

## Comparaisons de profils du Congo par des distances

Dominique MARTIN\* et Anne-Marie AUBRY\*\*

\*ORSTOM, Brazzaville, BP 181 (Congo)

\*\*Services Scientifiques Centraux de l'ORSTOM, Bondy (France)

### RÉSUMÉ

On expose les principes suivis pour comparer les profils à partir des informations de base que sont les données des variables descriptives de chaque horizon. Deux méthodes non-paramétriques sont proposées, utilisant le même programme pour ordinateur. Elles conduisent à deux relations de dissemblance entre profils, notées respectivement  $d$  et  $d_a$ . On expose en détail le processus programmé du calcul, on cite la méthode de regroupement utilisée et on explique le processus itératif suivi pour obtenir des résultats plus fins.

On décrit la façon dont ont été exploitées les données descriptives des 24 profils sujets de l'étude et les résultats obtenus. On étudie, en outre, à part, 11 profils regroupés sur leur ressemblance par  $d_a$ , puis 20 appumites. Les résultats, au risque 1 %, du test de corrélation entre les variables sont présentés.

### INTRODUCTION

La méthodologie élaborée par la Banque de Données Pédologiques de l'ORSTOM pour comparer statistiquement des profils a été suivie sur un fichier de 24 profils du Congo extraits de la bande magnétique qui supporte aujourd'hui environ 700 descriptions normalisées de profils de sols très divers. L'opération est décrite, étape par étape.

La méthodologie en question reste valable lorsque les analyses physico-chimiques et enzymatiques com-

### ABSTRACT

This is an outline of method to compare soil profiles using descriptive data from each horizon expressed as variables. Two non-parametric methods are suggested here, using the same computer programme. They lead to the establishment of relationships of dissimilarity between profiles, to be noted as  $d$  and  $d_a$  respectively.

The programmed calculation process is described in detail, the method of grouping cited, and the iterating process followed in order to obtain more detailed results is explained.

Descriptive data, from the 24 profiles studied has been treated, with explanation of method and results. Also studied are 11 other profiles grouped according to similarity by  $d_a$ , in addition 20 appumites. All the results are presented together with a correlation test of the variables which allows for an error of 1 % only.

plètent les descriptions, ou lorsque les effectifs sont supérieurs à celui de l'échantillon choisi ici. Elle reste inchangée, en outre, quel que soit le vocabulaire normalisé utilisé par les observateurs.

## 1. EXPOSÉ DES MÉTHODES STATISTIQUES

### 1.1. Introduction

En interrogeant la banque des profils (maximum 10 000 profils), on extrait le fichier dont les caractéris-

tiques — de nature géographique, classificatoire, morphologique, ou autre — satisfait les conditions données.

Chaque profil est décrit horizon par horizon. Pour chaque horizon, les données des variables descriptives — éventuellement et/ou de laboratoire — constituent un « enregistrement ». L'ensemble des enregistrements relatifs aux profils sélectionnés pour une étude donnée constituent un « fichier ». Ce fichier se présente à la fois sous forme de sortie listée en clair et de sortie perforée en codes numériques. Si certaines données manquent dans un enregistrement elles sont repérées par le code -1.

Le problème est de comparer les profils entre eux. Cette comparaison doit tenir compte de toutes les informations contenues dans les descriptions, c'est donc une comparaison statistique multivariable que l'on fait. D'autre part, la nature même des variables impose des méthodes non paramétriques. En effet, les profils sont caractérisés, dans chaque horizon, par les 3 types de variables suivants :

(a) *quantitatives* (« interval variables »), pour lesquelles il y a une métrique. Par exemple, dans le cas de pourcentages, le code 2 de la donnée libellée « 2PC DE MØ » est la moitié de 4.

(b) *ordonnées* (« ordinal variables »), pour lesquelles les données sont liées par des relations d'ordre. Par exemple, dans le cas de la clarté et de la saturation Munsell, le code 4 n'est pas le double de 2, mais seulement plus grand que 2.

(c) *binaires* (« binary variables »), dont les 2 seules données sont oui et non s'excluant mutuellement, ou présence-absence. Le code 1 indique la présence de la variable.

## 1.2. Méthode de distance $d$ séparant les profils

Une méthode non-paramétrique qui compare des profils est la mesure généralisée des « distances » les séparant dans l'espace multidimensionnel des variables qui les caractérisent.

Il faut donc d'abord définir cet espace multidimensionnel dont les points seront les  $m$  profils de l'étude. Chaque profil étant caractérisé par la succession de ses  $n$  horizons ( $n$  variant d'un profil à un autre), et chaque horizon par les données des  $v$  variables descriptives retenues pour l'étude, il est logique de caractériser l'espace des profils dans  $n \times v$  dimensions, en rendant  $n$  constant pour l'étude donnée.

Pour ce faire, le pédologue choisit, à partir du fichier de base initial sélectionné, une succession-type de  $n$  horizons qui « colle » le mieux à l'étude, c'est-à-dire qui permette de prendre en compte le maximum d'informations significatives.

Or 2 cas se présentent en fait :

*Cas (a)* Dans les descriptions, le nom ( $A_3$ ,  $B_{22}$ , par exemple) des horizons est fourni.

S'il reste des profils où la succession des horizons décrits est incomplète par rapport à la séquence de référence, ou va la compléter par des horizons fictifs où les données des variables sont « manquantes » (code -1, comme on l'a dit plus haut). Si, au contraire, un horizon décrit est « en plus » à première vue, mais que dans la séquence de référence s'en trouvent 2 (ou 3) correspondant à une subdivision d'un tel horizon, on répète 2 (ou 3) fois ses données. Un exemple est représenté dans le tableau 1.

On a  $m$  (= 4) profils,  $v$  (= 80) variables descriptives de l'horizon et on choisit  $n = 5$ , avec la séquence de référence  $H_1 = A_1$ ,  $H_2 = A_3$ ,  $H_3 = B_1$ ,  $H_4 = B_{21}$ ,  $H_5 = B_{22}$ .

*Cas (b)* Dans certains profils, le nom des horizons n'est pas fourni.

On choisit alors  $n$  niveaux de profondeurs :  $1 \times k$ ,  $2 \times k$ , ...,  $n \times k$  cm,  $k \in \mathbb{N}$ , par exemple, la profondeur  $n \times k$  étant la profondeur maximale des  $m$  profils de l'étude. Pour chaque profil, on prend les données des variables décrites à ces profondeurs, en complétant pour les moins profonds par des tranches fictives de « manquants », de façon similaire au cas précédent.

Dans les 2 cas, nous avons  $m$  points dans un espace à  $n \times v$  dimensions et nous calculons les distances non paramétriques entre ces points sur la base d'un tableau de données constitué ainsi qu'il est figuré tableau 2.

## 1.3. Processus programmé du calcul des distances

Par commodité nous allons appeler dans la suite du texte « distance »  $d$  qui est en fait une « relation de dissemblance » au sens strict mathématique.

Pour calculer la distance entre  $p$  points dans un espace à  $l$  dimensions, on part du tableau à  $p$  lignes et  $l$  colonnes des données (brutes ou calculées selon le cas),  $2 \leq p \leq 100$ ,  $1 \leq l \leq 400$ , les bornes supérieures des paramètres étant dues à la taille de mémoire centrale disponible de l'ordinateur UNIVAC 1108 qui effectue les travaux.

(a) On remplace le tableau  $X$  de données  $x_{ki}$  ( $1 \leq k \leq p$ ,  $1 \leq i \leq l$ ), à  $p$  lignes et  $l$  colonnes, par un tableau  $R$  de rangs  $r_{ki}$ , de mêmes dimensions, le « rangement » étant fait colonne par colonne indépendamment, de la façon suivante :

Si  $i$  désigne l'indice courant de colonne (donc de variable),  $i \in \mathbf{N}$ ,  $1 \leq i \leq l$ , pour chaque  $i$ ,

— on remplace les données non manquantes par des numéros d'ordre 1 à  $p_i$  ( $p_i = p$  - le nombre de données manquantes de  $i$ );

— on remplace les rangs des données identiques par les moyennes de ces rangs;

— les données « manquantes » gardent le code -1.

