

FIG. 1. — Deux Sapotacées à bois rouge et dense.
 A gauche, Mukulungu (*Autranella congolensis*).
 A droite, Moabi (*Baillonella toxisperma*). × 25.

SÉPARATION ANATOMIQUE DE BOIS DE SAPOTACÉES PAR DES MÉTHODES D'ANALYSE MULTIDIMENSIONNELLE

par Annie VITALIS-BRUN et Alain MARIAUX

Division d'Anatomie du C. T. F. T. ()*

(*) Cette étude a été présentée à la réunion conjointe I. A. W. A./I. U. F. R. O. Div. 5-1. invitée par la Société

Botanique Royale des Pays-Bas, à l'Institut Royal Tropical d'Amsterdam, 27-30 Août 1979.

SUMMARY

WOOD ANATOMICAL CLASSIFICATION OF SAPOTACEAE USING METHODS OF MULTIDIMENSIONAL ANALYSIS

Sapotaceae are a wood anatomically very homogeneous family and the identification of red heavy woods is sometimes very difficult, especially if the provenance is unknown.

It was thought that a statistical method could be more efficient for this purpose than a dichotomous key. A detailed analysis of 80 samples from 20 genera resulted in 26 varying characters. Two methods were tested: principal component analysis and factorial discriminant analysis. The first one gives the 2 or 3 better linear combinations of initial variables (the 26 characters) explaining a maximum of the total variation of the population, without consideration of the identity of the samples. The second one tries to compute the data in linear combinations of variables in such a way that a better clustering of each genus and a greater distance between them is obtained. The second method appears more efficient to identify a sample (that is to put it in a group) but the first method gives a natural clustering which may be of interest for taxonomic considerations. A point to be discussed is the use of secondary characters (as silica grains) sometimes very discriminant, together with structural and more basic characters.

RESUMEN

SEPARACION ANATOMICA DE MADERA DE SAPOTACEAS POR METODOS DE ANALISIS MULTIDIMENSIONALES

Las Sapotáceas constituyen una familia sumamente homogénea por lo que respecta a la anatomía de la madera y la identificación de las maderas rojas y densas resulta en algunos casos muy difícil, y fundamentalmente en caso de que se desconozca la procedencia. Se ha pensado que un método estadístico podría resultar más eficaz con este objeto que una clave dicotómica.

Se ha procedido a un análisis detallado de 80 muestras pertenecientes a 20 géneros con 26 características variables estudiadas. Se han sometido a prueba dos métodos: el análisis en cuanto a componentes principales y el análisis factorial discriminante. El primero proporciona las dos o tres mejores combinaciones lineales de las variables iniciales (las 26 características) que explica un máximo de la variación total de la población considerada y sin tener en cuenta para ello la identidad de las muestras. Por el segundo método se trata de procesar los datos en combinación lineal de las variables, con objeto de obtener la mejor agrupación de cada género y la mayor distancia entre los mismos. El segundo método parece ser más eficaz para la identificación de una muestra (es decir para situarla en un grupo), pero en cambio, el primer método proporciona una agrupación natural que puede tener un gran interés para las consideraciones taxonómicas. Uno de los puntos discutidos corresponde a la utilización de características secundarias (como los granos de sílice que en ciertos casos permiten una discriminación correcta) y ello, paralelamente a las características estructurales más fundamentales.

INTRODUCTION : LE PROBLÈME DES GROUPES COMPLEXES DE BOIS

Il existe des méthodes modernes d'analyse statistique des données très synthétiques sur des populations d'individus représentés chacun par de nombreux caractères. Elles sont maintenant connues dans les recherches sur le bois où elles révèlent, par exemple, des similitudes technologiques entre espèces sans liens taxonomiques.

Les méthodes que nous retiendrons ici sont l'analyse en composantes principales et l'analyse discriminante.

L'idée de départ est que leur application aux caractères anatomiques pourrait montrer, mieux que des clés dichotomiques arbitraires, si des groupements apparaissent en correspondance avec la classification bota-

nique. Un résultat favorable serait intéressant du point de vue systématique et conduirait, théoriquement tout au moins, à une identification plus logique et plus sûre dans les cas difficiles.

Comme matériel sur lequel tester ces méthodes, nous avons choisi un lot de Sapotacées. De nombreux bois rouges et denses de cette famille sont utilisés dans le monde, et leur identification précise est difficile en raison de la grande homogénéité de structure de la famille.

Le présent article condense le travail de thèse présenté par A. Vitalis-Brun pour l'obtention du titre de Docteur de l'Université (3).

1. — DESCRIPTION ANATOMIQUE GÉNÉRALE DES SAPOTACÉES A BOIS ROUGES ET DENSES

La description suivante leur est commune :

Vaisseaux généralement de forme ovulaire, en nombre moyen (10 à 20 par mm²), diamètre tangentiel moyen (80 à 200 μm), plus ou moins abondamment accolés radialement, jusqu'à plus de 4 et, en outre, avec une tendance à se disposer souvent en lignes

radiales ou obliques. Les ponctuations intervasculaires varient cependant de 4 à 11 microns. Les cloisons sont toujours à perforations uniques.

Rayons assez nombreux (7 à 19 par mm), petits (350 à 800 μm), et étroits (20 à 50 μm), avec une sériation le plus fréquemment de 2 et 3, rarement 1 ou 4. La

structure est toujours hétérogène avec 2 à 8 rangées de cellules carrées à chaque extrémité, et 5 à 25 rangées de cellules couchées au centre. Les ponctuations vaisseau-rayon sont toujours grosses, soit arrondies, soit allongées, mais d'une façon peu caractéristique. Jamais des cristaux n'ont été rencontrés, mais la silice en grain est courante dans certains genres.

Parenchyme en nombreuses lignes tangentielles à l'intérieur des accroissements. Elles sont plus ou moins fines, rectilignes, continues. Les moins continues donnent une image plus dispersée. Les files non étagées sont composées de 4 à 8 éléments dans le même échantillon. Ces éléments peuvent contenir, soit des

cristaux d'oxalate de calcium, soit des grains de silice.

Les fibres, dans le groupe étudié, varient du simple au double pour la longueur moyenne : 1.200 à 2.300 μm , et beaucoup moins pour la largeur ; de 22 à 27 μm .

Cette description montre bien que, pour différencier les bois, il faudra utiliser un faisceau de caractères à variations faibles, mais soigneusement analysées. Chaque caractère aura un faible pouvoir de différenciation des genres, mais l'accumulation des légères tendances de chacun donnera une meilleure discrimination si la méthode choisie correspond à nos espérances.

