

La banque de données pédologiques de l'ORSTOM

R. VAN DEN DRIESSCHE

S.S.C. de l'ORSTOM, 93140 Bondy, France

RÉSUMÉ

Après le traitement statistique sur ordinateur de données de laboratoire, la Banque de Données Pédologiques de l'ORSTOM a pu maîtriser, dès 1968, le traitement sans codage manuel des descriptions de profils.

Un système informatique en langage naturel, puissant et modulaire, a été mis en place autour d'un terminal TMF 300 (d'ordinateur UNIVAC 1108) installé dans les Services Scientifiques Centraux de l'ORSTOM.

Toutes les fonctions sont remplies par ce système et c'est après la sélection booléenne, commandée en français ou dans une langue étrangère, que sont appliquées — aux tableaux sélectionnés de données codées et de manquants — les méthodes statistiques suivantes : Mesures multivariées non-paramétriques de distances entre horizons, profils, groupes de sols, environnements, etc ; Tests non-paramétriques de corrélations entre variables qualitatives et/ou quantitatives ; Regroupements des horizons, profils, etc, en constellations répondant aux critères de l'infériorité des distances moyennes intra-constellation ; Estimation des limites de confiance des médianes de chaque variable.

Ce système convient à tous les vocabulaires normalisés.

ABSTRACT

A soil information system in plain language has been set up for the UNIVAC 1108 computer. The ORSTOM system is basically a data base of on-going research in soil science. The data base accepts profile descriptions in French and in foreign languages and incorporates the following features : (1) management of the data

base ; (2) updating ; (3) Boolean selection of horizons ; (4) nonparametric measurements of multivariate distances between horizons, profiles, or environments, followed by clustering ; (5) Spearman tests of correlations ; (6) Confidence intervals of the median.

The paper shows the modularity of the DP system but also the use and power of statistical analysis.

РЕЗЮМЕ

После статистической обработки на ЭВМ лабораторных данных, Банк почвенных данных ORSTOM смог освоить, уже с 1968-го года, обработку характеристик профилей без ручного кодирования.

Мощная, по вместительности и скорости, и модульная информационная система на естественном языке была размещена вокруг входной и выходной аппаратуры TMF 300 (ЭВМ UNIVAC 1108), установленной в Центральные Научные Службах ORSTOM.

Система выполняет все функции, и после булевого отбора, проведенного на французском или иных языках, применяются, к отборочным таблицам кодированных и отсутствующих данных, следующие статистические методы : многопеременные и непараметрические измерения расстояний между горизонтами, профилями, группами почв, окружающими средами и т. д. ; непараметрические тесты корреляций между качественными и/или количественными переменными ; группировки горизонтов, профилей и т. д. в констелляции отвечающие критериям, согласно которым средние расстояния между констелляциями превышают таковые внутри констелляции ; оценка доверительных границ медиан каждой переменной.

Система эта годится для всяких нормализованных словарей.

OBJECTIFS

La Banque de Données Pédologiques de l'ORSTOM est une cellule de recherche et de service, qui a pour objectifs de rassembler et de traiter sur ordinateur, avec des méthodes qui lui sont propres, des descriptions de sols. L'exploitation des résultats du traitement revient au pédologue.

Le besoin d'harmoniser le vocabulaire était ressenti par tous les pédologues. Sur l'initiative de la Délégation Générale à la Recherche Scientifique et Technique, un groupe de travail a examiné, en 1969, plusieurs projets de glossaire de pédologie et a suggéré un certain nombre d'améliorations. Le texte obtenu a été approuvé à l'unanimité par les experts qui ont décidé de recommander son adoption officielle par les organismes qu'ils représentaient. Un second glossaire de pédologie, pour la description de l'environnement du profil en vue du traitement informatique, a été préparé en 1970 et 1971 par une soixantaine d'experts sous l'égide de l'association Informatique et Biosphère et publié avec la participation du Conseil International de la Langue Française.

Ainsi, trois années de travail ont été nécessaires pour harmoniser au plan national le vocabulaire de terrain. Plusieurs pédologues de l'ORSTOM et le personnel de la Banque ont d'ailleurs pris part à la rédaction des glossaires.

