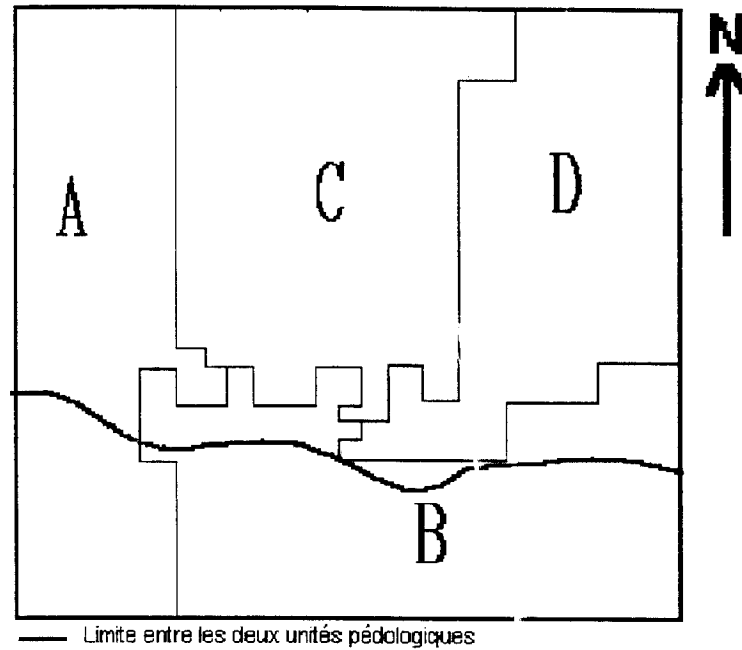


Figure 3 : Les zones homogènes d'après l'ACP

3. Cartographie multiparamétrique par krigeage

3.1. Aperçu bibliographique

Dans cette partie, nous avons tenté d'appliquer une approche géostatistique multiparamétrique regroupant l'ensemble des variables d'étude prises simultanément.

Cette méthode consiste à cartographier les valeurs des coordonnées des individus sur les différents axes pris en considération. Elle permet de dégager un fond régional des mesures, c'est à dire la tendance en grand de la parcelle d'étude (Goulard & al., 1987). Peu de travaux ont été effectués en pédologie dans ce sens. On retient le travail de Webster (1977) qui a utilisé cette méthode pour exprimer la périodicité des propriétés physico-chimiques sur des sols à relief gilgaï et celui de Voltz (1986) qui a aussi utilisé cette méthode pour des paramètres hydrodynamiques.

D'autres auteurs tels que Rambal (1978) et Mc Bratney & Webster (1981) ont aussi travaillé dans ce sens.

3.2. Résultat et discussion

3.2.1. Variographie des principaux axes

Le calcul des valeurs des coordonnées des individus sur les axes principaux, pour des variables centrées réduites, se fait de la façon suivante (Exemple pour l'individu 1)

Tableau III. Exemple de calcul des variables des coordonnées

Variables initiales	Variables centrées réduites	Vecteur propre 1	Vecteur propre 2
A H 1	$\frac{36.85-40.68}{2.21} = -1.73$	0.31	-0.043
L F H 1	$\frac{20.55-21.33}{4.03} = -0.19$	-0.16	-0.21
CEH2	$\frac{0.45-0.76}{0.21} = -1.47$	-0.046	-0.22

La coordonnée sur l'axe 1 est égale à :

$$(0.31)(-1.73) + (-0.16)(-0.19) + \dots + (-0.046)(-1.47) = -4.75$$

La coordonnées sur l'axe 2 est égale à :

$$(-0.043)(-1.73) + (-0.21)(-0.19) + \dots + (0.22)(-1.47) = 1.80$$

Les valeurs des coordonnées ainsi calculés sont considérées comme étant des variables régionalisées dont on essaye de déterminer les structures par l'étude de leurs variogrammes.

Le calcul des variogrammes a été fait pour les quatres premiers axes qui représentent un total de 61.71 % de la variance initiale.

