

# Traitement informatique des données pédologiques

## La bibliothèque Logos

D. KING<sup>(1)</sup>  
O. DUVAL<sup>(1)</sup>

---

### RESUME

Les données pédologiques se présentent d'une façon spécifique, liée à l'organisation du milieu pédologique et aux méthodes mises en œuvre pour l'appréhender. Ceci nécessite la création d'outils de traitements statistiques et graphiques qui respectent ces contraintes. Un ensemble de logiciels appliqués aux traitements cartographiques est réuni dans une bibliothèque nommée LOGOS.

**MOTS CLES :** Logiciels - Cartographie des sols - Statistique - Traitement des données.

---

### COMPUTERIZED TREATMENT OF SOIL DATA T. SOFTWARE LIBRARY LOGOS

*Soil data is organized in a specific way, due to the organization of pedologic context and to the methods used to describe it. It leads to the setting up of statistical treatments and graphic data handling, which respect these constraints. A software package applied to cartographic data handling was built into a library named LOGOS.*

**KEY WORDS :** Software - Soil survey - Statistics - Data handling.

---

## INTRODUCTION

Au cours des travaux de cartographie pédologique, il est nécessaire de rechercher les relations existant entre les caractères du sol et de déterminer les observations les plus caractéristiques. Ceci permet d'élaborer des hypothèses et de proposer des modèles d'organisation spatiale de la couverture pédologique.

La présentation et le traitement des données pédologiques doivent respecter l'organisation du sol. De nombreux logiciels ont été proposés pour aider les cartographes dans leur travail d'analyse des données. A la suite des recherches menées par M.C. Girard et D. King, un ensemble de programmes ont été réunis au sein de la bibliothèque LOGOS (KING et DUVAL, 1988) ; ils permettent la mise en forme des données (compatibles avec le logiciel STIPA, BERTRAND et al., 1984), la classification des horizons ou profils, enfin les traitements graphiques pour la recherche de structures spatiales.

Les programmes proposés ont été conçus dans le cadre d'une cartographie pédologique, mais certains d'entre eux peuvent traiter d'autres types de données.

## TRAITEMENT ET SYNTHÈSE DES DONNÉES PÉDOLOGIQUES

### A) NATURE DES DONNÉES

Le sol se présente sous la forme d'un continuum dont les caractères sont variables dans l'espace et le temps. Les observations sont nécessairement discontinues (discrètes).

(1) INRA, SESCOF, Ardon, 45160 Ollivet.

Les variables utilisées par le pédologue ne sont pas toutes de même nature. On distingue des variables quantitatives et des variables qualitatives ordonnées ou non. Ces variables peuvent présenter des relations conditionnelles (par exemple, on ne décrit la forme des taches que lorsqu'elles sont présentes). La codification de ces différentes variables est aussi bien numérique qu'alphabétique. Les programmes doivent donc accepter la diversité de leur nature. Des options particulières ont été prévues pour le cas de données manquantes.

La couverture pédologique est structurée. Il faut tenir compte de cette organisation dans la façon de relever et de codifier les informations. Le glossaire STIPA en est un exemple : la distinction est faite entre des variables liées à l'environnement, d'autres liées à une description synthétique du profil, d'autres décrivent l'horizon et les traits pédologiques. L'horizon est l'unité élémentaire privilégiée dans LOGOS.

## B) SAISIE ET MISE EN FORME DES DONNÉES

Les fiches de description et d'analyse se présentent sous la forme de bordereaux directement encodables sur support informatique (BONNERIC et NAVARRO, 1988). Il est actuellement possible de supprimer la fiche papier de terrain et par là même, la phase encodage. Des programmes de saisie ont été conçus sur microordinateur portable. Le gain de temps est appréciable pour des travaux à grande échelle et/ou dans des milieux relativement homogènes.

La réflexion sur l'organisation des données sous la forme d'un tableau statistique de type « individus-variables » (par ex. : horizon-caractères, GIRARD et KING, 1988) est une nécessité avant tout traitement. Les programmes EXTRAC ou PTRI liés à STIPA permettent d'organiser l'information de différentes façons : découpage du profil en tranches, description du profil par des variables synthétiques ou par des horizons diagnostics... Le tableau de données étant mis en place, il y a lieu, pour un grand nombre de traitements, d'harmoniser la nature des caractères que l'on veut traiter. Des programmes sont proposés pour établir de nouvelles variables, en particulier en rapportant les données à des variables qualitatives ordonnées, ceci à l'aide d'histogrammes (programme CODA). Le programme MATH permet le calcul de nouvelles variables à partir des variables initiales (fonction de type tableur). Enfin, le programme SELECT offre la possibilité de travailler sur un sous-ensemble d'individus, sélectionnés selon certains critères.