Soit, en illustration, l'exemple fictif d'une variable  $i$  dont les données sont extraites d'un tableau à  $p = 7$  lignes :

$$x_{.i} = \begin{bmatrix} 0.01 \\ 3 \\ 4.2 \\ 0.01 \\ -1 \\ 0.01 \\ 4.2 \\ 2 \end{bmatrix} \rightarrow \rightarrow \rightarrow \begin{bmatrix} 1 \\ 5 \\ 6 \\ 2 \\ -1 \\ 3 \\ 7 \\ 4 \end{bmatrix} \rightarrow \rightarrow r_{.i} = \begin{bmatrix} 2 \\ 5 \\ 6.5 \\ 2 \\ -1 \\ 2 \\ 6.5 \\ 4 \end{bmatrix}$$

(b) Pour chaque  $i$ , on calcule  $I_i$ , facteur de pondération des données manquantes :

$$I_i = \frac{1}{12} \left[ p_i^3 - p_i - \sum_{q_i=1}^{e_i} (t_{q_i}^3 - t_{q_i}) \right],$$

où :

$p_i$  est l'effectif réel de la variable  $i$ ,  $1 \leq p_i \leq p \leq 100$ ,

$e_i$  est le nombre de lots  $q_i$  de rangs identiques de  $i$ ,  $1 \leq e_i \leq p_i$ ,

$t_{q_i}$  est le nombre de données identiques du lot  $q_i$ ,  $1 \leq t_{q_i} \leq p_i$ .

Remarques : valeurs limites de  $I_i$

1° toutes les données de  $i$  sont distinctes :  $e_i = p_i$  et ( $\forall i, t_{q_i} = 1$ ), alors  $I_i = \frac{1}{12} (p_i^3 - p_i)$

2° toutes les données de  $i$  ont même valeur (la variable est « constante » sur l'unité d'étude) :  $e_i = 1$ ,  $t_{q_i} = p_i$ , alors  $I_i = 0$ .

(c) L'expression de la distance entre 2 unités

d'étude (profils, horizons ou autres) d'indice respectifs  $h$  et  $k$  ( $h \neq k$ ,  $1 \leq \frac{h}{k} \leq p$ ) s'écrit alors (Kendall, Stuart, 1966) :

$$d_{hk} = \frac{1}{l} \sum_{i=1}^l \frac{(r_{hi} - r_{ki})^2}{I_i} \quad (1)$$

Pour un  $i$  donné, si  $I_i = 0$ , ou si  $r_{hi}$  ou  $r_{ki} = -1$ , la variable  $i$  est abandonnée et  $l$  est diminué de 1 dans l'expression (1). A la limite, si  $l = 0$ ,  $d_{hk}$  est incalculable.

Comme on le voit cette distance est indépendante de la fonction de répartition des variables. Elle prend ses valeurs dans [0,2].

REMARQUES

Le programme offre la possibilité de choisir les variables à retenir pour un calcul parmi les  $l$  du fichier (constitué ainsi qu'on l'a représenté tableau 2, où  $p = m$  et  $l = n \times v$ ,  $n$  étant le nombre d'horizons types, ou nombre de tranches,  $v$  le nombre de variables descriptives de l'horizon ou tranche), sans toucher à ce fichier.

Cette facilité a été utilisée à plusieurs reprises dans cette étude. Elle permettrait également, par approximations successives, de reconnaître les variables — au sens large du terme — différenciant le plus les unités, quelles qu'elles soient.

1.4. Méthode de distance  $d_d$  séparant les profils

Quand on a la nomenclature des horizons décrits (cas (a) de 1.2.), on adjoint, à titre expérimental, une comparaison des  $m$  profils de l'étude par la méthode suivante, utilisant le même programme de distance, différemment.

(a) Reprenant les descriptions originelles des  $m$  profils, horizon par horizon (tabl. 1), pour chaque profil, on calcule les distances entre ses  $n$  horizons dans l'espace à  $v$  dimensions. D'où  $C_n^2$  résultats par profil. On les notera fréquemment par la suite « distances intra-profil ».

REMARQUE : Dans le calcul de l'expression (1),  $l = v$ ,  $p = n$ .

(b) Puis, dans l'espace à  $C_n^2$  dimensions (des distances entre horizons appariés), on calcule les distances entre les  $m$  points (profils) caractérisés par ce nouveau type de données. Ce ne sont plus, comme

TABLEAU 1

Forme des tableaux des données brutes d'origine relevées sur chaque profil de l'Etude, à partir desquels on constitue l'entrée pour le calcul des distances  $d$  entre profils

On a choisi ici  $n = 5$ ,  $m = 4$ ,  $v = 80$

Les données brutes sont codées, y compris les « manquants »

Variables	Profil 1			Profil 2			Profil 3			Profil 4		
	V1	...	V80	V1	...	V80	V1	...	V80	V1	...	V80
Horizons												
H1 → A1		Données		A1	Données		A1	Données		A1	Données	
H2 → A3		Données		A3	Données		A3	Données		A3	Données	
H3 → B1		Données		B1	Données			manquants		B1	Données	
H4 → B21		Données					B21	Données				
H5 → B22		Données		B2	Données		B22	Données		B2	Données	

TABLEAU 2

Tableau d'entrée pour un calcul de distances  $d$  entre  $m$  profils.

Il contient les données brutes codées, ainsi que les « manquants » d'origine, ou ajoutés (code : -1)

Variables	H1 ou tranche 1				H2 ou tranche 2				...	Hn ou tranche n			
	V1/H1	V2/H1	...	Vv/H1	V1/H2	V2/H2	...	Vv/H2		V1/Hn	V2/Hn	...	Vv/Hn
Profil 1			...				...		...			...	
Profil 2			...				...		...			...	
.			...				...		...			...	
.			...				...		...			...	
Profil $m$			...				...		...			...	

en 1.2., des données brutes, mais des données calculées, nombres réels compris entre 0 et 2, ou -1 (distances incalculables). Voir tabl. 3.

La relation de dissemblance entre profils ainsi obtenue peut être notée  $d_a$ .

REMARQUE : Dans le calcul de l'expression (1),  $l = C_n^2$ ,  $p = m$ .

### 1.5. Regroupement d'unités en constellations

La méthode est expliquée en détail par Van den Driessche (1974) pour la partie théorique, et un exemple numérique en est complètement traité par le même auteur (1965), en remarquant que dans cet exemple les mesures de ressemblance à partir desquelles sont faits les regroupements sont les  $D^2$  de

TABLEAU 3

Forme du tableau d'entrée pour un calcul de distances  $d_a$  entre  $m$  profils. Cet exemple correspond à celui présenté tableau 1. Les données sont ici des données calculées: les distances entre horizons appariés.

Variables	$d_{A1, A3}$	$d_{A3, B1}$	$d_{B1, B21}$	$d_{B21, B22}$	$d_{A1, B1}$	$d_{A1, B21}$	$d_{A1, B22}$	$d_{A3, B21}$	$d_{A3, B22}$	$d_{B1, B22}$
Profils										
Profil 1										
Profil 2				0		2 fois la même distance		2 fois la même distance		
Profil 3		-1	-1		-1					-1
Profil 4				0		2 fois la même distance		2 fois la même distance		

Mahalanobis. La méthode et le programme sont valables quel que soit le type de distance,  $d$  ou  $d_a$  en particulier.

### 1.6. Itération automatique programmée des calculs de regroupements

On utilise un processus itératif qu'on limite à 3 étapes.

(a) Calcul des distances  $d$  ou  $d_a$  entre les points initiaux sur la base du tableau des données (brutes ou calculées).

(b) Regroupement des points en constellations selon la méthode programmée citée au 1.5., à partir des distances obtenues au (a).

(c) Calcul des points moyens de chacune des constellations constituées au (b) (en faisant la moyenne des données de chaque variable). Chaque point moyen sera considéré comme représentatif de l'ensemble des points regroupés dans la constellation correspondante et s'y substitue.

(a') Reprise du calcul en (a). L'ensemble des points

initiaux de l'étape (a) a été réduit en prenant les points moyens obtenus au (c), etc.

Ce processus d'itérations est simplement commandé par cartes contrôle système, les programmes de calculs, pour leur part, ayant été écrits pour échanger, par l'intermédiaire de supports magnétiques auxiliaires, les informations nécessaires à sa réalisation.