2. — MATÉRIEL ÉTUDIÉ

TABLEAU 1
MATÉRIEL ÉTUDIÉ

| Groupes | Genres | Espèces | Nombre d'individus |
|--------------------------------|---|---|--|
| MIMUSOPOIDÉES (Manilkarées) | Faucherea | F. ambrensis Aubr. F. laciniata Cap. F. longepedicellata Aubr. F. parvifolia H. Lec. F. thouvenotii H. Lec. | 1 1 1 1 1 |
| | Labourdonnaisia Letestua Manilkara | L. calophylloides Bojer L. durissima H. Lec. M. amazonica Standl. M. bidentata A. Chev. M. foulloyana Aubr. & Pell. M. huberi Standl. M. lacera Dub. M. maboqueensis Aubr. M. multinervis Dub. M. obovata Hemsh. M. pellegriniana Sill. et Tiss. A. congolensis A. Chev. B. toxisperma Pierre M. commersonii Engl. M. elengi L. var. parvifolia Lam M. maxima Vaughan M. sambiranensis Aubr. T. africana Pierre T. heckelii A. Chev. V. paradoxa Gaertn. | 2 5 1 2 1 1 1 1 1 1 1 5 5 1 1 1 1 3 2 3 |
| (Mimusopées) | Austranella Baillonella Mimusops Tieghemella Vitellaria (Butyrospermum) | | |
| MADHUCOIDÉES | Madhuca | M. bejaudii Aubr. M. betis Merr. M. indica Gmel M. pasquieri H. J. Lam M. sp. P. leerii Kurz | 1 1 1 1 1 1 |
| SIDEROXYLOIDÉES (Kantouées) | Payena | | |
| | Kantou Neolemonniera Gluema Lecomtedoxa | K. guereensis Aubr. & Pell. N. clitandriifolia Heine G. ivorensis Aubr. & Pell. L. heitzeana Aubr. L. klaineana Pierre L. nogo Aubr. A. sapota L. C. antongiliensis Aubr. C. microlobum Aubr. C. microphyllum Aubr. C. rubrococostatum Aubr. C. sakalava Aubr. P. benai Aubr. et Pell. P. amygdalina Baehni P. guianensis Aubl. P. melanopoda Eyma | 5 1 4 1 4 1 4 1 1 1 1 1 1 1 1 1 1 1 |
| (Lecomtedoxées) | | | |
| (Poutériées) | Achras Capurodendron Podoluma Pouteria | | |
| OMPHALOCARPOIDÉES | Tridesmostemon | T. congoense Aubr. T. omphalocarpoides Engl. | 2 3 |

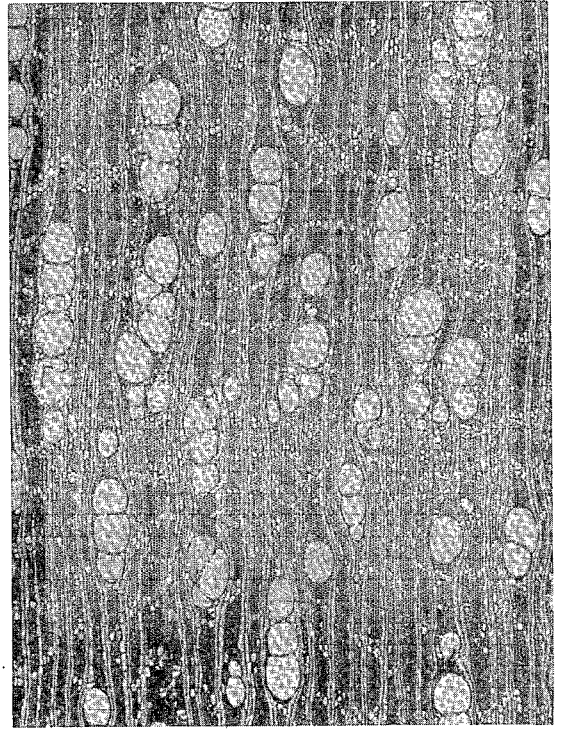
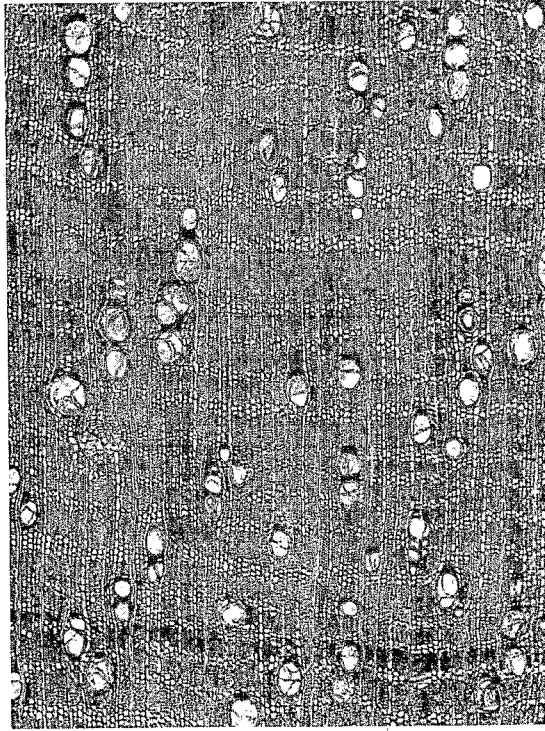


FIG. 2. — Différence dans la taille et dans le groupement des vaisseaux, dans la largeur et la continuité des lignes de parenchyme : à gauche, *Gluema ivorensis*, à droite, *Lecomtedoxa heitziana* (25 ×).

3. — CARACTÈRES ANATOMIQUES RETENUS POUR L'ANALYSE

Pour chaque échantillon, nous avons fait une étude qui diffère de la description classique par une analyse plus fine de certains caractères, et par leur découpage en classes (de valeur ou de disposition), adaptée au calcul statistique.

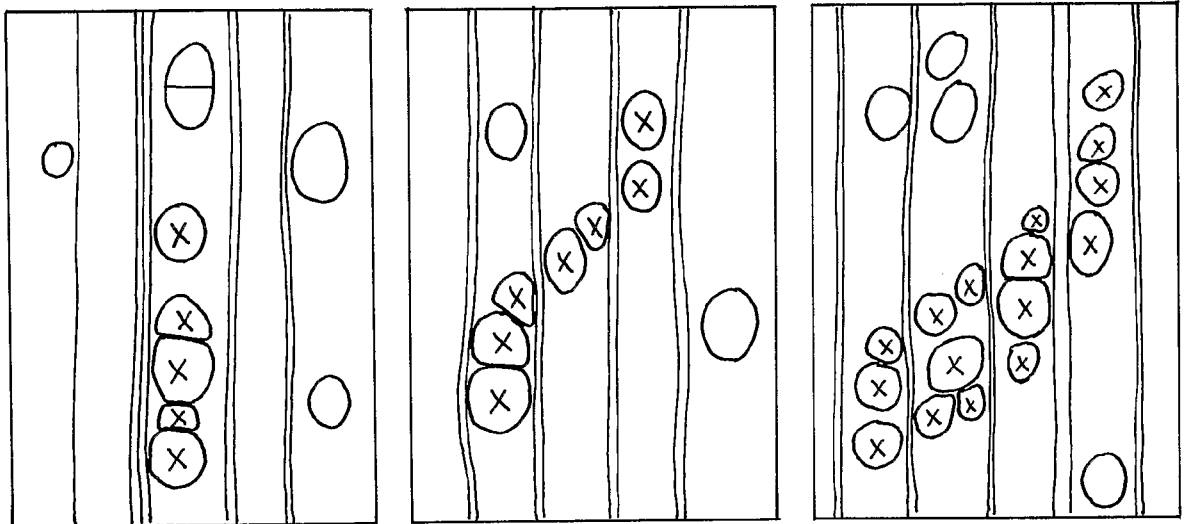
Ce n'est pas toujours la valeur moyenne d'un caractère qui a été exprimée, comme il est habituel, mais un développement « maximum courant », notion qui a

l'inconvénient d'être un peu subjective, mais l'avantage de mieux caractériser les particularités de chaque espèce.