La volonté de traiter des descriptions conformes aux glossaires de pédologie équivalait à mettre en place un système informatique puissant et modulaire incorporant des méthodes statistiques multivariées non-paramétriques.

DÉONTOLOGIE

Les correspondants de la Banque de Données Pédologiques de l'ORSTOM qui lui confient leurs données normalisées reçoivent personnellement, sous pli avion ouvert, un exemplaire du résultat de tous les traitements effectués et un tiré à part des méthodes utilisées. Les correspondants sont alors invités à interpréter ces résultats et à les publier, sous leur nom, dans la revue de leur choix, s'ils jugent la diffusion intéressante.

Ce traitement individuel ne saurait toutefois suffire, l'objectif de la Banque étant de mettre en commun les données harmonisées et de leur appliquer les mêmes méthodes de traitement — statistique ou graphique — de façon à *corrélér* (au sens pédologique) les sols et à estimer leurs paramètres sur des effectifs suffisants.

Le traitement collectif est toujours fait sur l'initiative de la Banque et avec l'accord de tous les intervenants. Il débouche sur une publication collective si les résultats la justifient.

Ces quelques règles simples doivent faciliter l'alimentation et le fonctionnement de la Banque.

Pour la mise au point de ses nouvelles méthodes, la Banque utilise exclusivement des données brutes publiées sans traitement statistique.

LE MATÉRIEL

Le matériel utilisé actuellement se compose d'un ordinateur, d'un terminal et d'une perforatrice.

Doté d'une unité centrale de 192 000 mots de 6 caractères, l'UNIVAC 1108 sous Exec 8 (cycle 750 ns) de la Société de Traitement Automatique des Données (STAD, 11 rue de la Vistule, 75013 Paris) est relié, depuis novembre 1973, par une ligne téléphonique spécialisée (4 800 bauds) à un terminal lourd TMF 300 de la Société des Ordoprocresseurs (113 rue Aristide Briand, 91400 Orsay) installé à 20 m de la Banque, dans les locaux des Services Scientifiques Centraux de l'ORSTOM. Ce terminal est modulaire. La configuration choisie comporte un lecteur de cartes (400 cartes par minute) et une imprimante à 136 caractères par ligne (400 lignes par minute).

Une perforatrice IBM 029 à bloc EBCDIC est accessible à mi-temps, en attendant l'arrivée d'une seconde perforatrice 029. Les travaux de perforation et de vérification sont, en partie, confiés à la Société d'Information de Gestion d'Analyse et de Programmation (SIGAP, 1 et 3 rue Lulli, 75002 Paris).

D'autres ordinateurs ont été utilisés par le passé. Dès 1962, l'ordinateur IBM 704 de l'Institut Blaise-Pascal du CNRS (23 rue du Maroc, 75019 Paris) pour les expériences sols-plantes du Bureau des Sols des Antilles, ensuite l'ordinateur CDC 3600 du même Institut pour les fichiers de données de laboratoire, et, à partir de 1968, pour les descriptions de profils. Pendant quelques années, on a aussi travaillé sur du matériel IBM : le 360/50 de l'Institut Blaise Pascal, le 360/75 et le 370/165 du CIRCÉ (CNRS) relié au terminal 2780 installé rue du Maroc.

LE LOGICIEL

En proposant, en 1968, de développer un système en langage naturel qui a la propriété de s'appliquer facilement à toutes les catégories de données, le res-

ponsable voulait répandre l'usage de l'informatique parmi les pédologues. De toute évidence, une fiche écrite et perforée comme ceci :

Forme·Versant·Rectiligne·A saillants anguleux·
·Recoupant les couches géologiques·Ravin·
·Dénivellation 50m·Exposition NW·Pente 30pc·
·Forme héritée se survivant·Non datée·etc·

est plus aisée à obtenir que la suite codée :

.AC1.AD1.BD8.BE3.AE2.BF16.FB50.CE8.FC30.
.CF2.CG0.etc.

En outre, les données sont nombreuses par variable, et les conditions de la collecte souvent mauvaises. C'est pourquoi, on ne pouvait mettre en circulation des fiches à graphiter convenant à un lecteur optique.