### 1.7. Corrélations totales entre variables

Le test de corrélation utilisé ici est celui de rangs publié par Spearman (1904). Il est indépendant de la fonction de répartition. La méthode est décrite en détail par Aubry *et al.* (1973).

## 2. EXPLOITATION D'UN FICHER DE 24 PROFILS

### 2.1. Le fichier

Le fichier des 24 profils a été constitué à partir d'une trentaine de profils du Congo : il s'agit de profils

observés classiquement en cartographie, répartis dans trois secteurs géographiques du pays, et qui n'ont pas fait l'objet d'observations plus détaillées que ne le permet le glossaire (Glossaire de Pédologie, 1969). La plus grande partie de ces profils a été observée par R. Bosseno, technicien de l'ORSTOM, puis mis sur fiches par le pédologue, après contrôle sur le terrain et au bureau. Quelques profils non ferrallitiques ont été retirés, ainsi que deux profils ne comprenant que deux horizons ( $A_1$  et  $B_2$ ). Le fichier des 24 profils ainsi constitué n'offrait pas a priori de « problème d'étude » particulier et son exploitation n'avait au départ pour but que d'expérimenter la méthodologie conçue par la Banque de Données Pédologiques de l'ORSTOM.

L'utilisation de méthodes statistiques et l'organisation du profil pédologique obligent à définir une unité sur laquelle portera les calculs : celle-ci peut être une tranche de profondeur comme déjà dit, mais l'utilisation de l'horizon est évidemment plus parlante. Les horizons sont groupés en profil, stade moyen universellement adopté de l'observation d'une organisation pédologique : se pose donc au départ le choix d'une nomenclature d'horizons et surtout le problème de la succession des horizons du profil et de la possibilité d'introduire des manquants par rapport à une succession-type.

L'examen rapide et visuel du fichier a permis de se rendre compte que la subdivision  $A_{11}+A_{12}$  de l'horizon  $A_1$  n'est opérée qu'une fois et que les horizons  $B_3$  et C sont présents deux fois ; en revanche, la subdivision  $A_{31}+A_{32}$  de l'horizon  $A_3$  est introduite trois fois ; deux profils n'ont pas d'horizon  $A_1$  (érosion de l'horizon humifère sous savane) ; les horizons  $A_3$ ,  $B_1$  et  $B_{23}$  manquent fréquemment ; l'horizon  $B_2$  n'est pas subdivisé dans 7 profils. Pour faciliter les comparaisons, il a paru intéressant de doubler les horizons  $B_2$  en  $B_{21}$  et  $B_{22}$ , en utilisant deux fois les mêmes descriptions. Le même processus n'a pas été suivi pour les horizons  $A_1$  et  $A_3$  en raison de l'absence, trop fréquente, des subdivisions  $A_{11}+A_{12}$  et  $A_{31}+A_{32}$  ; il en est de même pour l'horizon  $B_{23}$ , manquant à ce premier stade des calculs, celui-ci étant souvent gravillonnaire quand il est présent. Il est bien sûr impossible de remplacer des horizons  $A_1$ ,  $A_3$  et  $B_1$  manquants : il s'agit alors d'une caractéristique propre et importante du profil.

Ce stade de complètement du fichier peut se placer, comme ici, avant tout calcul, en raison du petit nombre de profils, mais peut aussi se faire, en présence d'un fichier plus important, après le calcul des distances intra-profil, ce qui permet de se faire une idée de la succession-type des horizons dans le fichier.

## 2.2. Distances $d_d$

### 2.2.1. DISTANCES INTRA-PROFIL

Chaque profil est traité à part, ce qui permet de mesurer les distances entre tous les horizons pris deux à deux, à partir des données descriptives. La sortie imprimée des résultats donne le nom de toutes les variables qu'on a choisies pour les calculs et indique également celles qui sont effectivement utilisées. Ne pas oublier qu'une variable dont les données seraient identiques dans tous les horizons du profil, ne serait pas utilisée à ce stade du calcul, bien que présente.

Par collationnement des 24 profils on a pu dresser une liste maximale d'horizons présents :

$A_{11}$ ,  $A_{12}$ , ( $A_1$ ),  $A_{31}$ ,  $A_{32}$ , ( $A_3$ ),  $B_1$ ,  $B_{21}$ ,  $B_{22}$ ,  $B_{23}$ , ( $B_2$ ),  $B_3$ , C.

L'ensemble des valeurs des distances intra-profil tirées des 24 listes de résultats a été rassemblé dans le tableau 4 qui appelle les remarques suivantes :

— les horizons  $B_3$  et C, très peu représentés, ont été retirés du tableau, ainsi que des calculs suivants ;

— les subdivisions  $A_{11}-A_{12}$  et  $A_{31}-A_{32}$ , peu utilisées, ne figurent pas dans le tableau pour ne pas le surcharger de données nulles ou identiques, mais les données correspondantes sont utilisées dans certains des calculs suivants.

Des données absentes, nulles ou identiques subsistent dans les cas suivants :

— absence de données, due à l'absence d'un des deux horizons : par exemple, pas de distance  $A_3-B_1$  pour le profil DIH 49 en raison de l'absence de l'horizon  $B_1$  ;

— donnée nulle évidente quand on a dédoublé un horizon (§ 2.1.) : horizon  $B_2$  remplacé par deux horizons identiques  $B_{21}$  et  $B_{22}$ .

Le tableau des distances intra-profils a été séparé en deux verticalement, selon que l'on a affaire à des horizons successifs ou non : on a d'ailleurs replacé dans la partie gauche du tableau des données normalement à droite, mais qui correspondent à des horizons successifs en raison de l'absence d'un horizon : par exemple distance  $A_3-B_{21}$  du profil DIH 49 en raison de l'absence de l'horizon  $B_1$ .

Dans le tableau 4 toutes les distances ont été multipliées par 100. Dans la partie gauche de ce dernier (distances entre horizons successifs), on a mis en

TABLEAU 4

Distances intraprofil multipliées par 100. Distances maximales en italique et distances  $\leq 10$  en gras

Numéro de profil																	Nombre de variables utilisées
	$d_{A1, A3}$	$d_{A3, B1}$	$d_{B1, B21}$	$d_{B21, B22}$	$d_{B22, B23}$	$d_{A1, B1}$	$d_{A1, B21}$	$d_{A1, B22}$	$d_{A1, B23}$	$d_{A3, B21}$	$d_{A3, B22}$	$d_{A3, B23}$	$d_{B1, B22}$	$d_{B1, B23}$	$d_{B21, B23}$		
1	SMB 1	38	26	15	16	16	45	56	54	77	29	38	61	19	38	38	46
2	A1, A31, A32, B1, B21, B22, B23	34	25	20	10	39	59	62	64	64	27	42	54	26	46	44	37
3	SMB 6	56	32	40	14		80	81	98		41	57		57			35
4	A1, A3, B1, B21, B22	25	46	32	9	39	65	61	64	47	45	52	57	46	63	43	45
5	A1, A3, B1, B21, B22, B23	44	41	5	16		92	97	106		40	49		22			27
6	DIH 8	29	36	44	20	35	48	62	61	62	49	55	51	40	55	42	48
7	A1, A3, B1, B21, B22, B23	35	65	29	0		103	109	109		66	66		29			29
8	A1, A3, B1, B2	52	33		69			71	108		33	80					37
9	DIH 49	59	36	71	0		100	122	122		51	51		71			37
10	A1, A3, B1, B2	52	63	18	0		96	116	116		76	76		18			31
11	A1, A3, B1, B2	35	14	7	35	7	52	64	73	75	27	57	61	40	49	43	33
12	A1, A3, B1, B21, B22, B23	36	37	7	31		71	82	98		45	67		41			33
13	DIH 21	41	21	37	51		61	90	94		48	62		47			39
14	A1, A3, B1, B21, B22	31	33	39	4		73	77	82		58	47		47			32
15	DIH 64	26	40	27	10		65	87	89		61	64		41			29
16	A1, A3, B1, B21, B22	43	63	39	0		94	111	111		65	65		39			32
17	DIH 61	30	37	20	34	1	54	69	60	63	52	57	57	44	43	33	49
18	A1, A3, B1, B21, B22, B23	39	59	0			81	81		59	59						44
19	A1, A3, B2	28	46	53	13		63	47	62	46	34	55		64	13		45
20	CMB 15	32	44	17	47		55	82	91	44	67				58		37
21	A1, A3, B21, B22, B23	32	10	0			37	37		10	10						34
22	CMB 22																
23	A11, A12, A3, B2		72	90						72	152						32
24	CMB 23																
25	A3, B21, B22	111	79	0			111	131	131		79						32
26	CMB 25																
27	A1, B1, B2		21	13	33					28	55		37				40
28	CMB 28																
29	A3, B1, B21, B22 B3C																

italique les distances maximales et en gras les distances faibles ( $\leq 10$  à l'exception de 0 dont on a déjà donné l'explication). On peut ainsi faire les remarques suivantes concernant les distances maximales :

