

Agence de Coopération Culturelle et Technique, Paris. 1978.
Normalisation et échange de données pédologiques. 173
par ordinateur.

(N)

ANNEXE IV

**LA BANQUE DE DONNEES PEDOLOGIQUES
DE L'O.R.S.T.O.M.
ET SES PROCEDURES OPERATIONNELLES
EN STATISTIQUE ET INFORMATIQUE
POUR DONNEES EN LANGAGE NATUREL
(POSEIDON)**

par Raymond VAN DEN DRIESSCHE, *ir. Eaux et forêts*,
Ana GARCIA GOMEZ, *Dr. ès Sc. Chimie des sols*,
Anne-Marie AUBRY, *Lic. Math., maitr. informatique*,
André GIEY, *ing.-conseil en Informatique*,
O.R.S.T.O.M., Bondy, France.

1. MATERIEL

Utilisation en libre-service à l'O.R.S.T.O.M., 70, route d'Aulnay, F 93140 Bondy, d'un terminal français modèle TMF 400 loué à la SFENA ORDO-PROCESSEURS (11 000 F/mois). Il s'agit d'un lecteur de cartes, console de visualisation pour le contrôle des passages en machine et d'une imprimante. Ce terminal est relié par une ligne à 4 800 bauds à un ordinateur UNIVAC 1108 à 192 K-mots de 36 bits, sous exec 8, compilateur FORTRAN 5 niveau 11, OS1100 version 32. Le même terminal est aussi relié par le réseau Caducée à des ordinateurs I.B.M. 370-168 sous MVS + JES3.

Perforatrices et perforieuses O.R.S.T.O.M. à temps partiel pour la banque de données pédologiques.

Facturation par GSI STAD du temps d'unité centrale 1108 (temps CPU) et des pistes Fastrand. 5 100 F (janvier 1977).

Facturation par C.N.R.S. CIRCE des ressources centrales 370 et périphériques, services et unités spéciales.

2. QUELQUES FICHIERS

Des fichiers de descriptions ont été constitués en se servant de plusieurs vocabulaires.

O.R.S.T.O.M. Fonds Documentaire
N° : 216 ex 1
Cote : B

DONNEES : INRA Science du Sol - Montpellier

PROGRAMME : Bibliothèque
Cartolab J.-L. MALLET

174 Fichier 112 : Ce fichier contient des enregistrements constitués à partir de 112 variables binaires, nominales, ordinales, et à intervalles, toutes descriptives de l'horizon du profil. Il contient à ce jour 835 profils de 5 à 10 horizons. Le temps CPU moyen d'interrogation est de 0,36 s/horizon.

Fichier 14 : Ce fichier est constitué d'enregistrements de 14 variables temporo-spatiales et classificatoires pour chaque profil de la B.D.P. A ce jour, 1 065 références de profils. Interrogation en 0,12 s/profil.

Fichier 110 : Fichier expérimental créé en 1976 avec des descriptions d'environnement. Les enregistrements portent 110 variables nominales et ordinales. 68 environnements à ce jour. Interrogation en 0,61 s/environnement.

3. PROGRAMMES USUELS

★ Interrogation des fichiers par sélection booléenne dans le langage d'un vocabulaire de référence ; sortie imprimée et encodée en Fieldata au format 15 sur bande magnétique BASF, 7 pistes, 800 bpi. Le langage d'interrogation est naturel et convient à l'utilisateur non informaticien. Le programme SGYT est en FORTRAN pour 1108, 906 lignes d'instructions, 80 K-mots ; accepte jusqu'à 100 conditions booléennes.

★ Programme équivalent SGY, 893 lignes, 74 K-mots, pour une sortie cartes perforées.

★ Sélection et représentation graphique de l'emplacement des profils à l'intérieur d'un périmètre géographique. Programmes enchainés COOR, 240 lignes, 9 K-mots, SGYT, et CARTO, 507 lignes, 17 K-mots, à couverture mondiale, 0,32 s/profil.