L'avantage du langage naturel apparaît encore mieux lors de l'écriture des expressions de sélection booléenne.

Pour les méthodes statistiques, il était fait appel au concept du tableau lignes x colonnes occupé par des données qualitatives et/ou quantitatives et par des manquants.

Les principales fonctions du système sont : Collecter, Répertoire, Perforer, Collationner, Corriger, Editer, Traduire, Actualiser, Sélectionner, Annuler, Caractériser, Distancer, Regrouper, Corréler.

COLLECTER

La collecte des données scientifiques, qualitatives ou quantitatives, doit utiliser un vocabulaire normalisé, au moins au plan national.

Par convention, on appelle FICHE l'ensemble des données d'une même discipline, collectées par un même observateur, dans un même lieu et à une date déterminée.

Cette fiche peut être présentée de trois façons :

1) Sous la forme d'une énumération ordonnée ou non, manuscrite ou dactylographiée, des seules données observées, séparées par des points.

2) Sous la forme d'une liste de référence type, qui est exhaustive, et sur laquelle les données observées sont pointées ou, le cas échéant, complétées.

3) Sous la forme d'un enregistrement phonique des données (sur minicassette par exemple) effectué lors de l'observation des faits. Les données doivent être espacées. Leur ordre est indifférent.

Telles que présentées dans les vocabulaires normalisés, et par voie de conséquence dans les répertoires, les données se suffisent à elles-mêmes et s'enchaînent, même en l'absence de certaines variables. Le nom de la variable n'apparaît pas dans la fiche. De même, les différentes données d'une variable doivent s'exclure mutuellement et il est erroné d'en noter deux. Et d'autres mots, une seule donnée représente la variable. Selon l'objectif de l'étude, il arrive que l'observateur ne fasse appel qu'à un sous-ensemble de variables. Dans ce cas, les variables non utilisées ne font l'objet d'aucune inscription. En effet, toutes les variables ne doivent pas nécessairement être représentées pour conserver aux méthodes utilisées leur validité.

RÉPERTORIER

Un répertoire exhaustif d'équivalences *données en langage naturel = codes numériques* et *variables en langage naturel = codes alphabétiques*, à l'usage exclusif de l'ordinateur, doit être établi pour chaque domaine.

Des données synonymes peuvent figurer sous le même code. Toutes sont acceptées à l'entrée des données observées, mais seulement la première apparaît en sortie.

Les versions en langue étrangère doivent contenir les mêmes codes que la version française. Toutes les versions se succèdent sur la même bande magnétique.

Les répertoires établis pour des domaines différents sont chargés sur des bandes différentes.

PERFORER

Les données de la fiche sont perforées en langage naturel et en continu, avec un point comme séparateur. L'ordre de perforation est sans importance.

Les caractères spéciaux ayant été volontairement limités aux six qui ont les mêmes codes sur tous les matériels (point, virgule, astérisque, trait d'union, barre oblique, franc), la perforation peut être décentralisée.

A la réception des cartes perforées, la séparation des fiches est marquée définitivement par une carte de numérotage séquentiel continu.

COLLATIONNER

Lors du chargement des fiches sur bande magnétique, toutes les discordances avec le répertoire consulté

sont annotées. Le bilan des annotations est indiqué au bas de la fiche sur la liste de l'imprimante.

Etablie au format 157 × 195 mm, en double exemplaire sur la même page, la fiche respecte l'ordre initial des données tel qu'adopté lors de la collecte.

Le collationnement est obtenu par une carte de référence au répertoire et par une carte d'insertion, suivies des cartes perforées avec les données. Celles-ci doivent se suivre dans l'ordre croissant des numéros de fiches.

CORRIGER

Pour la correction sur bande d'une fiche inexacte il suffit : (1) de remplacer dans le jeu de cartes perforées correspondant à la fiche inexacte — extrait manuellement du fichier — chaque carte inexacte par une ou deux cartes exactes (selon le rapport de longueur des données) mettant à profit l'équivalence entre la suite *point espacements point* et le *point* séparateur de données ; (2) d'introduire dans le lecteur le nouveau jeu de cartes précédé de la carte de contrôle modification et de la carte de référence au répertoire.