— plus de 60 % des distances maximales sont observées entre les horizons  $A_1$ - $A_3$  ou  $A_3$ - $B_1$ , ce qui semble normal : l'horizon humifère représente la plus importante différenciation d'un profil ferrallitique dans ses horizons supérieurs (appumite et structichron) (Chatelin et Martin, 1972) ;

— 25 % des profils présentent une distance maximale dans l'horizon B en raison de la présence d'un horizon grossier (gravolite ou gravelon) dans les horizons  $B_{22}$  ou  $B_{23}$ .

L'examen des distances faibles est également intéressant : on les note entre divers horizons successifs à l'intérieur de l'horizon B pour 9 profils. On peut donc se demander, à la limite, s'il était bien nécessaire de distinguer deux horizons, qui se révèlent aussi proches. L'examen des fiches descriptives permet de préciser quelles sont les variables réellement différentes et deux cas sont donnés à titre d'exemple (ne sont mentionnées, à l'exception de la profondeur et de l'épaisseur, que les variables effectivement utilisées dans le calcul de distances et différentes entre les deux horizons) :

Profil	Distance	Variables	Données	
DIH 64	$B_{21}$ - $B_{22} = 4$	Couleur humide Netteté de la structure Taille de la structure	$B_{21}$ 7,5 YR 5/7  Nette  Moyenne et fine	$B_{22}$ 7,5 YR 5/8  Peu nette  Fine
DIH 77	$B_{22}$ - $B_{23} = 4$	Cailloux  Teneur en argile en sable	$B_{22}$ Pas de cailloux  40 % 30 %	$B_{23}$ Cailloux  45 % 25 %

Le tableau des distances entre horizons non successifs (partie droite du tableau 4) est plus difficile à utiliser. Les distances maximales soulignées montrent

bien que les deux horizons les plus éloignés ou les plus différenciés du profil sont les horizons  $A_1$  et  $B_2$  ( $B_{21}$ ,  $B_{22}$  ou  $B_{23}$ ) : dans un seul cas, il s'agit de l'horizon  $B_1$  (profil SMB 12) mais les distances  $A_1$ - $B_1$  et  $A_1$ - $B_{22}$  sont très proches. Les horizons les plus éloignés sont évidemment  $A_3$  et  $B_{22}$  (profils CMB 23 et CMB 28) en l'absence de l'horizon  $A_1$ .

## 2.2.2. DISTANCES INTER-PROFILS

### 2.2.2.1. Données utilisées

Les distances intra-profil du tableau servent de « données » pour un nouveau calcul de distances, cette fois-ci, entre profils, selon la méthode expliquée précédemment. La distance inter-profil, calculée à partir des distances intra-profil mesure en fait le degré de différenciation des horizons d'un profil par rapport à celle de l'autre profil, donc en fait la dissemblance entre deux profils en ce qui concerne la succession et la différenciation de leurs horizons : deux profils ayant une distance faible présentent des successions semblables d'horizons de même importance ; deux profils à distance élevée ont, en revanche, des horizons d'importance inégale et assez différents.

Ces distances inter-profil peuvent ensuite servir de coefficient d'agrégation pour un calcul de constellations, qui groupera les profils à distances inter-profil réduites ou à différenciation d'horizons semblable.

Les distances du tableau 4, complété comme suit, ont donc servi de données et ont été perforées en même temps que les manquants codés -1.

— la succession d'horizons utilisée est la plus complète possible d'après les résultats du § 2.2.1. et après la seule élimination des horizons  $B_3$  et C :  $A_{11}$ ,  $A_{12}$ ,  $A_{31}$ ,  $A_{32}$ ,  $B_1$ ,  $B_{21}$ ,  $B_{22}$ ,  $B_{23}$  (8 horizons) ;

— il existe un maximum de  $C_8^2 (= 28)$  horizons appariés 2 à 2 ;

— les distances  $A_{11}$ - $A_{12}$  et  $A_{31}$ - $A_{32}$  sont nulles quand  $A_1$  et  $A_3$  sont seuls observés ;

— les distances  $A_{11}$ - $B_1$  et  $A_{12}$ - $B_1$  sont identiques à celles de  $A_1$ - $B_1$ , quand n'est observé que le  $A_1$  ;

— les distances  $B_{21}$ - $B_{22}$  sont nulles et  $B_1$ - $B_{21}$  et  $B_1$ - $B_{22}$  identiques quand l'horizon  $B_2$  a été dédoublé avant les premiers calculs (§ 2.1.) ;

— les distances à l'horizon  $B_{22}$  ont été répétées pour l'horizon  $B_{23}$  en l'absence de celui-ci et la distance  $B_{22}$ - $B_{23}$  est considérée comme nulle ;

— quand un horizon manque, les distances incalculables à cet horizon sont notées -1.

2.2.2.2. Résultats obtenus

Le calcul a utilisé toutes les données que sont les distances intra-profil entre les 8 horizons possibles :  $C_8^2 (= 28)$  distances d'horizons appariés. Le tableau de données ainsi constitué (tableau 4 complété pour  $A_{11} + A_{12}, A_{31} + A_{32}$  et les autres distances appariées) comprend de nombreuses valeurs nulles ou identiques, mais aucune des distances intra-profil calculées (§ 2.2.1) n'est perdue.

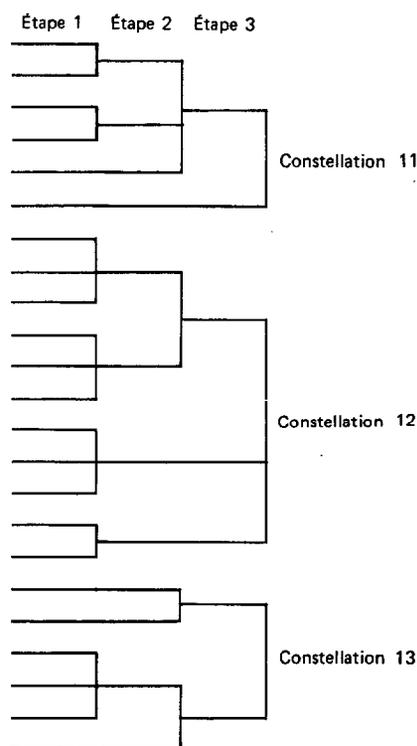
Les programmes calculent les distances et les constellations de profils, puis les moyennes intra-constellations qui permettent une deuxième et troisième étape de regroupement. Les résultats sont donnés par le dendrogramme du tableau 5. Les 24 profils sont ainsi regroupés en 3 constellations + 1 profil et la succession réelle des horizons pour chaque profil

permet de faire les constatations suivantes :

- la constellation 11 est formée de profils à succession  $A_1, A_3, (B_1), B_{21}, B_{22}$  ou  $A_3, (B_1), B_{21}, B_{22}$  : le dernier horizon peut ou non être gravillonnaire mais est toujours peu profond ( $< 120$  cm);
- la constellation 12 comprend les profils qui ont la succession la plus complète d'horizons et comprend tous les profils à horizon  $B_{23}$  :  $A_1, A_3, (A_{31}-A_{32}), B_1, B_{21}, B_{22}, (B_{23})$ ; l'horizon  $B_{23}$  est très gravillonnaire et profond (95 à 340 cm);
- la constellation 13 a une succession d'horizons de type  $A_1, A_3, B_1, B_2$  et comprend le profil DIH 8 qui a une distance  $B_1-B_{21}$  très faible ( $= 5$ );
- le profil isolé CMB 22 représente une succession très particulière d'horizons, avec subdivision de  $A_1$ , absence de  $B_1$  et présence du seul  $B_2$ .