3.1. Vaisseaux.

Le nombre de vaisseaux par mm^2 , leur groupement (accolement maximum courant), leur diamètre tangentiel moyen et le diamètre de l'aréole des ponctuations

FIG. 3. — Exemple d'analyse de la disposition des vaisseaux : à gauche, file radiale de 5 vaisseaux isolés ou accolés. Au centre, file oblique 1 — sériee sur 3 interrayons. A droite, file oblique 3 sériees sur 4 interrayons.



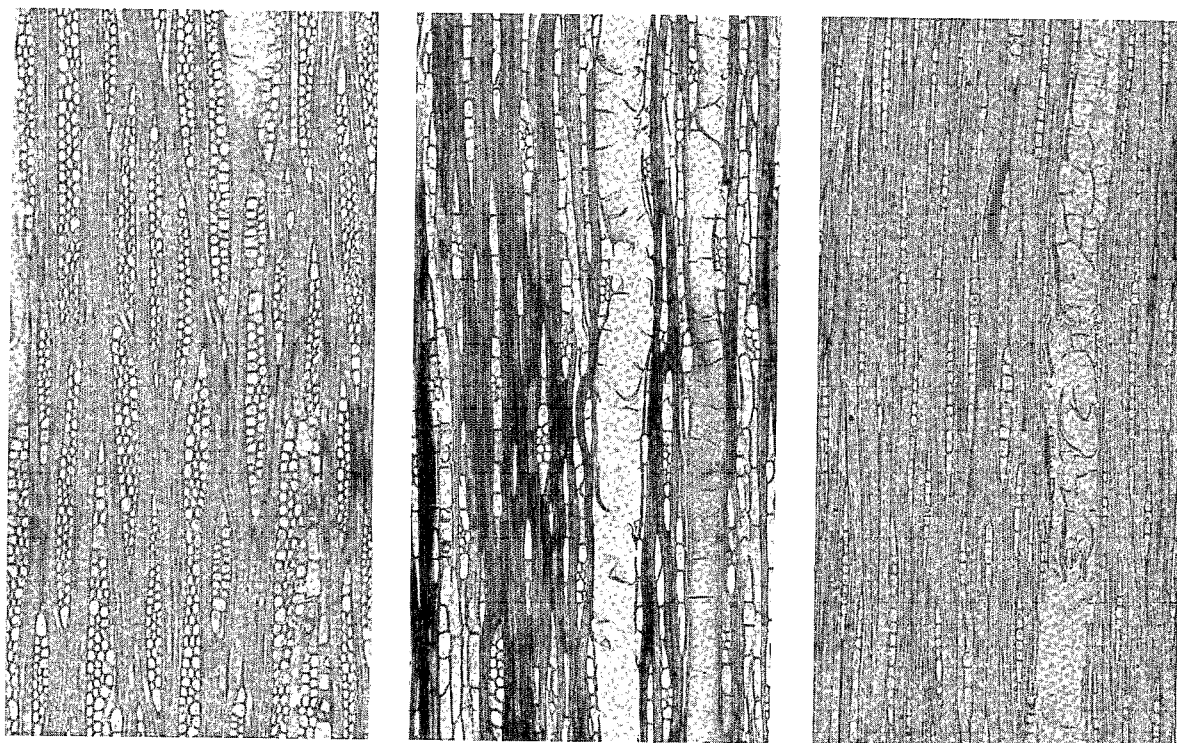


FIG. 4. — Divers types de rayons vus en section tangentielle (55 ×). De gauche à droite : *Payena leerii*, *Faucherea longepedicellata*, *Lecomtedoxa heitzeana*.

intervasculaires, font partie des observations et mesures d'une étude classique.

Par contre, la disposition a été classée beaucoup plus précisément pour dépasser une qualification générale qui s'appliquerait à la plupart des Sapotacées, telle que : Vaisseaux disposés plus ou moins en files obliques. Nous avons déterminé 6 classes de disposition des vaisseaux (toujours observées à la même échelle macroscopique X 10) en files radiales, ou obliques, ou disséminés. Puis nous avons essayé de préciser quantitativement, quand il y a « file » (fig. 3), sa dimension et sa constitution, en comptant séparément :

- le nombre de vaisseaux rencontrés successivement dans le sens radial (sans intervalle supérieur à la taille d'un vaisseau),

- la « sériation », nombre de vaisseaux côte à côte tangentiellement,

- enfin, le nombre d'inter-rayons occupés par la file, d'autant plus grand que celle-ci est plus oblique et plus longue (un inter-rayon est l'intervalle, vu en section transversale, entre deux rayons).

3.2. Rayons.

Le nombre par mm, la hauteur et la largeur ont été examinés comme pour une étude classique.

Par contre, c'est le maximum courant qui est exprimé pour la sériation des rayons (nombre de cellules dans la plus grande largeur du rayon) ainsi que pour la hauteur comptée en nombre de rangées de cellules, d'une part dans la partie à cellules couchées, d'autre part dans un des prolongements à cellules dressées ou carrées (fig. 4).

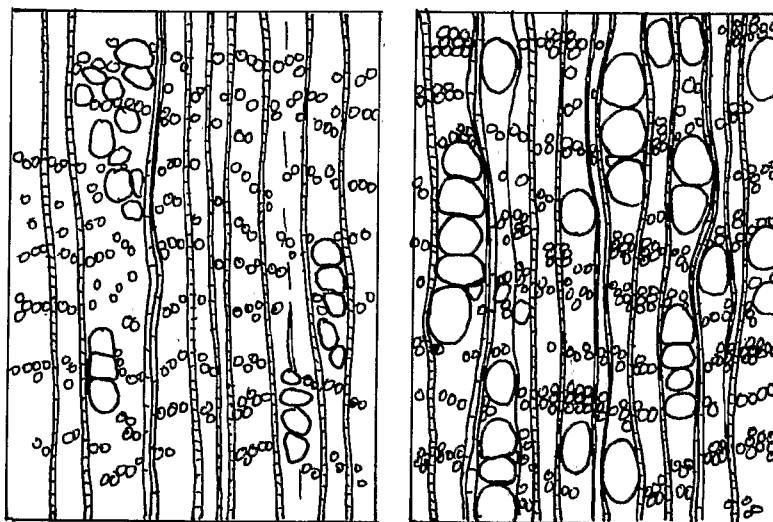


FIG. 5. — Lignes de parenchyme. A gauche, 1-sériées couvrant 3 interrayons. A droite, 3-sériées couvrant jusqu'à 6 interrayons.

Enfin, on a noté la présence de grains de silice ; jamais celle de cristaux, et quelquefois, la présence de contenus d'aspect huileux assez typiques, observables sur bois massif, sur coupe ou sur dissociation avant traitement par un solvant organique.

3.3. Parenchyme.

Dans le bois des Sapotacées, le parenchyme apparaît macroscopiquement en lignes plus ou moins continues, plus ou moins fines et plus ou moins espacées. Une légère sinuosité éventuelle des lignes est difficile à exprimer objectivement. Nous nous y sommes efforcés.