La correction simultanée de plusieurs fiches ne réclame qu'une seule carte de contrôle.

La sortie imprimée en double est annotée si des discordances subsistent.

La destruction des cartes perforées est donc retardée jusqu'au moment où les fichiers bandes sont reconnus exacts et sont trois à être utilisés en bascule.

EDITER

L'édition se distingue de l'impression par un ordre logique des données, par la séparation des sujets traités, par l'exclusion des données non normalisées.

L'édition d'une fiche s'obtient par une carte perforée avec la référence du répertoire approprié, suivie d'une carte portant le numéro de la fiche. Pour éditer plusieurs fiches, il faut perforer une carte par fiche et les présenter dans l'ordre croissant.

La sortie de l'imprimante est paginée et dédoublée sur la moitié gauche et droite, avec une justification de 62 caractères (157 mm) et une hauteur maximale de 195 mm.

TRADUIRE

La traduction d'une ou plusieurs fiches dans une langue étrangère est commandée par une carte per-

forée avec la référence du répertoire établi dans cette langue et par une ou plusieurs cartes portant le numéro des fiches.

La sortie est éditée.

ACTUALISER

On peut à tout instant actualiser la langue de l'édition. L'actualisation est considérée comme la traduction d'un répertoire ancien dans un répertoire nouveau, quel qu'en soit la langue. Elle est en effet commandée de la même façon.

Les sorties sont éditées.

SÉLECTER

La fonction de sélection peut être commandée sur n'importe quel fichier bande par une carte de référence au répertoire utilisé, par une carte numéro d'ordre du travail de sélection et par des cartes perforées avec les conditions de sélection.

Variables et données, exprimées en langage naturel, le cas échéant dans l'une des langues étrangères ou dans l'une des versions actualisées, sont unies par les relations d'ordre :

GE plus grand ou égal
PE plus petit ou égal
PG plus grand
PP plus petit
EG égal
IG inégal

et les opérateurs logiques :

OU
ET (qui est prioritaire)

en une série maximale de 100 conditions, pour extraire sélectivement d'une bande magnétique déterminée les fiches qui satisfont les conditions imposées. Plusieurs expressions entières de sélection sont acceptées dans le même passage en intercalant des cartes de rebobinage.

Un type particulier de sélection consiste à extraire les fiches ayant en commun une liste déterminée de variables, dont les données elles-mêmes ne sont pas prises en considération. Une telle sélection est commandée par autant de conditions *INEGAL.MAN-QUANT* qu'il y a de variables dans la liste ; l'opérateur ET reliant les conditions.

Les fiches sélectionnées sont éditées et la langue est celle de l'expression de sélection. Divers diagnostics permettent de déceler les fautes éventuelles.

Une sortie perforée en codes numériques dans des zones assignées par le répertoire et identifiée par les numéros de travail et de fiche est, en outre, considérée comme standard. Un code de perforation (-1) remplace les manquants.

ANNULER

L'annulation d'une fiche enregistrée sur bande est commandée par une carte de référence au répertoire et une carte d'annulation perforée avec le numéro de la fiche. Pour annuler plusieurs fiches il faut la carte du répertoire et, dans l'ordre croissant, une carte d'annulation par fiche.

CARACTÉRISER

La médiane des données d'une variable, qualitative ou quantitative, est la donnée observée qui a dans l'échantillon sélectionné autant de données qui lui sont inférieures que de données qui lui sont supérieures.

Les rangs sont des numéros d'ordre 1 à m affectés aux m données observées de l'échantillon. Les données identiques sont toutes remplacées par la moyenne des rangs correspondants. Le rang $\frac{1}{2}(m+1)$ est le rang de la donnée observée qui est considérée comme la médiane de l'échantillon.

La médiane vraie est la médiane de la population théorique échantillonnée. Elle est incalculable. En revanche, les limites de confiance de la médiane vraie sont calculées à partir de l'effectif avec un risque inférieur à 5 % :

$$\frac{1}{2}(m+1) - m^{1/2} \quad \frac{1}{2}(m+1) + m^{1/2}$$

On obtient ainsi deux rangs de données observées : le premier est le rang de la donnée considérée comme limite inférieure ; le second est le rang de la donnée considérée comme limite supérieure de la médiane vraie.