TABLEAU 5

8	DIH 49	A1 A3 B21 B22
12	DIH 21	A1 A3 B1 B21 B22
3	DIH 7	A1 A3 B1 B21 B22
13	DIH 60	A1 A3 B1 B21 B22
24	CMB 28	A3 B1 B21 B22 B3 C
22	CMB 23	A3 B21 B22
14	DIH 64	A1 A3 B1 B21 B22
15	DIH 68	A1 A3 B1 B21 B22
18	CMB 10	A1 A3 B2
4	SMB 12	A1 A3 B1 B21 B22 B23
6	DIH 43	A1 A3 B1 B21 B22 B23
17	DIH 77	A1 A3 B1 B21 B22 B23
1	SMB 1	A1 A31 A32 B1 B21 B22 B23
2	SMB 6	A1 A31 A32 B1 B21 B22 B23
19	CMB 15	A1 A31 A32 B21 B22 B23
11	DIH 53	A1 A3 B1 B21 B22 B23
20	CMB 16	A1 A3 B21 B22 B23
9	DIH 50	A1 A3 B1 B2
23	CMB 25	A1 B1 B2
7	DIH 47	A1 A3 B1 B2
10	DIH 52	A1 A3 B1 B2
16	DIH 61	A1 A3 B1 B2
5	DIH 8	A1 A3 B1 B21 B22
21	CMB 22	A11 A12 A3 B2 B3 C



Dans un deuxième calcul  $d_q$ , on a réduit le nombre d'horizons aux cinq horizons les plus fréquents ( $A_1, A_3, B_1, B_{21}$  et  $B_{22}$ ); puis, en utilisant toutes les données que sont les distances entre horizons appariés

ce qui réduit fortement la présence de valeurs nulles ou identiques, le nombre de données utilisées par profil n'est plus que de  $C_5^2 (= 10)$ , mais on a une perte d'informations.

Les calculs sont conduits de la même manière que précédemment et aboutissent au dendrogramme du tableau 6. Les constellations sont ainsi caractérisées :

— la constellation 21 regroupe 12 profils de type  $A_1, A_3, (B_1), B_{21}, B_{22}, B_{23}$  et est assez peu homogène : l'horizon  $B_1$  manque 4 fois et l'horizon  $B_{23}$  est présent 4 fois, mais il ne faut pas oublier qu'il n'a pas été utilisé dans les calculs ;

— la constellation 22 ne comprend que deux profils assez incomplets, de type  $A_1, (A_3), B_1, B_2$  ;

— la constellation 23 est formée de profils assez variés depuis le plus complet  $A_1, A_{31}, A_{32}, B_1, B_{22}, B_{23}$  jusqu'au profil très incomplet  $A_{11}, A_{12}, A_3, B_2$  ;

— la constellation 24 est surtout formée de profils du type  $A_1, A_3, B_1, B_2$  (le profil DIH 8 à distance  $B_1-B_{21} = 5$  en fait partie comme précédemment) et est en fait très proche de la constellation 22 ;

— le profil isolé CMB 23 a un type très particulier :  $A_3, B_{21}, B_{22}$ .

La comparaison entre les deux calculs de constellations, selon que l'on utilise l'ensemble des données de toutes les distances disponibles ( $l = 28$  pour 8 horizons appariés) ou que l'on se limite aux données des distances entre les 5 horizons principaux appariés ( $l = 10$ ), montre la supériorité du premier calcul ;

— les constellations 12 et 21 comprennent de nombreux profils communs mais pas tous, et il en est de même pour les constellations 11 et 23 ;

— par contre, les constellations 13 et 22+24 sont exactement identiques.

Il semble donc que l'on a intérêt à utiliser le maximum d'informations disponibles, quitte à compléter par des valeurs de distances identiques dans tous les cas où l'on peut remplacer des horizons manquants sans faire d'erreur de logique : possibilité de dédoubler un horizon  $A_1, A_3$  et  $B_2$  en  $A_{11}+A_{12}, A_{31}+A_{32}$  et  $B_{21}+B_{22}$ , mais impossibilité de remplacer un horizon  $A_1, A_3$  ou  $B_1$  manquants, que l'on code en conséquence comme « manquant ».

TABLEAU 6

			Étape 1	Étape 2	Étape 3	
14	DIH 64	A1 A3 B1 B21 B22				Constellation 21
15	DIH 68	A1 A3 B1 B21 B22				
18	CMB 10	A1 A3 B2				
4	SMB 12	A1 A3 B1 B21 B22 B23				
6	DIH 43	A1 A3 B1 B21 B22 B23				
17	DIH 77	A1 A3 B1 B21 B22 B23				
19	CMB 15	A1 A31 A32 B21 B22 B23				
20	CMB 16	A1 A3 B21 B22 B23				
8	DIH 49	A1 A3 B21 B22				
12	DIH 21	A1 A3 B1 B21 B22				
3	DIH 7	A1 A3 B1 B21 B22				
13	DIH 60	A1 A3 B1 B21 B22				
9	DIH 50	A1 A3 B1 B2				
23	CMB 25	A1 B1 B2				Constellation 23
11	DIH 53	A1 A3 B1 B21 B22 B23				
24	CMB 28	A3 B1 B21 B22 B3 C				
1	SMB 1	A1 A31 A32 B1 B21 B22 B23				
2	SMB 6	A1 A31 A32 B1 B21 B22 B23				Constellation 24
21	CMB 22	A11 A12 A3 B2				
7	DIH 47	A1 A3 B1 B2				
10	DIH 52	A1 A3 B1 B2				
16	DIH 61	A1 A3 B1 B2				
5	DIH 8	A1 A3 B1 B21 B22				
22	CMB 23	A3 B21 B22				

### 2.2.3. CONCLUSION PARTIELLE

Une première conclusion peut être tirée des calculs exécutés aux § 2.2.1 et 2.2.2. et basés sur les distances intra-profil et leur utilisation comme données caractérisant le profil. Il ne faut pas oublier, en effet, que cette méthode ne consiste pas à comparer des profils d'après leurs données morphologiques, mais compare la différenciation des profils. Une telle possibilité est intéressante quand on veut étudier un fichier important de profils (autour de la centaine), dont l'utilisateur ne peut avoir une connaissance approfondie. La démarche proposée est la suivante :

- calculs des distances intra-profil (§ 2.2.1.) de tous les profils à étudier ;
- examen des résultats de chaque profil : existence d'horizons à distance faible, succession anormale d'horizons, mauvaise description ou caractérisation ;
- détermination des types d'horizons présents et d'une succession-type avec le maximum d'horizons ;
- complètement ou retranchement d'horizons dans la limite des connaissances en pédologie ;
- utilisation de la méthode  $d_d$  en utilisant le maximum possible d'horizons présents, pour regrouper les profils à différenciation d'horizons semblable ;
- utilisation des groupes de profils ainsi constitués pour les calculs suivants.

## 2.3. Distances $d$ entre profils d'après les données morphologiques de tous les horizons

### 2.3.1. DONNÉES ET VARIABLES UTILISÉES

Les données utilisées dans les calculs de ce paragraphe sont les données descriptives des horizons. Une indication sur le nombre de variables réellement utilisées est disponible dans le tableau 4, ce nombre varie de 25 à 50 par profil. (Ne pas oublier que des variables, qui seraient constantes à l'intérieur d'un profil, n'auraient pas été utilisées dans le calcul du § 2.2.1. — distances intra-profil — mais pourraient l'être ici car, pour être abandonnée, une variable relevée sur un type d'horizon, doit prendre la même valeur dans tous les profils du fichier : cette remarque ne doit pas toutefois augmenter fortement le nombre de variables présentes).

Cependant, ces variables connues par profil ne représentent pas l'ensemble des variables des 24 profils : leur décompte exact n'a pas été fait mais il doit être inférieur à la centaine.