La finesse des lignes est donnée par la sériation (fig. 5).

Leur continuité, comptée en inter-rayons, doit être rigoureuse, et elle est considérée comme interrompue si la traversée d'un rayon provoque un décrochement d'au moins la hauteur d'une cellule de parenchyme, ou encore si une ligne s'interrompt, et que son cheminement reprend l'espace de 2 ou 3 cellules plus loin.

4. — APPLICATION DE L'ANALYSE EN COMPOSANTES PRINCIPALES.

4.1. Rappel sur les méthodes d'analyse multidimensionnelle des données.

Les principes des méthodes d'analyses factorielles : analyse en composantes principales et analyse discriminante, sont largement décrits dans des publications spécialisées (1) (2).

L'analyse factorielle considère chaque individu comme représenté par un point dans un espace à n dimensions, ses n coordonnées étant les facteurs qui le caractérisent ; en ce qui nous concerne, ce seront 26 caractéristiques anatomiques. L'ensemble des individus forme un nuage de points.

Le principe de l'analyse en composantes principales (A. C. P.) est de visualiser le nuage des individus par projection dans un espace à 2 dimensions, c'est-à-dire un plan. Celui-ci est recherché par le calcul, de façon à représenter au mieux le nuage (en minimisant la somme des carrés des distances entre les individus et leur projection sur le plan).

Le plan est défini par 2 axes de coordonnées appelées composantes principales, qui représentent elles-mêmes plus ou moins bien un faisceau de caractères. On calcule successivement plusieurs plans principaux sur lesquels les projections du nuage apportent une quantité d'informations décroissantes.

L'A. C. P. est une méthode visant à voir les proximités entre individus et les corrélations entre caractères ; elle ignore le fait que les individus appartiennent à des groupes différents. Les groupements de points qui apparaîtront éventuellement seront donc « naturels » en fonction des facteurs utilisés.

Les résultats fournis par l'A. C. P. sont très riches et donnent un grand nombre d'informations : d'abord par les calculs préliminaires, puis par les nuages de points.

3.4. Fibres.

Les mesures étant effectuées sur matériel dissocié, la largeur est mesurée au ventre de la fibre, dans la partie la plus large, ainsi que l'épaisseur des deux parois cumulées. Le « coefficient de souplesse » exprime le diamètre de la cavité par rapport à celui de la fibre : $\frac{100 C}{1}$.

3.5. Synthèse de la description : codage des caractères.

Pour permettre les analyses statistiques ultérieures, les caractères anatomiques ont été codés.

Les caractères *quantitatifs* sont notés par leur valeur propre, et l'étendue de leurs valeurs est découpée en classes par le programme de calcul.

Les caractères *qualitatifs*, plus délicats à classer, ont été codés de la façon la plus logique possible, ce qui entraîne un nombre de classes assez inégal, tel que : disposition des vaisseaux : 6 classes ; silice dans les rayons : 2 classes (absente ou présente).

On obtient ainsi 26 caractères récapitulés dans le tableau 2.

4.2. Les calculs préliminaires en A. C. P.

Le programme de l'analyse permet, en préliminaire, d'obtenir des informations classiques d'une grande importance.

L'*histogramme de chaque variable* permet de juger de la qualité de sa répartition, et de détecter les valeurs exceptionnelles, soit réelles, soit erronées.

La *matrice de corrélation des caractères initiaux* montre des relations soit naturelles, soit inattendues :

— pour certains caractères, la corrélation était logique ; par exemple : VS/VI = 0,63 entre la sériation d'une file de vaisseaux et le nombre d'inter-rayons qu'elle couvre ;

— pour d'autres, elle souligne des oppositions remarquables : ainsi, corrélation RS/PC = - 0,63 entre la silice dans les rayons et des cristaux dans le parenchyme qui sont rarement présents dans le même échantillon. Seuls les genres *Austranella*, *Gluema*, *Letestua*, *Mimusops* et *Manilkara* d'Afrique (2 individus sur 6) ont à la fois de la silice dans les rayons et des cristaux dans le parenchyme.

Certaines corrélations fortes peuvent se produire entre caractères tout à fait indépendants, mais assez spécifiques de quelques genres. Ainsi, la présence de globules huileux dans les rayons (RG) et la disposition des vaisseaux en plusieurs séries (VS). Le fait que RG ne se rencontre que chez *Capurodendron*, a entraîné de fortes corrélations de ce caractère avec ceux qui sont le plus spécifiques de ce genre, par exemple VS.

Le tableau des corrélations entre les variables initiales et les composantes principales calcule le poids de chacune des composantes dans la représentation de l'ensemble des variations. On y verrait que les trois

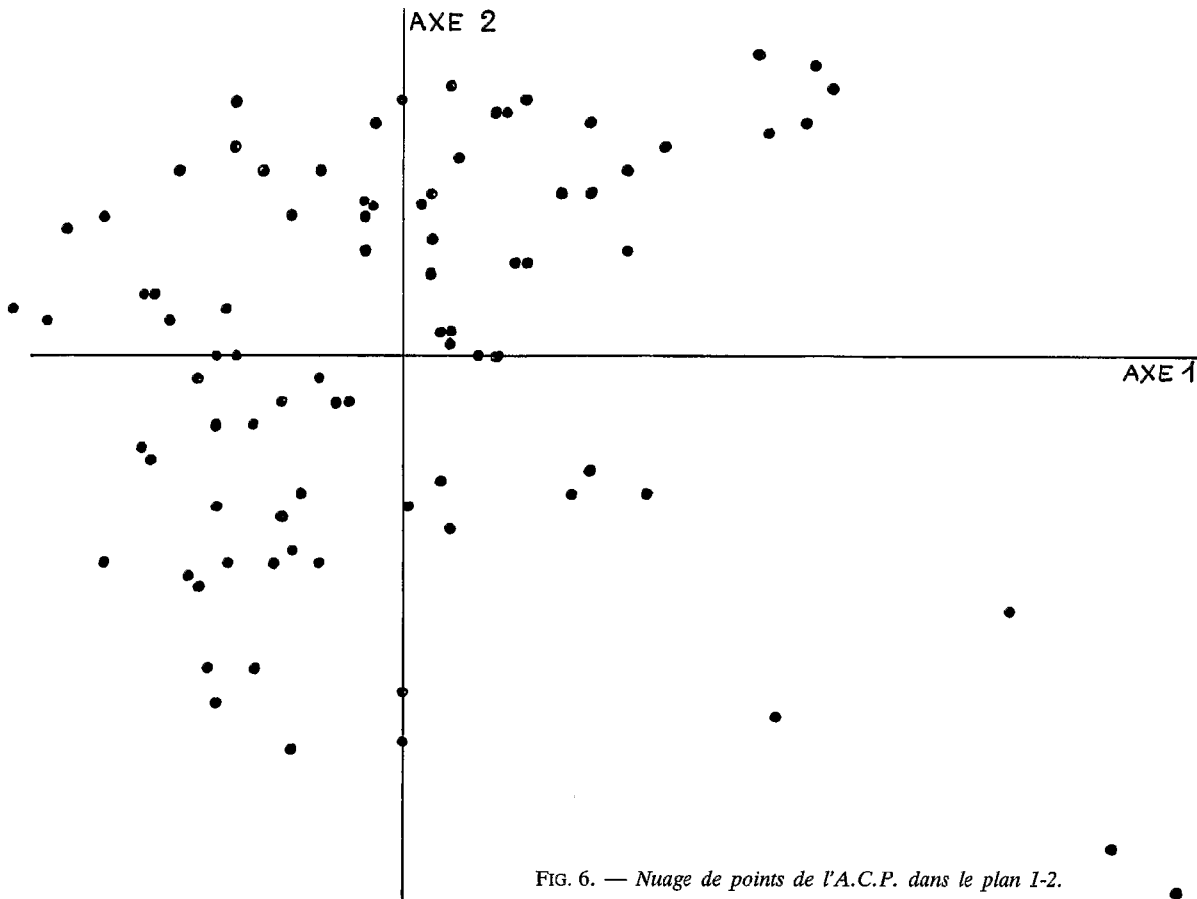


FIG. 6. — Nuage de points de l'A.C.P. dans le plan 1-2.