DISTANCER

La distance est une mesure multivariable de dissemblance entre fiches. Elle peut être non-paramétrique

(indépendante de la fonction de répartition des variables) en la calculant à partir des rangs r . Elle varie de zéro à deux.

Si v est le nombre de variables qualitatives et/ou quantitatives

m_i l'effectif de la variable i (effectif total m moins manquants)

e_i le nombre de lots de données identiques de la variable i

t_{qi} le nombre de données identiques dans le lot q de la variable i

la distance entre les données de la fiche h et celles de la fiche k est :

$$D_{hk} = \frac{1}{v} \left\{ \sum_{i=1}^v (r_{hi} - r_{ki})^2 / \frac{1}{12} \left[m_i^3 - m_i - \sum_{q=1}^{e_i} (t_{qi}^3 - t_{qi}) \right] \right\}$$

Entre m fiches on calcule $\binom{m}{2}$ distances.

REGROUPER

Après l'établissement d'une matrice de distances entre fiches, il y a lieu de se demander à quel point ces distances différencient les fiches et si certaines fiches très proches, du fait des distances faibles qui les relient, ne forment pas des ensembles indissociables, ou des ensembles que les variables utilisées ne parviennent pas à dissocier, et qui portent le nom de constellations.

L'examen attentif d'un tableau de distances généralisées obtenues sur une dizaine de fiches permet, à lui seul, de déceler certaines constellations.

Pour quelques dizaines de fiches cela devient laborieux et il est de toute façon plus rigoureux et plus rapide de procéder par calcul.

Une approche numérique du problème de la recherche des constellations a été tentée. Des conditions précises ont été choisies ; elles permettent de décider si des fiches entrent ou non dans des constellations.

La première constellation

Les distances sont présentées, par fiche, dans l'ordre croissant. Il y a, au total, $(m-1)m/2$ distances entre les m fiches. La différence entre la distance maximale et la distance minimale est appelée étendue. La distance la plus faible du tableau relie deux fiches dont l'appartenance à une première constellation n'est

mise en doute que lorsque cette distance dépasse l'étendue.

L'adjonction d'une troisième fiche se fait sur la base d'un accroissement minimal de la première distance. Cette première distance ne peut être accrue, pour une troisième fiche, que de la somme de deux distances, mais il existe $m-2$ façons de l'accroître. Ces $m-2$ accroissements possibles sont à calculer comme suit :

$$\text{Un accroissement} = \left\{ \begin{array}{l} 1^{\text{re}} \text{ distance, distance d'une} \\ \text{troisième fiche à la pre-} \\ \text{mière} \\ + 2^{\text{e}} \text{ distance, distance de cette} \\ \text{troisième fiche à la deu-} \\ \text{xième.} \end{array} \right.$$

C'est la fiche qui provoque l'accroissement minimal qui est retenue, comme troisième fiche de la constellation, si toute distance mesurée entre une de ces trois fiches et une des $m-3$ autres fiches dépasse la distance moyenne intra-constellation :

$$\boxed{D \text{ intra-extra-constellation} > \bar{D} \text{ intra-constellation}}$$

Quand, au contraire, une des distances intra-extra-constellation d'avère inférieure à la distance moyenne intra-constellation $\Sigma D/3$, il faut en conclure que la troisième fiche est à l'extérieur de la constellation et qu'il n'y a pas d'autre fiche dans la première constellation : les $m-2$ fiches entrent dans d'autres constellations ou restent dissociées.