Il faut remarquer aussi que seules peuvent être utilisées pour des comparaisons entre profils (calcul de distances) des variables ordonnées, quantitatives ou binaires, qui ont une signification réelle dans cette comparaison ; les variables qualitatives (nominal variables) ne sont pas utilisées parce qu'on ne peut effectuer avec elles aucun classement, ni aucune mesure et que le fait qu'elles soient présentes ou absentes n'est pas interprétable sauf si on a pris soin de transformer une telle variable à  $k$  données en  $k$  variables binaires.

La nécessité pratique d'avoir  $n \times v$  ou  $n_1 \times v_1 + n_2 \times v_2 \leq 400$  (§ 1.3.) oblige à définir de façon stricte le nombre d'horizons et de variables à utiliser : les variables présentes peuvent en particulier être sélectionnées à partir des listings des calculs précédents.

Les deux types de processus utilisés ont donc présenté des modalités différentes et ont nécessité des manipulations plus ou moins importantes pour homogénéiser les profils et permettre des comparaisons.

### 2.3.2. ETUDE DES 11 PROFILS DE LA CONSTELLATION 12 OBTENUE À PARTIR DES $d_d$ (§ 2.2.2.)

#### 2.3.2.1. Données et variables utilisées

La constellation 12 comprend 11 profils à différenciation d'horizons à peu près homogène : il a donc paru intéressant de savoir si ce groupe de profils pouvait de nouveau se diviser selon des critères uniquement morphologiques.

Les données et variables présentent donc les caractéristiques suivantes :

- les profils ont des successions d'horizons voisines, ce qui diminue l'importance des manipulations d'horizons ;
- la présence fréquente de l'horizon  $B_{23}$  permet d'utiliser 6 horizons ;
- les variables ont été sélectionnées à partir des listings des calculs de distances intra-profil (§ 2.2.1.) et ont été ainsi partagées :

horizons $A_1$ et $A_3$	: $54 \times 2 = 108$
horizons $B_1, B_{21}, B_{22}, B_{23}$	: $71 \times 4 = 284$
Total	392

On a procédé aux manipulations d'horizons suivantes :

- suppression de deux horizons  $A_{32}$  et passage d'un horizon  $A_{32}$  en horizon  $B_1$ , qui manquait ;

- complètement de deux horizons  $B_{23}$  par l'horizon  $B_{22}$  ;
- complètement, dans un profil, des horizons  $B_{22}$  et  $B_{23}$  par l'horizon  $B_2$  transformé en horizon  $B_{21}$  ;
- seuls deux horizons  $B_1$  ont finalement manqué (code -1).

### 2.3.2.2. Résultats obtenus

Ce calcul sur 11 profils à différenciation et nombre d'horizons à peu près homogènes n'a permis qu'un faible regroupement à partir du calcul de leurs distances  $d$  : à la première étape on n'obtient que deux constellations de respectivement 3 et 2 profils, tous les autres profils restant isolés ; il n'y a pas de constellations à la deuxième étape, tous les profils, non suffisamment différenciés ou pas assez nombreux, étant regroupés.

Une des deux constellations décelées est formée des profils SMB 6, SMB 12 et CMB 16 (constellation 31) :

- les deux premiers profils sont des sols ferrallitiques appauvris et profonds observés sur grès et ce rapprochement avait déjà été évidemment fait sur le terrain ;
- le profil CMB 16 observé sous savane vient s'adjoindre aux deux précédents.

Ce dernier rapprochement n'ayant été fait ni sur le terrain, ni à l'examen rapide des profils, on peut se demander quelles sont les raisons de ce regroupement :

- les trois profils sont profonds et ont un horizon  $B_{23}$  à éléments grossiers d'oxydes et hydroxydes indurés en nodules entre 3 et 4,5 cm ;
- CMB 16 est un peu plus jaune (7,5 YR) que les deux autres profils (5 YR) ;
- aucune tache dans les trois profils, ni aucun élément grossier avant l'horizon  $B_{23}$  ;
- CMB 16 a une texture un peu plus argileuse que les deux autres profils (25 à 40 % d'argile, contre 15 à 35 %), mais avec une différenciation granulométrique voisine ;
- les structures des différents horizons sont assez voisines.

Sans l'aide du calcul, le rapprochement du profil CMB 16 des profils SMB 6 et SMB 12 n'était pas évident, à moins d'un examen approfondi de la description des profils.

### 2.3.3. ETUDE DE 20 APPUMITES (HORIZONS $A_1 + A_3$ ) (PAR LA MÉTHODE $d$ )

#### 2.3.3.1. Données et variables utilisées

Il a paru intéressant, avant d'étudier le profil complet sur l'ensemble du fichier, de l'étudier uniquement sur les appumites (horizons humifères) : le groupe « horizon  $A_1 +$  horizon  $A_3$  » est présent dans 20 profils.

En raison de la limitation à deux horizons, le problème du nombre de variables à entrer dans le calcul ne se pose plus : celles-ci ont été choisies sur les listings des calculs précédents en fonction de leur utilisation réelle. Les variables furent ainsi réparties et utilisées :

- 50 variables sur l'horizon  $A_1$  ;
- 60 variables sur l'horizon  $A_3$  ;
- seulement 9 variables sur 110 furent finalement inutilisées.

Aucune manipulation de données n'a été indispensable, puisqu'on a choisi les profils à appumite complet (horizons  $A_1$  et  $A_3$  présents).

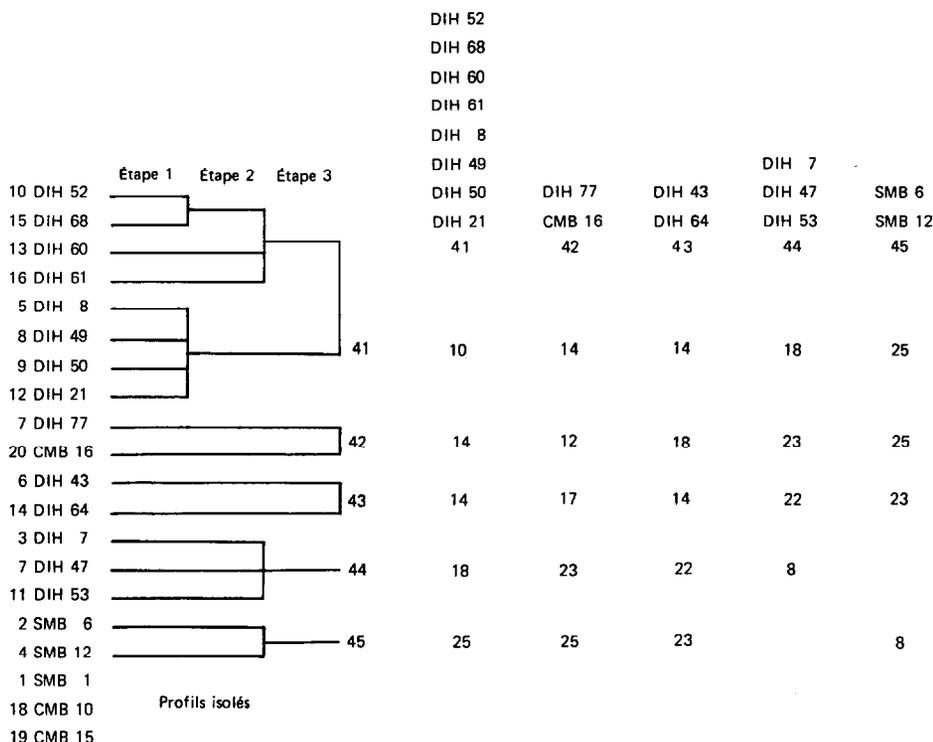
#### 2.3.3.2. Résultats obtenus

Le tableau 7 présente le dendrogramme des 3 étapes du calcul, ainsi que la matrice des distances moyennes (multipliées par 100) inter et intra-constellations : 5 constellations sont ainsi décelées et l'examen rapide des profils concernés permet de faire certaines constatations sur les caractéristiques qui les différencient.