L'inférence des différents variogrammes correspondant aux différents axes a donné deux variogrammes linéaires pour l'axe 1 et l'axe 2, un variogramme aléatoire pour l'axe 3 et un variogramme à palier pour l'axe 4. Les fonctions d'ajustement sont les suivantes :

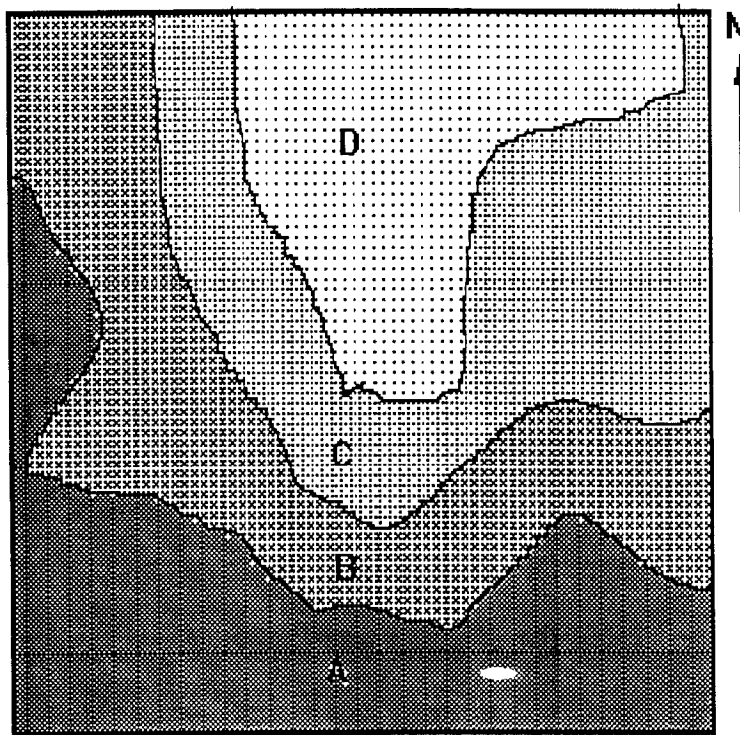
$$\text{Axe 1 : } y = 0.041 x + 0.8$$

$$\text{Axe 2 : } y = 0.036 x + 0.55$$

Axe 4 : modèle sphérique avec un effet de pépite de 0.5, un palier de 1.4 et une portée de 60 m

3. 2. 2. Cartographie des axes principaux

La cartographie de l'axe 1 (fig. 4), liée beaucoup plus aux limons et aux argiles et à un degré moindre au CaCO_3H_1 et aux sables du second horizon, présente une structure spatiale formée de quatre zones pédologiques homogènes avec un gradient dans la direction SO-NE. Ce gradient exprime le maximum d'hétérogénéité spatiale des sols de la parcelle d'étude. L'orientation de ce gradient est perpendiculaire au sens des courbes de niveau.