Dans l'éventualité de l'adjonction d'une troisième fiche à la première constellation, il y a lieu de calculer, dans une troisième étape, les accroissements de distance provoqués par chacune des $m-3$ fiches restantes.

$$\text{Un accroissement} = \left\{ \begin{array}{l} 1^{\text{re}} \text{ distance, distance d'une} \\ \text{quatrième fiche à la troi-} \\ \text{sième} \\ + 2^{\text{e}} \text{ distance, distance de cette} \\ \text{quatrième fiche à la pre-} \\ \text{mière} \\ + 3^{\text{e}} \text{ distance, distance de cette} \\ \text{quatrième fiche à la deu-} \\ \text{xième.} \end{array} \right.$$

La somme de la deuxième et troisième distances est déjà calculée à l'étape précédente. La quatrième fiche envisagée pour la constellation est celle entraînant l'accroissement le plus faible. Elle n'est retenue que

si toutes les distances intra-extra-constellation sont supérieures à la nouvelle distance moyenne intra-constellation $\Sigma D/6$.

En résumé, le processus d'adjonction de fiches à l'ébauche de constellation imposée par la distance minimale est poursuivi jusqu'à l'obtention d'une distance moyenne intra-constellation supérieure à une quelconque distance intra-extra-constellation. La dernière fiche envisagée est alors rejetée et le processus recommence à partir d'une ébauche de deuxième constellation.

La deuxième constellation

La distance la plus petite entre les fiches qui n'entrent pas dans la première constellation définit l'ébauche d'une deuxième constellation si la distance moyenne interconstellations est supérieure à chacune des deux distances moyennes intra-constellation.

$$\boxed{\bar{D} \text{ interconstellations} > \bar{D} \text{ intra-constellation}}$$

Une distance moyenne entre deux constellations s'obtient en divisant la somme des distances, mesurées entre chaque fiche de la première constellation et chaque fiche de la seconde, par le nombre de distances.

L'adjonction éventuelle d'autres fiches aux deux premières de cette deuxième constellation doit remplir les deux conditions :

$$\boxed{\begin{array}{l} \bar{D} \text{ interconstellations} > \bar{D} \text{ intra-constellation} \\ D \text{ intra-extra-constellation} > \bar{D} \text{ intra-constellation} \end{array}}$$

L'adjonction de fiches est interrompue dès que l'une ou l'autre de ces conditions n'est plus remplie.

Les constellations suivantes

La distance la plus faible entre les fiches se trouvant à l'extérieur des deux premières constellations sert, à nouveau, de point de départ et les deux conditions précitées restent d'application jusqu'à épuisement de toutes ces fiches. Les distances moyennes interconstellations sont calculées, par conséquent, lors de chaque tentative d'adjonction de nouvelles fiches.

Le critère de l'étendue, auquel il n'est fait appel qu'une fois, lors de l'acceptation des deux premières fiches de la première constellation, est arbitraire. Son abandon reviendrait, toutefois, à déceler toujours au moins une constellation, ce qui irait à l'encontre de l'objectif fixé à certaines études, à savoir la recherche du pouvoir maximal de dissociation.

On peut présenter sous forme matricielle les distances moyennes intra- et interconstellations. Les distances moyennes interconstellations sont également calculées à partir des accroissements de distances.

L'interprétation des résultats doit porter, à la fois, sur le pouvoir de dissociation des variables, mesuré par les distances respectives entre constellations, entre constellations et fiches individualisées, et aussi, entre fiches individualisées.

CORRÉLER

La corrélation est un test d'interdépendance entre deux variables qualitatives, quantitatives, ou qualitative et quantitative, dans un échantillon sélectionné d'effectif m . Seule la méthode de corrélation de rangs est applicable. L'indice calculé est testé aux risques 5 % et 1 %. Le signe de la corrélation significative est positif ou négatif.

Si m_{ij} est l'effectif commun des variables i et j

e_i le nombre de lots de données identiques de la variable i

e_j le nombre de lots de données identiques de la variable j

t_{qi} le nombre de données identiques dans le lot q de la variable i

t_{qj} le nombre de données identiques dans le lot q de la variable j

la corrélation entre les données de la variable i et celles de la variable j est :

$$R_{ij} = \frac{1}{2} (I_i I_j)^{-1/2} \left[I_i + I_j - \sum_{h=1}^m (r_{hi} - r_{hj})^2 \right]$$

où

$$I_i = \frac{1}{12} (m_{ij}^3 - m_{ij}) - \sum_{q=1}^{e_i} \frac{1}{12} (t_{qi}^3 - t_{qi})$$

$$I_j = \frac{1}{12} (m_{ij}^3 - m_{ij}) - \sum_{q=1}^{e_j} \frac{1}{12} (t_{qj}^3 - t_{qj})$$

Le test de signifiante de R_{ij} se fait par consultation de tables aux risques choisis.