★ Création et mises à jour des fichiers avec contrôle de concordance des données, suivi d'impression, annotation, et traduction éventuelle. Les fiches sont perforées en langage naturel. L'ordre des données est quelconque et peut être différent d'un horizon à l'autre. Quand les corrections éventuelles ne touchent qu'un faible pourcentage des horizons d'un lot de profils, on ne relit que les horizons corrigés. Programme DGY, 943 lignes, 69 K-mots, acceptant 10 000 profils (unités) à 1-10 enregistrements (horizons, profondeurs, sous-unités, etc.). Temps moyens : 1,8 s/112 variables, 0,9 s/14 variables, 2,4 s/110 variables.

★ Création et mises à jour des vocabulaires de référence. Programme RGY, 763 lignes, 68 K-mots, moins de 500 variables, moins de 60 000 mots par vocabulaire, moins de 150 caractères (chiffres, lettres, caractères ★ / , - \$ et espacements) par donnée. Temps : 40 s/112 variables, 85 s/14 variables, 32 s/110 variables.

★ Classification à partir de tout indice de dissemblance ; sortie d'une matrice de dissemblances moyennes intra et interconstellations. La définition d'une constellation repose sur des relations d'ordre entre les divers types de dissemblances moyennes. Programme AGGLOM, 427 lignes, 89 K-mots, $m \leq 100$ enregistrements, manquants acceptés, 47 s/100 enregistrements.

145
★ Classification à partir de dissemblances de rangs, processus agrégatif pas à pas, dendrogramme en sortie. Programme CAVT, 282 lignes, 87 K-mots, $m \leq 127$, $v \leq 2 000$, $mv \leq 38 100$, manquants acceptés, 35 s/25 enregistrements de 234 variables.

★ Dissemblances multivariées à partir de probabilités. Version GDL adaptée du programme reçu du Dr Goodall, 852 lignes, 38 K-mots, $m \leq 100$, $v \leq 100$ (dont 50 variables nominales ou ordinales à moins de 20 données), manquants acceptés, 1 mn/52 enregistrements de 22 variables.

★ Dissemblances multivariées normées par les données maximales. Programme DISSIM, 201 lignes, 47 K-mots, $m \leq 250$, $v \leq 400$, manquants acceptés, 7 mn/100 profils à 399 variables, 10 s/25 enregistrements à 234 variables.

★ Classement multivariable d'un profil dans un fichier référence de 230 profils d'Afrique centrale à 209 variables de terrain. Programmes enchainés SGYT, MERGE, 39 lignes, 8 K-mots, IDENT ou IDENT2, 99 lignes, 12 K-mots, $m \leq 750$, $v \leq 400$, 30 s/profil, manquants acceptés.

★ Test des corrélations de rangs. Programme CSAT, 478 lignes, 60 K-mots, $m \leq 127$, manquants acceptés, 12 s/43 enregistrements de 16 variables.

★ Limites de confiance des médianes. Programme MEDT, 187 lignes, 57 K-mots, $m \leq 127$, $v \leq 2 000$, $mv \leq 38 100$, manquants acceptés, 17 s/43 enregistrements de 50 variables.

★ Outre ces douze programmes FORTRAN pour 1108, seize programmes FORTRAN sont utilisés sur 1108 ou sur 370. Certains n'acceptent pas de manquants.

4. TRANSPOSITION EVENTUELLE DES PROGRAMMES

Les programmes FORTRAN SGYT, SGY, DGY, RGY, COOR et CARTO ne sont pas transposables sur 370. La version 1108 des programmes 370 n'est pas encore disponible. Une version CDC 3600 des programmes SGY, DGY, et RGY est disponible depuis 1970. La transposition des programmes AGGLOM, CAVT, DISSIM, CSAT et MEDT sur 370 ou 360 ne poserait aucun problème.

5. OBJECTIFS DE LA B.D.P.

Dès sa création en 1968, la banque de données pédologiques s'était fixé pour objectifs la recherche et le service dans la gestion automatisée et le traitement statistique de profils décrits outre-mer.