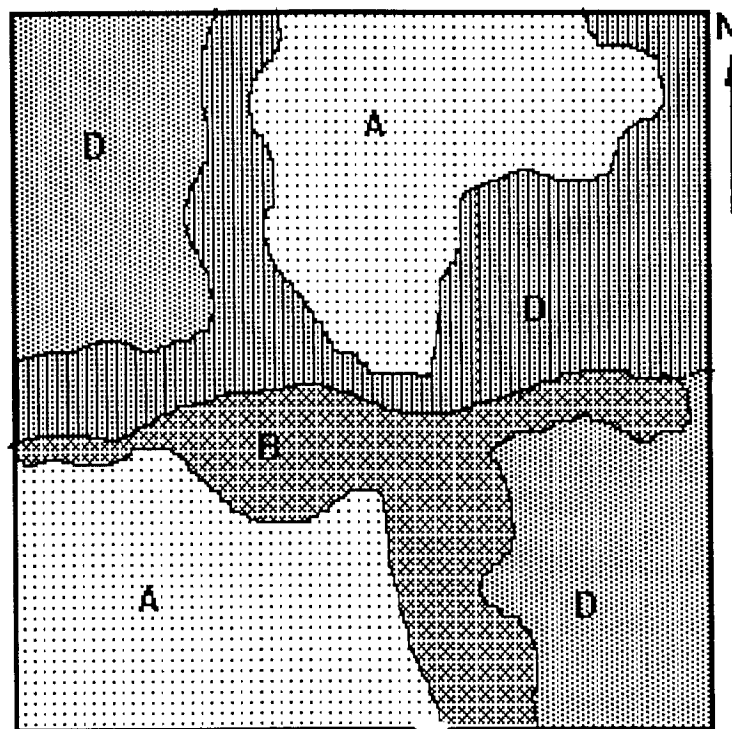


1cm = 15m

Figure 4 : carte des zones homogènes de la première composante principale

Une comparaison de cette carte avec la carte pédologique montre que la zone située au sud de la parcelle correspond dans sa plus grande partie aux sols peu évolués, elle est suivie par une autre zone qui représente la zone de transition entre les deux types de sols. Les deux dernières zones correspondent aux sols calcimagnésiques.

La cartographie de l'axe 4, liée principalement au pH et aux sables fins, présente un fond régional sans gradient apparent. Cette structure spatiale a été déjà mise en évidence par le variogramme qui présentait un palier et donc une distance d'autocorrélation. Sur cette carte, on peut voir une structure spatiale différente de celle de l'axe 1 dont l'organisation se traduit par une certaine disposition des zones homogènes selon le sens de la plus grande pente et celui de la pente la plus faible.



1cm = 15 m

Figure 5. carte des zones homogènes de la quatrième composante principale

Enfin, on peut en conclure que la cartographie géostatistique multiparamétrique qui tient compte des corrélations spatiales, montre une répartition spatiale des zones pédologiques homogènes avec une certaine organisation en relation étroite avec la topographie.

CONCLUSION

Les résultats statistiques obtenus par l'ACP n'ont pas dégagés des groupes bien distincts. Du point de vue statistique, ceci est dû à la dispersion élevée des variables et des individus. En pédologie, cette dispersion est expliquée par la transition progressive entre les unités de la parcelle. La cartographie des composantes principales par l'utilisation de la géostatistique qui prend en compte les corrélations spatiales entre les individus a permis de mieux représenter cette hétérogénéité spatiale du sol. C'est ainsi qu'on a pu voir, en plus de l'organisation spatiale du sol, que le maximum d'hétérogénéité se fait dans la direction SO-NE ce qui correspond à la ligne de la plus grande pente.

Références

ARROUAYS D. , 1987 – Cartographie des sols et comportement des données. Comparaison des données de cartographie et d'enquête agronomique. Science du sol, vol. 25 (1), 43–58.

FENELON J.P. , 1981 – Qu'est-ce que l'analyse des données ? Lefonen, 311 p.

GIRARD M.C. , 1983 – Recherche d'une modélisation en vue d'une représentation spatiale de la couverture pédologique. Thèse d'état. 430 p., INA-PG, SOLS, n° 12

GIRARD M.C. & KING D. , 1988 – Un algorithme interactif pour la classification des horizons de la couverture pédologique. Science du sol, vol. 26(2), 81–103.

GOULARD M., VOLTZ M & MONETIEZ P. , 1987 – Comparaison d'approches multivariées pour l'étude de la variabilité spatiale des sols. Agronomie, 7(9), 657–665.

Mc BRATNEY A.B. & WEBSTER R. , 1981 – Spatial dependence and classification of soil a long a transect in northeast scotland. Geodema, 26, 63–82.

OLIVER M.A. , 1984 – Soil variation in the myre forest : its elucidation and measurement. Ph. D thesis, University of Birmingham, 230 p.

PHILIPPEAU G. , 1992 – Comment interpréter les résultats d'une analyse en composantes principales ? collection STAT- ITCF, 63 p.

RAMBAL S. ,1978 – Analyse en composantes principales des variations du stock d'eau du sol. Bull. Groupe français d'humidimétrie neutronique, 3, 74–81.

VOLTZ M. , 1986 – Variabilité spatiale des propriétés physiques en milieu alluvial. Essai de cartographie quantitative des paramètres hydrodynamiques. Thèse de doct. Ing, ENSAM, 198 p.

WALTER C. , 1990 – Estimation des propriétés du sol et quantification de leur variabilité à moyenne échelle : cartographie pédologique et géostatistique dans le sud de l'Ille et Vilaine (France). Thèse de doctorat de l'université de Paris 6, 172 p.

WEBSTER R. , 1977 – Spectral analysis of gilgai soil. Australian J. Soil Res., 15 (3), 191–204.