TABLEAU 2
Récapitulation et abréviations des caractères utilisés dans l'analyse en composantes principales (A. C. P.) et l'analyse factorielle discriminante (A. F. D.)

| | Caractères | A. C. P. | A. F. D. |
|--------------------------|-------------------------------|----------|----------|
| FIBRES | nombre par mm ² | VN | VN |
| | disposition | V | V |
| | nombre de pores dans 1 file | VF | |
| | sériation par file | VS | |
| | nombre d'inter-rayons | VI | |
| | groupement | VG | VG |
| | diamètre des vaisseaux | VD | VD |
| | diamètre des ponctuations 1/V | VP | VP |
| PARENCHYME RAYONS | nombre par mm | RN | RN |
| | hauteur | RH | RH |
| | largeur | RL | |
| | sériation | RSE | |
| | nombre de cellules couchées | RCO | RCO |
| | nombre de cellules carrées | RCA | |
| | silice | RS | RS |
| | globules huileux | RG | |
| VAISSEAUX | nombre d'inter-rayons | PI | PI |
| | sériation | PSE | PSE |
| | nombre de lignes par mm | PN | |
| | sinuosité ou raideur | PR | PR |
| | silice | PS | PS |
| | cristaux | PC | PC |
| | longueur | FLO | FLO |
| | largeur | FLA | FLA |
| coefficient de souplesse | CS | CS | |
| masse volumique | D | | |

premières composantes principales ensemble expliquent 44 % de la variation totale :

- 23 % par la 1^{re} ;
- 12 % par la 2^e ;
- 10 % par la 3^e composante.

4.3. Les nuages de points en A. C. P.

Ils sont l'expression la plus synthétique et la plus instructive de cette analyse. La figure 6 montre le nuage brut projeté dans le plan 1-2 (la machine inscrit les numéros des points, mais ils ne sont pas reproduits ici pour alléger la figure), et la figure 7 montre les points réunis par genres. L'intérêt est de voir si les points appartenant à un même genre forment un petit nuage bien groupé, et comment se répartissent les genres.

4.3.1. Qualité des groupements.

Certains genres apparaissent homogènes : *Kantou*, *Autranella* ; d'autres, au contraire, très hétérogènes : *Capurodendron*, *Faucherea*. Cette hétérogénéité peut entraîner une révision critique de certains individus ; exemple : un *Capurodendron*, très loin des autres et inséré au voisinage des *Mimusops*, est-il bien analysé sans erreur ? et si oui, sa détermination est-elle indis-

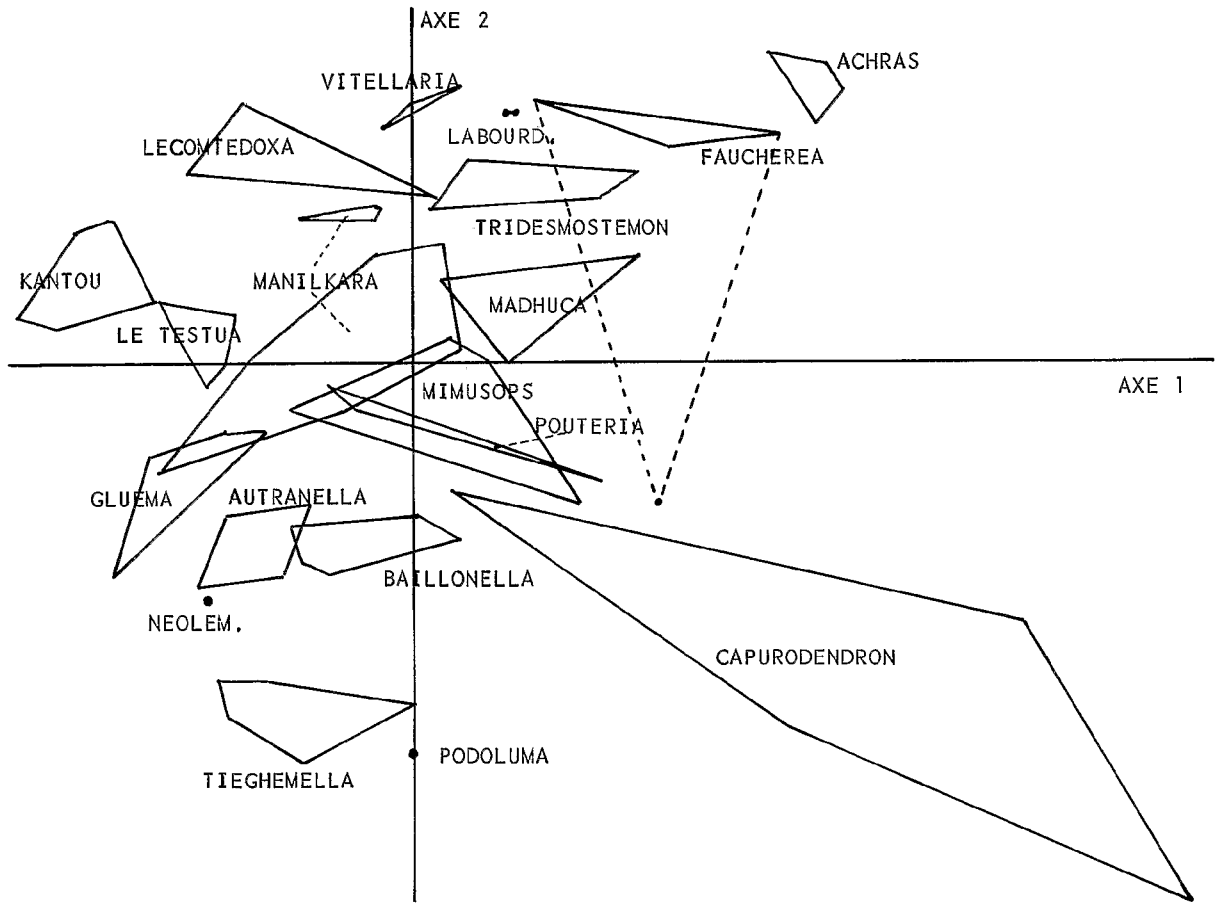
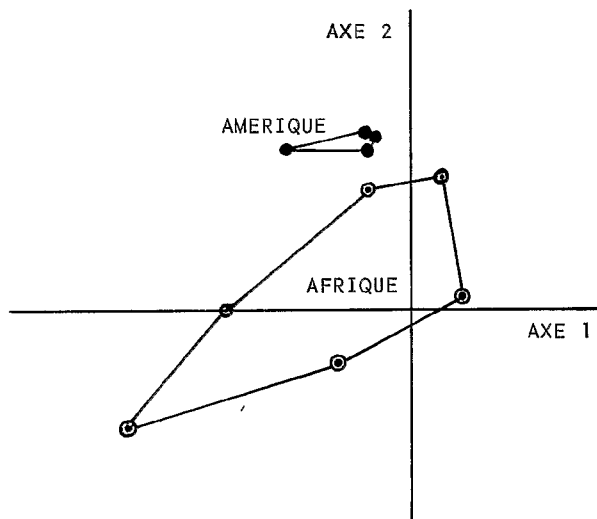