L'encodage manuel des descriptions ne pouvait être envisagé faute de personnel. L'encodage automatique impliquait l'utilisation d'un glossaire par tous les pédologues de l'Office. Ce glossaire fut rédigé en collaboration avec tous les organismes français, sur l'initiative de la DGRST, et édité en 1969.

L'objectif gestion a pu être atteint par la création d'un SGBD performant, dont toutes les fonctions (création d'enregistrements, mises à jour et inter-

rogations) ont la propriété d'être écrites dans le langage naturel du glossaire. En outre, ce SGBD reste indépendant du contenu des glossaires. Le logiciel de gestion a été entièrement écrit et mis au point début 1970, par l'un de nous (A. Giey), sur l'ordinateur UNIVAC 1108 et la calculatrice CDC 3600.

La statistique mathématique reste un objectif majeur après les opérations menées à bien ces deux dernières décennies.

Un objectif est venu s'ajouter sous la forme d'une fusion de données de laboratoire avec les données de terrain. La diversité des méthodes analytiques limite toutefois cet objectif à des sous-ensembles homogènes.

6. UTILISATEURS

Outre le personnel permanent (2 unités) qui prend en charge les travaux de service pour certains pédologues, des utilisateurs accèdent seuls au matériel après avoir maîtrisé les modes d'emploi des programmes de leur choix. Une expérience d'accès par voie postale a réussi.

7. STAGIAIRES

Les possibilités d'accueil de stagiaires se résument en un effectif maximal de 10 pendant trois semaines consécutives une fois par an. Le stage est gratuit, sans hébergement. Les stagiaires ont accès au matériel, aux fichiers et aux programmes du système POSEIDON pendant un temps CPU total limité.

8. BIBLIOGRAPHIE SOMMAIRE

Biom.-Prax., 2, 131-132 ; 245-259.

An. Universid. Hisp., 23, 51-69.

Oléagineux, 18, 35-42.

Oléagineux, 19, 175-177.

Biom.-Prax., 5, 20-25.

An. Edaf. y Agrob., 23, 305-321 ; 447-460 ; 461-469 ; 563-572 ; 697-704.

1^{er} coll. eur. contr. nutr. min. et fert., 273-286, 1964.

Biom.-Prax., 6, 36-47.

Fertilité, 25, 41-45.

Cah. O.R.S.T.O.M., sér. pédol., 3, 79-88.

Cah. O.R.S.T.O.M., sér. pédol., 4, 91-96.

Bull. serv. carte géol. Als.-Lorr., 20, 257-276.

An. Edaf. y Agrob., 26, 733-762.

Cah. O.R.S.T.O.M., sér. pédol., 6, 21-34.

Sc. Sol, 8, 55-78.

2 col. eur. contr. fert., 190-213, 1968.

Stat. theory and method abstr., 10, 510.

Oecol. Plant., 4, 99-122 ; 237-266.

Init. doc. tech. O.R.S.T.O.M., n° 13.

An. Edaf. y Agrob., 29, 317-341.

U.N.E.S.C.O.- Igu exp. meet. envir. data proc., 1970, 31.

3 col. eur. contr. fert., 1971.

Allionia, 18, 13-32.

Rev. Ecol. Biol. Sol, 9, 617-628.

F.A.O. U.N.E.S.C.O. consult. comput. soil data, 26, 1972.

2nd U.N.E.S.C.O.-Igu symp. geogr. inf. syst., 1131-1200, 1972.

Bull. Ecol. Res. Comm., 17, 433-442.

Cah. O.R.S.T.O.M., sér. pédol., 11, 257-264.

Sem. ét. sol et fert., 1973.

Cah. O.R.S.T.O.M., sér. pédol., 12, 125-132.

Init. doc. tech. O.R.S.T.O.M., n° 25.

Cah. O.R.S.T.O.M., sér. pédol., 13, 175-190 ; 173 ; 223-226.

Init. doc. tech. O.R.S.T.O.M., n° 30.