La constellation 41 ne comprend que des appumites de sols de savanes peu arbustives à dense végétation graminéenne (8 profils) :

- la teinte est supérieure à 8,75 YR et la saturation passe de 2 à 3 ou 3 à 4 entre les deux horizons ;
- la texture est argileuse à très argileuse : 40 à 70 % d'argile ;
- la structure grenue de l'horizon  $A_1$  (caractéristique d'un horizon humifère de savane argileux et correctement pourvu en matière organique) passe à une structure polyédrique subanguleuse ou polyédrique dans l'horizon  $A_3$  ;
- l'ensemble des deux horizons est poreux à peu poreux ;
- le chevelu est très fréquent dans l'horizon  $A_1$ , mais il n'y a plus que quelques racines dans l'horizon  $A_3$ .

TABLEAU 7



La constellation 42 (CMB 16 et DIH 77) est formée aussi d'appumite de sols de savanes. Par rapport à la constellation précédente, elle se distingue par les caractéristiques suivantes :

- la texture est nettement moins argileuse : 25 à 40 % d'argile ;

- la structure grenue très nette et grossière à sous-structure plus fine est constante dans les deux horizons.

La constellation 43 semble assez voisine de 41 :

- la texture est argileuse à très argileuse ;
- la structure est moins bonne, plus polyédrique que grenue ;
- le chevelu est présent dans l'horizon A<sub>1</sub>, ainsi que de nombreuses racines dans l'horizon A<sub>3</sub>.

La constellation 44 est assez nettement différenciée des précédentes (3 profils) :

- la teinte (10 YR) et le chroma (passant de 2 à 3 de A<sub>1</sub> à A<sub>3</sub>) sont très homogènes ;
- la texture varie de 30 à 60 % d'argile ;
- la structure est mauvaise dans les deux horizons, à éclats émoussés puis anguleux ;

— l'appumite est peu poreux et se caractérise par une certaine dégradation de la structure, sans doute en liaison avec un faible appauvrissement.

La constellation 45 comprend les appumites sous forêts des deux sols ferrallitiques appauvris sur grès, (SMB 6 et SMB 12). Ce type d'appumite est nettement différent de ceux étudiés ci-dessus et se caractérise ainsi :

- texture sableuse et net appauvrissement en argile ;
- structure particulière ou massive, mais bonne porosité ;
- organisation racinaire différente de celle des sols sous savane.

En comparant les constellations 42 et 45 et la constellation 31 du calcul précédent, on peut conclure que les trois profils SMB 6, SMB 12 et CMB 16 sont des sols ferrallitiques appauvris et profonds, assez voisins au point de vue morphologique, mais que l'un d'entre eux, CMB 16, se différencie des deux autres par un appumite spécial, ce qui est normal compte-

tenu de la différence de végétation : cette différence n'est cependant pas suffisante pour que, dans l'étude de l'ensemble du profil, CMB 16 soit séparé de SMB 16 et SMB 12.

### 2.3.4. ETUDE DES 24 PROFILS DU FICHIER (PAR LA MÉTHODE *d*)

#### 2.3.4.1. Données et variables utilisées

Ce calcul a été en fait le premier effectué et n'a pas bénéficié de résultats antérieurs et, en particulier, on n'a pas utilisé la possibilité de choisir des variables différentes pour deux groupes d'horizons. Aussi, la limitation  $l = n \times v \leq 400$  et l'hétérogénéité des 24 profils quant à la succession et au nombre d'horizons présents ont conduit à un choix :

— comparaison des profils limitée aux 5 horizons :  $A_1, A_3, B_1, B_{21}, B_{22}$  ;

— limitation à 80 variables (400/5) par horizon.

Le choix des variables s'est fait sur le listing du calcul des distances intra-profil (§ 2.2.1.), en éliminant

certaines variables peu représentées ou ne paraissant pas avoir un grand intérêt (comme les traces d'activité humaine ou le mode de liaison entre racines et agrégats).

L'homogénéisation des profils a par conséquent entraîné les suppressions et ajouts suivants :

— suppression d'un horizon  $A_{12}$  et de trois horizons  $A_{31}$  ;

— dédoublement de sept horizons  $B_2$  en horizons  $B_{21}$  et  $B_{22}$  ;

— remplacement de trois horizons  $A_1$ , d'un horizon  $A_3$  et de six horizons  $B_1$  absents par des codes -1 pour manquants ;

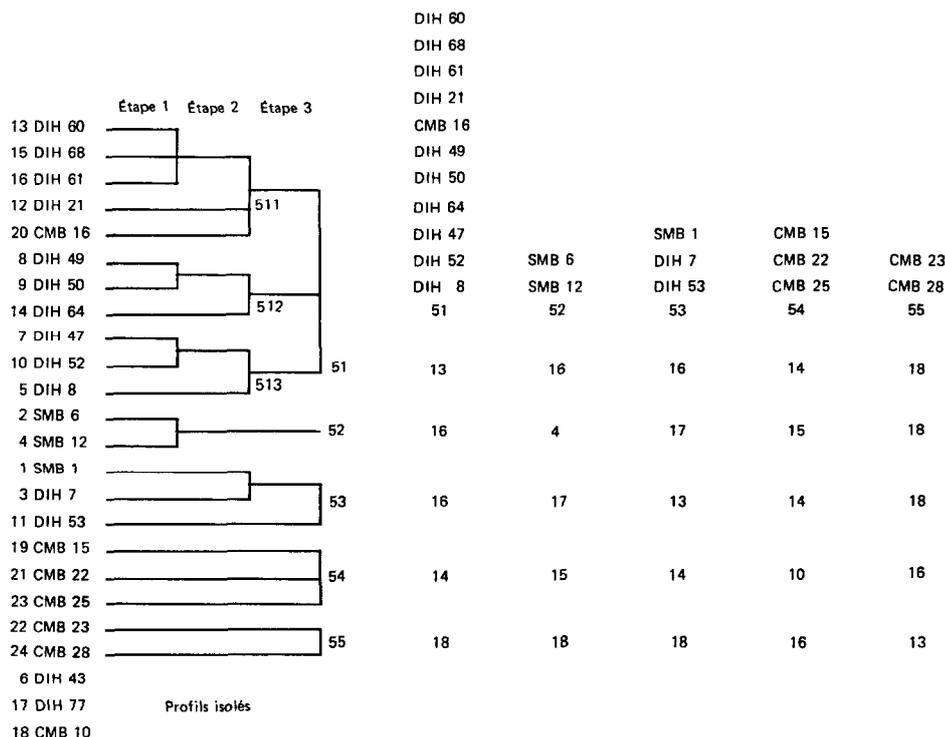
— abandon de 7 horizons  $B_{23}$ .

#### 2.3.4.2. Résultats obtenus

Le tableau 8 présente le dendrogramme des 3 étapes du calcul, ainsi que la matrice des distances moyennes (multipliées par 100) inter et intra-constellations.

Ne pas oublier que les comparaisons entre tous ces profils sont limitées à l'horizon  $B_{22}$  (5 horizons)

TABLEAU 8



et que les profils sont étudiés jusqu'à des profondeurs très variables : 80 à 460 cm. (Bien que la profondeur et l'épaisseur des horizons soient utilisées comme variables, elle ne peuvent jouer qu'un faible rôle devant les 50 à 60 autres variables utiles : il peut donc en résulter que des profils d'épaisseurs très variables fassent partie d'une même constellation).

La constellation 51 comprend 11 profils dont il est difficile de discerner a priori les principaux caractères communs. Aussi est-il plus intéressant d'étudier les 3 sous-constellations de la deuxième étape.

Constellation 511 (profils DIH 60, DIH 68, DIH 21, CMB 16) : il s'agit de sols ferrallitiques jaunes (7,5 YR), d'épaisseur variable (100 à 340 cm) avec ou sans taches, exceptionnellement à éléments grossiers en  $B_{22}$  ; la structure grenue dans les premiers horizons passe systématiquement à une surstructure polyédrique se résolvant en très fins polyèdres ; la teneur en argile des horizons B varie de 40 à 70 % avec de faibles variations depuis la surface.

Constellation 512 (profils DIH 49, DIH 50, DIH 64) : les trois profils ferrallitiques sont également jaunes (7,5 YR), mais peu épais ( $B_{22} < 120$  cm), avec ou sans taches et éléments grossiers, très argileux (65 à 75 %) dans les horizons B avec faibles variations depuis la surface, à structure polyédrique fine sans surstructure en profondeur.