FIG. 7. — Contour des sous-images des genres dans le plan 1-2.

table ? De même, les *Faucherea* seraient bien groupés si l'un des points n'avait une ordonnée aberrante dans le premier plan (à cause de contenus anormaux). Le

groupe des *Manilkara* se partage en 2 sous-groupes généralement bien différenciés, notamment par la présence de silice dans les rayons des Africains (fig. 8).



4.3.2. Séparation des genres.

Il apparaît dans le 1^{er} plan un assez bon isolement de certains genres : *Achras*, *Vitellaria*, *Kantou*, *Tieghemella*, *Capurodendron* (malgré un point gênant).

D'autres genres, moins bien séparés par le 1^{er} plan, le sont parfaitement à l'aide du 2^e plan (fig. 9), ainsi : *Autranella* et *Baillonella*. Chaque genre n'étant représenté ici que par quelques points, il ne serait pas réaliste de considérer comme étant séparés, des groupes qui restent très proches ou légèrement imbriqués. Mais l'intérêt est surtout d'examiner la proximité relative des genres, en fonction de leur position systématique. Ainsi le groupe des Mimosopées (*Mimusops*, *Autranella*, *Vitellaria*, *Tieghemella*, *Baillonella*) n'apparaît pas comme une zone exclusive dans le nuage général et s'étale largement sur l'axe 2. Néanmoins,

FIG. 8. — A.C.P. — Plan 1-2. Les nuages de points des *Manilkara* d'Afrique et d'Amérique.

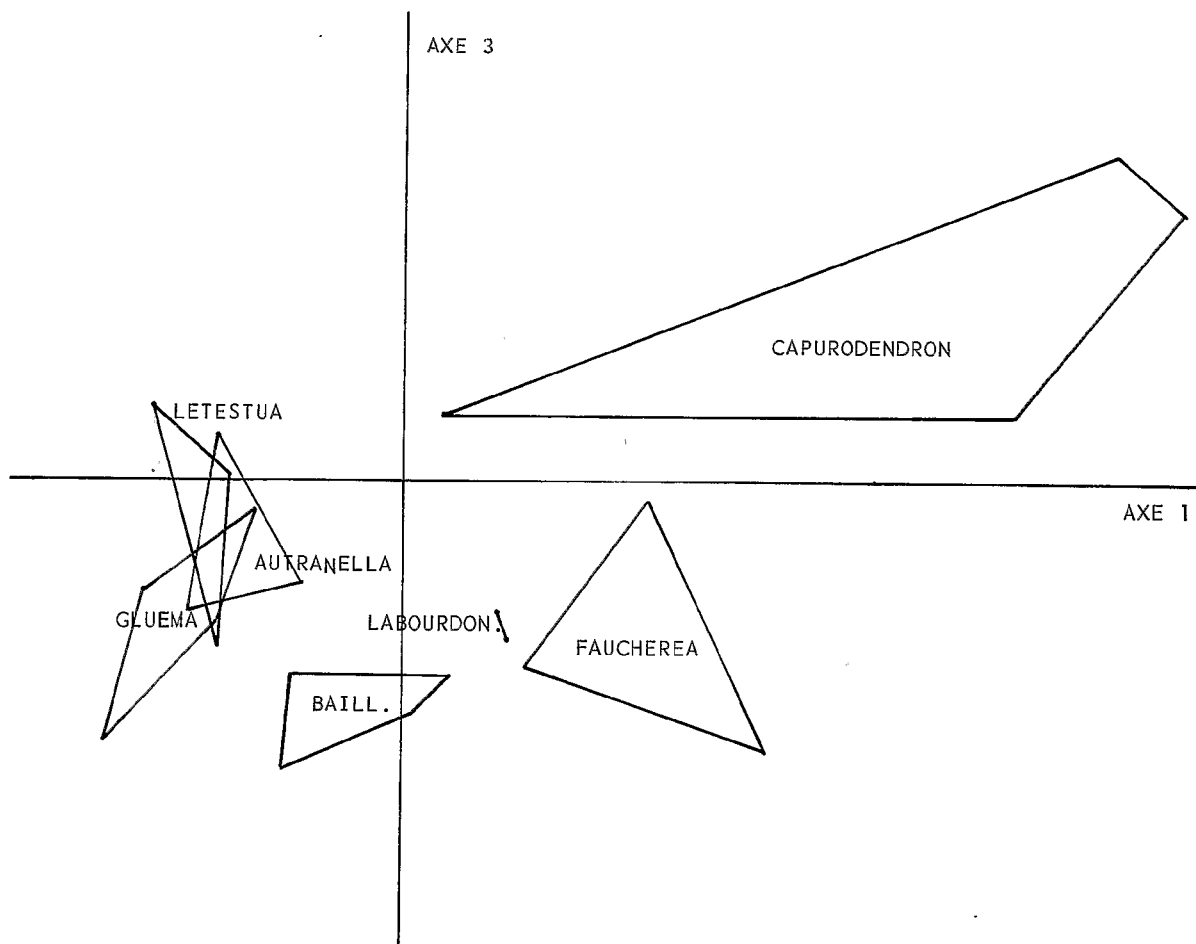


FIG. 9. — Quelques sous-nuages de genres dans le plan 1-3.

certains genres forment des petits groupes bien localisés et assez peu emmêlés dans leurs voisins : *Labourdonnaisia*, *Letestua*, *Gluema*.

4.3. Critique du choix des caractères, A. C. P. réduite.

Remontant de la critique individuelle de quelques points à la réflexion d'ensemble sur la qualité du nuage et de ses composants, on observerait par l'étude des cercles des corrélations, non abordée ici, que les caractères de *structure* sont peu représentés dans les composantes principales 2 et 3 où dominant, d'une part les contenus, d'autre part l'épaisseur des parois. Pour un essai de panorama global des relations entre classification et structure, ces facteurs intéressants, mais néanmoins secondaires, ont pris trop de poids. Aussi avons-nous recommencé l'exécution du programme A. C. P. avec un choix réduit de variables, ne retenant que les 19 caractères de structure, excluant les cristaux, la silice, les globules huileux, la masse volumique et même le coefficient de souplesse des fibres.