Isss catal. soil inf. syst., F-1, 1975.

1st meet. Isss soil inf. syst., 17-21, 1975.

H. s. O.R.S.T.O.M., 1976.

2nd meet. Isss soil inf. syst., 1977.

9. EXTRAIT DE LA FICHE DE TERRAIN DE L'O.R.S.T.O.M.

(Voir page 178.)

Extrait de la fiche de terrain de l'O.R.S.T.O.M.

HORIZON/HRZ 1/DE	A	CM/UN	//
- HORIZON/HRZ 2/DE	A	CM/UN	//
- HORIZON/HRZ 3/DE	A	CM/UN	//
- HORIZON/HRZ 4/DE	A	CM/UN	//
- HORIZON/HRZ 5/DE	A	CM/UN	//

Fiche 1977

12345 sec
12345 légèrement humide
12345 humide
12345 très humide
12345 noyé

1	.	/.	humide
2	.	/.	humide
3	.	/.	humide
4	.	/.	humide
5	.	/.	humide
1	.	/.	sec
2	.	/.	sec
3	.	/.	sec
4	.	/.	sec
5	.	/.	sec

12345 sans taches
12345 quelques taches
12345 taches
12345 nombreuses taches
12345 très nombreuses taches

12345 peu étendues
12345 étendues

1	*	/.	*
2	*	/.	*
3	*	/.	*
4	*	/.	*
5	*	/.	*

12345 liées aux faces des unités structurales
12345 associées aux vides

12345 apparemment non organique
12345 à matière organique non directement décalable
12345 à matière organique directement décalable
12345 à débris organiques

1 teneur en matière organique voisine de PC
2 teneur en matière organique voisine de PC
3 teneur en matière organique voisine de PC
4 teneur en matière organique voisine de PC
5 teneur en matière organique voisine de PC
12345 moins de 1PC de matière organique

12345 aucune effervescence
12345 faible effervescence
12345 effervescence
12345 vive effervescence

12345 localisée
12345 généralisée
12345 généralisée irrégulièrement répartie

12345 moins de 2PC
12345 de 2 à 15PC
12345 de 15 à 30PC
12345 de 30 à 60PC
12345 plus de 60 PC

12345 éléments carbonatés
12345 éléments gypseux
12345 éléments carbonatés et gypseux

12345 diffus
12345 en pseudomycélium
12345 en amas
12345 en nodules friables
12345 en nodules
12345 en pédodes
12345 en septarias
12345 en encroûtement
12345 en croûtes
12345 en dalles
12345 en pellicules rubanées
12345 lithiques
12345 en macrocristaux
12345 en revêtements autour d éléments grossiers
12345 en pellicules d arrachement

12345 et diffus
12345 et en pseudomycélium
12345 et en amas
12345 et en nodules friables
12345 et en nodules
12345 et en pédodes
12345 et en septarias

12345 très peu de cailloux
12345 cailloux peu abondants
12345 cailloux
12345 cailloux abondants
12345 cailloux très abondants

12345 très peu de blocs
12345 blocs peu abondants
12345 blocs
12345 blocs abondants
12345 blocs très abondants

12345 de roche sédimentaire détritique
12345 de roche sédimentaire argileuse
12345 de roche sédimentaire calcaire
12345 de roche sédimentaire dolomitique
12345 de roche sédimentaire phosphatée
12345 de roche sédimentaire siliceuse non détritique
12345 de roche sédimentaire saline
12345 de roche sédimentaire
12345 de roche métamorphique
12345 de roche ignée grenue
12345 de roche ignée microgrenue
12345 de roche ignée microlithique
12345 de roche ignée ultrabasique
12345 de roche ignée

1

2

3

4

5

12345 basique
12345 acide

12345 tendre
12345 dure

12345 de forme arrondie
12345 de forme allongée
12345 de forme aplatie
12345 de forme irrégulière

12345 à arêtes anguleuses
12345 à arêtes émoussées

12345 non altérée