Ces différentes opérations sont indispensables dans le choix des variables prises en compte et donc dans l'interprétation finale des résultats.

## C) TRAITEMENTS STATISTIQUES DES DONNÉES

### 1) Programmes généraux d'analyse de données

Les premiers traitements peuvent être très simples : établir l'histogramme d'une variable. Si l'on désire comparer deux variables, le travail se fait à deux dimensions, soit par le calcul d'un coefficient de corrélation, soit par l'analyse d'un tableau à double entrée (programme CROISA). Pour trois variables et plus, on se situe dans un espace mathématique à  $n$  dimensions ; l'on peut faire appel à des techniques comme l'analyse en composantes principales, l'analyse factorielle. Il s'agit de programmes très répandus (par ex. GENSTAT, SPAD, STATGRAPHICS...).

### 2) Classement et classification des données

Un classement de données consiste à ranger des individus dans des classes préalablement établies (par exemple un référentiel national). Une classification de données réalise une partition au sein d'un ensemble étudié. Les programmes proposés s'inspirent de la méthode utilisée implicitement par le pédologue cartographe (RAYNER, 1966), proche de la méthode dite des nuées dynamiques

(DIDAY, 1971). Le pédologue dispose au départ d'un référentiel pré-établi très général. Il peut alors procéder à un classement de ses données. Le plus souvent, il lui faut construire ses propres références régionales en établissant une classification. Toute nouvelle observation est classée au sein de ce référentiel élargi, mais elle peut induire des modifications. La démarche proposée s'inspire de cet aller-retour entre les données recueillies et les références mises en place ; on procède ainsi par approximations successives.

La méthode est divisée en deux niveaux d'approche :

— d'une part, au niveau de l'horizon qui est l'entité structurale minimale couramment utilisée pour les descriptions pédologiques (programme DIMITRI, GIRARD, 1983) ;

— d'autre part, au niveau du profil qui correspond à une approche plus globale (programme VLADIMIR, KING et GIRARD, 1988).

Dans le programme DIMITRI, le rattachement de chaque horizon étudié à un horizon de référence se fait à partir du calcul d'une distance mathématique. Le programme VLADIMIR utilise ce même principe mais il découpe chaque profil en un nombre de niveaux de comparaison fixé par le pédologue. La distance entre un profil « réel » et un profil de référence est égale à la somme des distances calculées à chaque niveau. La valeur de cette distance est un moyen d'évaluer la validité du classement.

Tant pour DIMITRI que pour VLADIMIR, un retour aux données de base après classement des horizons ou des profils est possible grâce à la production d'histogrammes. Simples à interpréter, ils permettent le contrôle de la classification et d'éventuels réajustements des références.

Des programmes annexes permettent d'appréhender les relations entre horizons appartenant à un même profil. Ils facilitent en particulier la construction des profils de référence à partir des horizons de référence établis par le programme DIMITRI. Le programme CONTRA calcule les distances entre les horizons adjacents au sein d'un même profil ; la visualisation graphique de cette distance fournit une évaluation des contrastes intra-profil. Le programme SUSOU recherche le nombre de fois où deux horizons de référence sont adjacents ou présents au sein d'un même profil.

## D) TRAITEMENTS GRAPHIQUES DES DONNÉES

Les techniques graphiques sont couramment utilisées en Pédologie pour mieux visualiser les informations.

### 1) Projection sur un plan mathématique

La projection des données sur un tel plan est l'une des opérations les plus couramment utilisées. Le triangle de texture est un exemple parmi de nombreux cas. Le programme PROJET effectue automatiquement le dessin d'une projection sur deux axes dont les variables  $x$  et  $y$  sont choisies par l'utilisateur. Ces axes peuvent être bornés et inscrits négativement.

### 2) Projection sur un plan géographique

Le plan géographique est un sous-ensemble du plan mathématique, où l'axe des  $x$  est la longitude et l'axe des  $y$  la latitude. Dans un plan géographique, l'axe  $z$  possède un sens précis, soit l'altitude du point, soit la profondeur du profil observé. Sur cet axe  $z$ , on peut représenter une variable pédologique de son choix. Une telle projection forme un document appelé cartogramme.