On obtient de nouvelles corrélations entre variables et axes principaux, mais ceux-ci n'apparaissent pas

représentatifs d'un faisceau naturel de caractères, avec un axe bien corrélé aux données des vaisseaux, un autre aux rayons, etc, ... Par exemple, le 1^{er} axe est lié fortement au diamètre des vaisseaux comme au nombre des rayons.

Les sous-nuages de genres sont au moins aussi imbriqués que dans la première analyse, c'est pourquoi ils ne sont pas illustrés ici, sauf quelques genres de Madagascar (voir fig. 11 gauche). Mais cela n'enlève rien à l'intérêt de cette analyse, quand on étudie un groupe précis de quelques genres, ou d'espèces dans un genre. On regardera leur position relative dans 2 à 3 plans, pour juger de leurs affinités structurales.

4.5. A. C. P. modifiée : variables rangées

On peut remarquer sur les figures des nuages de points, que ceux-ci ne sont pas répartis d'une façon équilibrée dans les plans principaux. Cela est dû à la dissymétrie des histogrammes de valeur de certains caractères.

Une méthode différente a été essayée. Elle consiste à ranger les individus par valeur croissante d'un

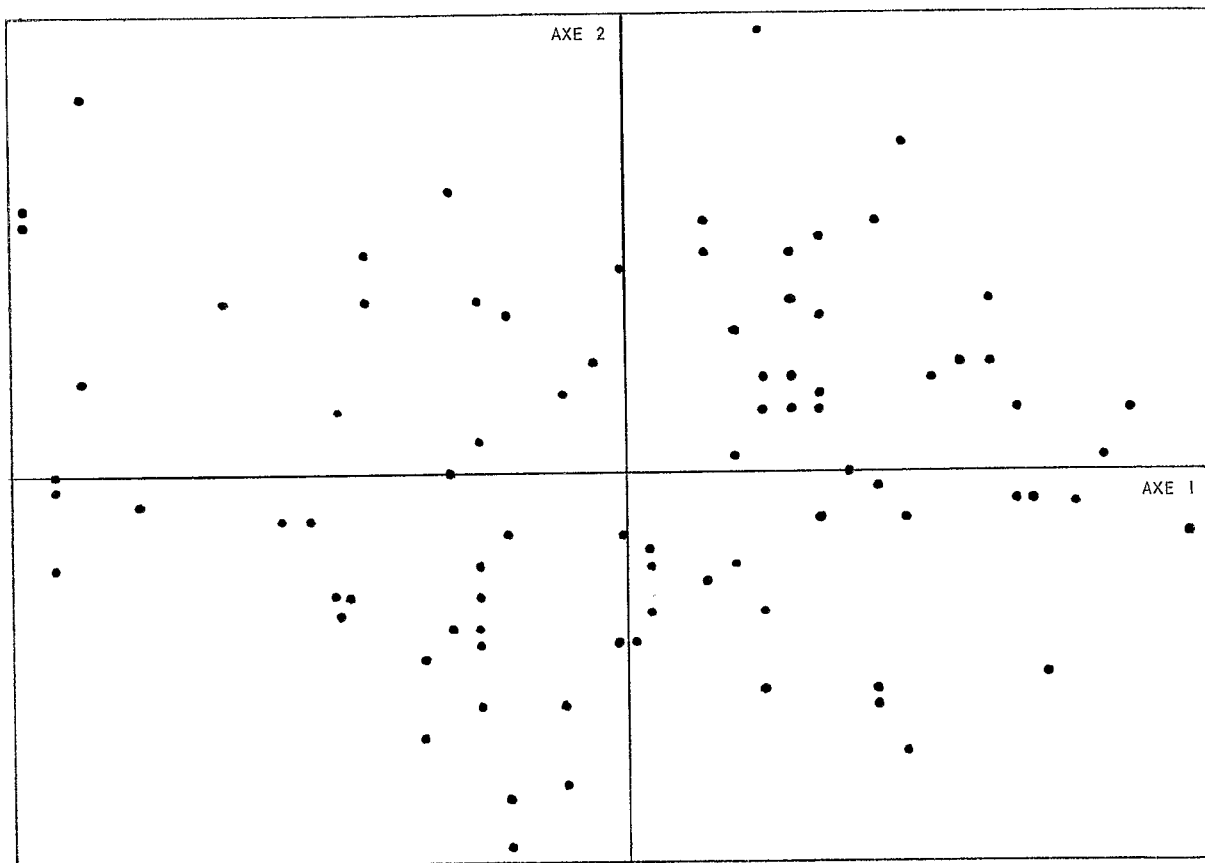


FIG. 10. — A.C.P. par rang. Plan 1-2, nuage général de 84 points.

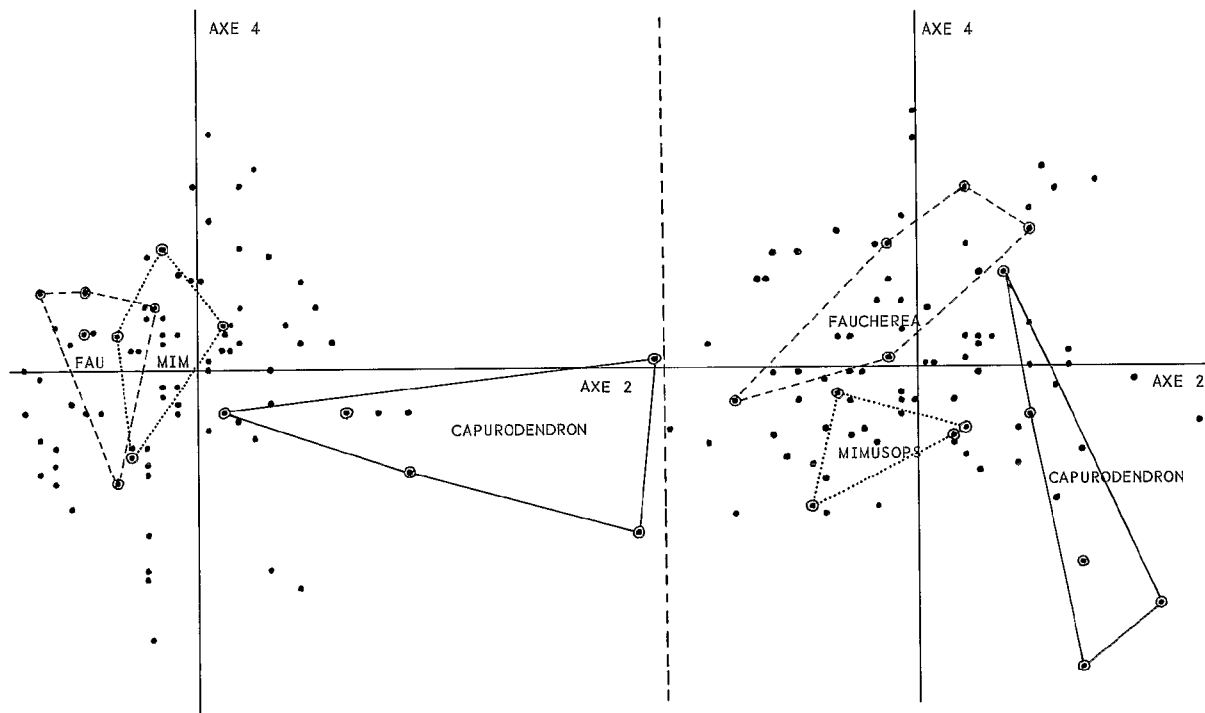


FIG. 11. — A.C.P. à 19 variables. A droite A.C.P. par rangs, dans le plan 2-4, avec figuration des 3 genres de Madagascar. Meilleure répartition du nuage à droite.

caractère et à donner à chacun des 84 individus son propre rang comme valeur de ce caractère. Ainsi, le plus grand nombre de vaisseaux (53 par mm²) se rencontre sur le *Capurodendron* n° 17 qui aura donc, comme valeur de ce caractère, 84 et non 53, et ne diffèrera que de 3 unités de *Capurodendron* n° 19, qui a 28 vaisseaux par mm², et se place au rang 81. On rassemble ainsi les positions trop étalées et l'on espace les individus trop resserrés sur l'échelle des valeurs.