Constellation 513 (DIH 47, DIH 52, DIH 8) : ces sols ferrallitiques sont très jaunes (10 YR) et peu épais ( $B_{22}$ , 150 cm), à taches plus ou moins abondantes et variées et sans éléments grossiers ; la teneur en argile est élevée en profondeur (60 à 75 %) et la structure grenue de surface devient polyédrique fine avec ou sans surstructure ; ces profils ont été classés sur le terrain dans un sous-groupe hydromorphe.

La constellation 52 (SMB 6 et SMB 12) regroupe, dès la première étape et sans autre adjonction ultérieure, deux profils de sols ferrallitiques appauvris et profonds sur grès et sous forêt : il s'agit des deux mêmes profils déjà rapprochés sur le terrain et qui font partie de la constellation 31 (§ 2.3.2.2.). Il faut remarquer que le profil CMB 16 n'est plus joint à ceux-ci, mais fait partie de la constellation 51 et de la sous-constellation 511 : on peut donc conclure que les trois profils SMB 6, SMB 12 et CMB 16 se ressemblent plus par l'ensemble de leur profil que par leurs appumites ou leurs profils limités à 5 horizons.

La constellation 53 (SMB 1, DIH 7, puis DIH 53 en troisième étape) est formée de sols appauvris et/ou lessivés, avec taches plus ou moins abondantes, sans éléments grossiers jusqu'à  $B_{22}$ , les teneurs en argile sont variables, mais toujours nettement différenciées

selon la profondeur ; la structure est massive dans les horizons A puis polyédrique dans l'horizon B.

Les trois profils (CMB 15, CMB 22 et CMB 25) de la constellation 54 ne se regroupent qu'à la troisième étape, aussi leurs caractères communs ne sont pas très précis : il s'agit de sols jaunes (7,5 YR ou plus jaune), moyennement épais (100 à 190 cm pour  $B_{22}$ ), à taches et éléments grossiers variables ; leur texture assez homogène (20 à 25 % d'argile) est caractérisée par une forte teneur en limon fin et une faible différenciation granulométrique ; la structure est variable.

Les deux profils (CMB 23 et CMB 28) de la dernière constellation 55 ne se regroupent également qu'à la troisième étape et paraissent assez proches des précédents ; ils sont cependant un peu plus rouge (7,5 à 5 YR), sans taches et à éléments grossiers siliceux ; la texture plus argileuse (35 à 50 % d'argile) et peu différenciée est également caractérisée par la présence de limon fin.

#### 2.4. Corrélations

Le test de corrélation a été effectué sur l'effectif maximal de 100 horizons pour 43 variables choisies parmi les plus intéressantes. L'ensemble des résultats des 903 tests ( $C_{43}^2$ ) de corrélation est partitionné en quatre sous-ensembles :

— corrélations incalculables	: 92
— absences de corrélation au risque 5 %	: 551
— corrélations positives	: 135
— corrélations négatives	: 125

A titre d'exemple, on donne ci-après la liste de toutes les corrélations au risque 1 % avec la variable « structure polyédrique », variable du type absence-présence, donc absente pour tout autre type de structure.

##### *Corrélations positives*

— Humidité : la structure polyédrique est présente en cas d'humidité élevée.

— Clarté et saturation état humide : la structure polyédrique est associée à des valeurs élevées de ces deux caractéristiques (couleurs vives).

— Abondance des taches : présence concomitante de structure polyédrique et de taches.

— Clarté et saturation des taches : les taches associées à la structure polyédrique sont de couleur vive.

— Nature générale de la structure : corrélation évidente : la structure polyédrique est obligatoirement une structure fragmentaire.

— Netteté de la structure : la structure polyédrique est en même temps très nette.

— Profondeur minimale : la présence de structure polyédrique augmente avec la profondeur.

— Teneur en argile ; la structure polyédrique est associée à des teneurs élevées en argile.

#### Corrélations négatives

— Taille de la structure : la structure polyédrique est en même temps une structure fine.

— Chevelu : pas de structure polyédrique associée à la présence de chevelu.

— Activité : faible activité en liaison avec la structure polyédrique.

— Teneur en sable : la structure polyédrique est associée à de faibles teneurs en sable.

On en conclut que la structure polyédrique est présente dans les horizons profonds normalement humides, à forte teneur en argile et faible teneur en sable, à couleur vive et en présence de taches de couleur vive et associée à l'absence de chevelu et à une faible activité : cette structure est nette et fine.

Sur les 260 (135+125) corrélations recensées, seules les 76 corrélations positives et les 80 corrélations négatives au risque 1 % sont interprétables. On trouve des corrélations obligatoires ou normales pour les sols étudiés :

#### Corrélations obligatoires

— Éléments ferro-manganésifères et Oxydes et hydroxydes.

— Graviers (et cailloux) et Teneur en éléments grossiers.

#### Corrélations normales

— Humidité et Clarté, Saturation, Profondeur : les sols ferrallitiques sont humides et de couleur vive en profondeur.

— Activité et Chevelu, Teneur en sable, Vides, Présence de matière organique : l'activité est liée à la présence de matière organique, de vides, de chevelu et de forte teneur en sable, toutes caractéristiques de l'horizon humifère où se concentre l'activité dans les sols ferrallitiques.

#### CONCLUSIONS

Cet article a été établi à l'issue du traitement sur ordinateur d'un fichier de profils décrits au Congo. Il contient l'exposé des méthodes statistiques choisies, le déroulement des procédures informatiques et, surtout, les résultats obtenus. Le nombre élevé de variables prises simultanément approche la capacité maximale autorisée par l'ordinateur auquel est raccordé la banque de données pédologiques. En revanche, le nombre de profils retenus pourrait monter à 100 dans un travail similaire. En outre, les programmes restent inchangés quand il y a substitution de variables morphologiques par des variables de laboratoire. La programmation de l'enchaînement des programmes donnés dans un autre article (Aubry, 1975) et la présence d'un terminal ont facilité le traitement. L'accès à des ordinateurs ayant une mémoire centrale plus grande permettrait d'augmenter facilement la capacité de ces programmes écrits en Fortran IV. Les regroupements obtenus par l'algorithme cité dans le texte constituent une information pédologique intéressante alors que les tests de corrélation n'apportent, comme souvent, que peu d'information.

#### BIBLIOGRAPHIE

- AUBRY (A.M.) *et al.*, 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 (Ivory Coast). *Bull. Ecol. Res. Comm.*, 17, 433-442.
- AUBRY (A.M.), 1975. — Programmes Fortran pour distances de rangs, constellations et corrélations. *Init. Doc. Tech. ORSTOM*, à paraître.
- CHATELIN (Y.), MARTIN (D.), 1972. — Recherche d'une terminologie typologique, applicable aux sols ferrallitiques. *Cah. ORSTOM, sér. Pédol.*, vol. X, n° 1, 25-43.
- Glossaire de Pédologie, 1965. — ORSTOM, *Init. Doc. Tec.*, n° 13 (H.S.), 82 p.
- KENDALL (M.C.), STUART (A.), 1966. — *The advanced theory of statistics*. Vol. 3, Design and analysis, and Time series, Griffin, London, 552 p.
- SPEARMAN (C.), 1904. — The proof and measurement of association between two things. *Am. J. Psych.*, 15 : 72-101.
- VAN DEN DRIESSCHE (R.), 1965. — La recherche des constellations de groupes à partir des distances généralisées  $D_2$  de Mahalanobis. *Biométrie-Praximétrie*, vol. VI, n° 1, 36-47.
- VAN DEN DRIESSCHE (R.) et GARCIA GOMEZ (A.), 1972. — Distances non paramétriques entre profils. *Rev. Ecol. Biol. Sol.*, IX, 4, 617-628.
- VAN DEN DRIESSCHE (R.) et GARCIA GOMEZ (A.), 1973. — Comparaison multivariable non paramétrique des profils sans nomenclature d'horizons. *Cah. ORSTOM, sér. Pédol.*, XI, 34, 257-267.
- VAN DEN DRIESSCHE (R.), 1974. — La Banque de Données Pédologiques de l'ORSTOM. *Cah. ORSTOM, sér. Pédol.*, XII, 1, 125-132.