Entre v variables on calcule $\binom{v}{2}$ coefficients de corrélation.

PUBLICATIONS DE LA BANQUE EN STATISTIQUE ET INFORMATIQUE APPLIQUÉES

CHARBONNEL (M.), VAN DEN DRIESSCHE (R.), 1964. — Analyse de la covariance d'expériences 3³ en 27 unités sur ordinateur IBM 704. *Biom.-Prax.*, vol. V, n° 1 : 20-25.

VAN DEN DRIESSCHE (R.), 1965. — La recherche des constellations de groupes à partir des distances généralisées D^2 de Mahalanobis. *Biom.-Prax.*, vol. VI, n° 1 : 36-47.

VAN DEN DRIESSCHE (R.), MAIGNIEN (R.), 1965. — Application d'une méthode de la statistique approfondie à la pédologie. *Cah. ORSTOM, sér. Pédol.*, vol. III, n° 1 : 79-88.

VAN DEN DRIESSCHE (R.), 1966. — Un problème de classification numérique. *Cah. ORSTOM, sér. Pédol.*, vol. IV, n° 4 : 91-96.

DEFOSSEZ (M.), MANGIN (J.-Ph.), PINTA (M.), VAN DEN DRIESSCHE (R.), 1967. — Répartition de quelques éléments traces dans les eaux de surface en zone intertropicale (Côte d'Ivoire). *Bull. Ser. Carte géol. Als.-Lorr.*, vol. XX, n° 4 : 257-276.

VAN DEN DRIESSCHE (R.), MASBOU (M.C.), 1968. — La recherche des variables les plus utiles à la mesure des distances D^2 . *Cah. ORSTOM, sér. Pédol.*, vol. VI, n° 1 : 21-34.

BAUZON (D.), VAN DEN DRIESSCHE (R.), DOMMERGUES (Y.), 1968. — Caractérisation respirométrique et enzymatique des horizons de surface des sols forestiers. *Sc. Sol.*, vol. VIII, n° 2 : 55-78.

VAN DEN DRIESSCHE (R.), MASBOU (M.C.), 1969. — Determination of the best subset of variables with respect to D^2 distance. *Statistical Theory and Method Abstracts*, vol. X, n° 2 : 510.

BAUZON (D.), VAN DEN DRIESSCHE (R.), DOMMERGUES (Y.), 1969. — L'effet litière. I. - Influence in situ des litières forestières sur quelques caractéristiques biologiques des sols. *Oecol. Plant.*, vol. IV : 99-122.

BECK (G.), DOMMERGUES (Y.), VAN DEN DRIESSCHE (R.), 1969. — L'effet litière. II. - Etude expérimentale du pouvoir inhibiteur des composés hydrosolubles des feuilles et des litières forestières vis-à-vis de la microflore tellurique. *Oecol. Plant.*, vol. IV : 237-266.

VAN DEN DRIESSCHE (R.), 1971. — A soil information system. First Seminar in Geographic Data Processing. Vancouver, Simon Fraser University, 11-12 décembre 1971, multigr. 4 p.

AUBRY (A.M.), 1971. — Test programmé de la corrélation des rangs. ORSTOM, Bondy, multigr., 50 p.

VAN DEN DRIESSCHE (R.), GARCIA GOMEZ (A.), 1972. — Distances non-paramétriques entre profils. *Rev. Ecol. Biol. Sol.*, vol. IX : 617-628.

VAN DEN DRIESSCHE (R.), 1972. — Un système en langage naturel. ORSTOM, Bondy, multigr., 4 p.

AUBRY (A.M.), VAN DEN DRIESSCHE (R.), BAUZON (D.), PERAUD (A.), DOMMERGUES (Y.), 1973. — Measures of rank distances followed by repeated clustering and tests of rank correlations in the study of biological and chemical data from tropical forest soils (Ivory Coast). *Bull. Ecol. Res. Comm.*, vol. XVII : 433-442.