Un premier programme (STILO) réalise une représentation symbolique simple pour un nombre infini de points. Il est particulièrement adapté aux sondages à la tarière. Un second programme nommé DOP (COHEN-BRISSON, 1984) dessine des schémas conventionnels de profils pédologiques à l'aide de nombreuses options. L'altitude est prise en compte ; elle permet de dessiner des transects topographiques ou des blocs diagrammes.

### 3) Généralisation spatiale

Un programme d'interface (SERPE) permet de présenter les données sous une forme compatible avec les méthodes de la géostatistique (GASCUEL-ODOUX, 1984). Il est possible par exemple de sélectionner en chaque point d'observation :

- l'horizon rencontré à une profondeur choisie (tomographie),
- l'horizon le plus semblable à une référence,
- la profondeur d'apparition d'un caractère, etc.

## CONCLUSION

Plusieurs travaux scientifiques ont été réalisés à l'aide de la bibliothèque LOGOS (SALIN, 1985 ; KING, 1986 ; MICHELIN, 1986). Ils ont montré l'intérêt du développement de logiciels statistiques et graphiques adaptés aux besoins spécifiques de la Pédologie. Les programmes sont interactifs, laissant l'utilisateur maître du choix des nombreuses options proposées et de l'orientation de certains logiciels statistiques. Nous espérons ainsi satisfaire au mieux les besoins des cartographes qui doivent assurer le traitement de leurs données au fur et à mesure de la prospection. Un tel ensemble permet enfin de comparer, avec plus de rigueur, diverses approches cartographiques selon des choix liés à des objectifs thématiques différents.

## BIBLIOGRAPHIE

- BERTRAND R., FALIPOU P., LEGROS J.-P.**, 1984. — STIPA, notice pour l'entrée des descriptions et analyses de sols en banque de données. ACCT, Paris. 136 p.
- BONNERIC P., NAVARRO R.**, 1988. — STIPA micro, version 2. Procédures d'installation et d'utilisation. CUTI Montpellier - INRA - ACCT. Doc. ronéo. 20 p.
- COHEN-BRISSON N.**, 1984. — Elaboration d'un programme de dessin automatique de profils de sol : le programme DOP. Rapport de stage. INRA-SESCPF. Doc. ronéo. 18 p.
- DIDAY E.**, 1971. — La méthode des nuées dynamiques. Rev. Stat. Appl., 19 (2) 19-34.
- GASCUEL-ODOUX C.**, 1984. — Application de la géostatistique à l'étude de la variabilité spatiale des propriétés hydriques du sol. Thèse Doct. Ing. ENSMP Fontainebleau, 235 p.
- GIRARD M.-C.**, 1983. — Recherche d'une modélisation en vue d'une représentation spatiale de la couverture pédologique. Application à une région des plateaux jurassiques de Bourgogne. Thèse Doct. Etat, Paris 7, 430 p.
- GIRARD M.-C., KING D.**, 1988. — Un algorithme interactif pour la classification des horizons de la couverture pédologique : DIMITRI. Science du Sol, 26(2) 81-100.
- KING D.**, 1986. — Modélisation cartographique du comportement des sols basée sur l'étude de la mise en valeur du Marais de Rochefort. Thèse (Docteur-Ingénieur) INA Paris-Grignon, 243 p.
- KING D., DUVAL O.**, 1988. — LOGOS. Logiciels pour l'étude de la géographie des sols. Version 3.1. INRA-SESCPF, 112 p.
- KING D., GIRARD M.-C.**, 1988. — Réflexions sur la classification des profils de la couverture pédologique. Proposition d'un algorithme : VLADIMIR. Science du Sol, 26(4) 239-254.
- MICHELIN J.**, 1986. — Recherche informatisée d'unités de sols sur schistes briovériens en Basse-Normandie. Programme d'aide à l'aménagement. DEA, INA Paris-Grignon, 75 p.
- RAYNER J.H.**, 1966. — Classification of soils by numerical methods. Journal of Soil Science. Vol. 17, 79-82.
- SALIN R.**, 1985. — Etude pédologique de marais en Charente Maritime. Carte des sols et notice explicative. INRA-SESCPF, EPR Poitou-Charentes. 168 p.