VAN DEN DRIESSCHE (R.), GARCIA GOMEZ (A.), 1973. — Comparaison multivariable non-paramétrique des profils sans nomenclature des horizons. *Cah. ORSTOM, sér. Pédol.*, vol. XI, n° 3/4 : 257-264.

VAN DEN DRIESSCHE (R.), GARCIA GOMEZ (A.), 1974. — Un système informatique en langage naturel : application didactique aux coloris ISCC-NBS. *Init. Doc. tech. ORSTOM*, sous presse.

RENCONTRES

Une participation à plusieurs rencontres internationales a permis de présenter le système à des collègues informaticiens, statisticiens, géographes et pédologues.

- septembre 1970 : UNESCO/IGU First Symposium on Geographical Information Systems
Ottawa
- novembre 1971 : IGU Commission on Geographical Data Sensing and Processing Working Group 2. System Definition and Review
Saskatoon
- décembre 1971 : First Seminar in Geographic Data Processing
Vancouver
- février 1972 : IGU Commission on Geographical Data Sensing and Processing Working Group 2. System Definition and Review
Athènes
- avril 1972 : FAO/UNESCO ad hoc Consultation on Computerized Soil Data Interpretation for Development Purposes
Rome
- juin 1972 : IBP Symposium on Modern Methods in the Study of Microbial Ecology
Uppsala
- août 1972 : UNESCO/IGU Second Symposium on Geographical Information Systems
Ottawa
- septembre 1973 : FSAE Semaine d'Etude Sol et Fertilisation
Gembloux

C'est aux contacts qui ont été pris à l'occasion de ces rencontres que l'on doit certains aspects de la Banque. Que les animateurs de ces réunions et en particulier MM. les Professeurs Boyle, Dudal, Peucker, Swindale, Tomlinson, Yoeli veuillent bien trouver ici l'expression de notre gratitude.

LES CORRESPONDANTS

Une quinzaine de correspondants appartenant aux cadres de l'ORSTOM apportent leur soutien à la

Banque en lui envoyant des profils décrits en termes normalisés (plus de 500 profils en 1973).

En outre, des fichiers de données enzymatiques sont traités en collaboration avec le Centre de Pédologie Biologique du CNRS (Nancy).

Enfin, quelques correspondants de France, d'Italie et d'Espagne confient des dossiers à la Banque dans le but de leur appliquer les méthodes disponibles et de s'associer dans l'interprétation des résultats.

STAGIAIRES

Le meilleur accueil est réservé aux stagiaires. Un souci d'efficacité qui se répand dans le monde scientifique donne la priorité aux stages de longue durée. Le stagiaire participe aux travaux et aux recherches en cours. La langue de travail peut être indifféremment le français, l'anglais, l'espagnol, le néerlandais, le portugais.

Manuscrit reçu au SCD le 20 février 1974.

BIBLIOGRAPHIE

- Glossaire de pédologie description des horizons en vue du traitement informatique*, 1969. — Rédigé sur l'initiative de la DGRST en collab. avec CNRS, CACG, CNARBRL, etc. Init. Doc. Tech. ORSTOM h.s., Paris, 82 p.
- Glossaire de pédologie description de l'environnement en vue du traitement informatique*, 1971. — Rédigé en collab. avec CNRS, CACG, CNARBRL, etc., Informatique et Biosphère, Paris, 172 p.
- DOMMERS (Y.), 1968. — *La biologie des sols*. Coll. « Que sais-je ? », n° 399, PUF, Paris, 128 p.
- SIEGEL (S.), 1956. — *Nonparametric statistics : for the behavioral sciences*. Mc Graw-Hill, New York, 312 p.
- TOMLINSON (R.F.), ed., 1970. — *Environment information systems*. IGU Commission on Geographical Data Sensing and Processing, Ottawa, 161 p.
- TOMLINSON (R.F.), ed., 1972. — *Geographical data handling.*, 2 vol. IGU Commission on Geographical Data Sensing and Processing, Ottawa, 1389 p.