A partir de là les calculs de corrélations des caractères sont donc les corrélations des rangs.

Les nuages de points qui en résultent sont assez

sensiblement modifiés, mieux répartis (Fig. 10, plan 1-2). La figure 11 montre l'influence de ce traitement sur la forme des groupes et leur position relative dans un même plan principal. Il se trouve que les 3 genres représentés sont un peu mieux séparés, mais rappelons qu'il s'agit moins, dans cette méthode, de *séparer* que de *représenter* naturellement les genres en fonction de leur structure.

S'il s'agit, au contraire, de progresser dans leur séparation, il est plus intéressant d'aborder l'autre analyse, destinée à obtenir *volontairement* le meilleur groupement de chaque genre.

5. — APPLICATION DE L'ANALYSE DISCRIMINANTE

5.1. L'analyse discriminante.

Le principe de l'analyse discriminante est différent de celui de l'A. C. P., car on tient compte de l'appartenance des individus à des groupes.

Comme l'A. C. P., l'analyse discriminante consiste à projeter les nuages de points sur un plan, mais le nuage projeté — donc le plan de projection — doit répondre à une double condition :

— que les centres de gravité des sous-nuages des genres soient le plus dispersés possible (le centre de gravité a pour coordonnées les moyennes des caractères des individus du genre),

— que les sous-nuages (genres) projetés soient le mieux groupés possible autour de leurs centres de gravité.

On définit un premier axe discriminant par une recherche « pas à pas » : quelle est la variable la plus discriminante à elle seule ? puis quelle est la variable qui, s'ajoutant à la première, améliore le plus la discrimination ? puis la troisième, etc... On arrive ainsi à un axe qui maximise le rapport de la variance inter-genres à la variance intra-genres.

Puis on recherche un 2^e axe discriminant de la même façon dans le plan orthogonal au 1^{er} axe ; puis un 3^e axe.

Les 2 premiers plans ainsi définis donnent les meilleures projections discriminantes du nuage de points.

L'analyse discriminante qui a été effectuée utilise le programme BMDP 1-M de la Bibliothèque de l'Université de Californie (Biomedical Computer Programs).

5.2. Choix des caractères.

Comme pour la 2^e A. C. P., une sélection a été faite parmi les caractères anatomiques utilisés, mais dans un autre esprit visant surtout à alléger un peu le programme : les contenus, très discriminants, ont été conservés (sauf les globules huileux, trop restreints au genre *Capurodendron*), mais on a éliminé des caractères redondants comme la masse volumique, bien corrélée avec le coefficient de souplesse, ou des caractères apparus dans l'A. C. P. comme particulièrement peu explicatifs. Au total, 17 variables sont maintenues.

D'autre part, les calculs s'appliquant à des groupes ont entraîné l'élimination de 3 genres représentés par un individu unique, et il reste 81 bois pour constituer notre population.

5.3. Calculs préliminaires en A. D.

Certaines étapes du calcul donnent des résultats intéressants par eux-mêmes ; ils sont brièvement décrits ici.

5.3.1. Le plus important est le *pouvoir discriminant individuel* de chaque variable, donné par l'analyse de variance à un facteur. Dans le cas de groupes suffisamment nombreux et de variables distribuées normalement, il est exprimé par la variable de Fischer présentée au tableau 3. Ici, les conditions de validité n'étant pas remplies, le test n'a qu'une valeur indicative.

TABLEAU 3

POUVOIR DISCRIMINANT DE CHAQUE VARIABLE

| | Variable | |
|-----|---------------------------------------|-------|
| PR | Rectitude du parenchyme | 23,93 |
| VP | Diamètre des ponctuations | 23,85 |
| RS | Silice dans les rayons | 20,44 |
| PS | Silice dans le parenchyme | 17,57 |
| CS | Coefficient de souplesse | 15,50 |
| PC | Cristaux dans le parenchyme | 14,85 |
| PSE | Sérialisation du parenchyme | 14,84 |
| PI | Nombre d'interrayons du parenchyme | 9,70 |
| RN | Nombre de rayons | 9,26 |
| VN | Nombre de vaisseaux | 8,54 |
| VG | Groupement des vaisseaux | 6,49 |
| VD | Diamètre des vaisseaux | 6,28 |
| RH | Hauteur des rayons | 4,60 |
| FLO | Longueur des fibres | 4,32 |
| FLA | Largeur des fibres | 3,73 |
| RCO | Nombre de cellules couchées de rayons | 3,23 |
| V | Disposition des vaisseaux | 2,58 |

C'est la rectitude du parenchyme qui discrimine le plus, et la disposition des vaisseaux qui discrimine le moins ; résultat tout à fait inattendu.

5.3.2. Pourcentage des « bien classés ».

Les résultats de l'analyse discriminante relative à chaque pas étant volumineux, ne sont pas reproduits.

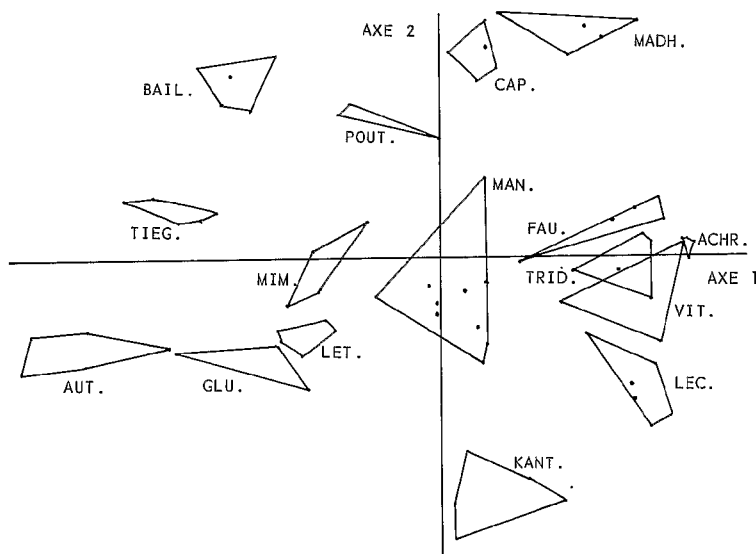


FIG. 12. — Analyse discriminante, plan 1-2.

Parmi ces résultats figure la comparaison de la classification botanique vraie (en genres) et de la classification fournie par l'analyse discriminante. Dans cette dernière classification, chaque individu est placé dans le genre du centre de gravité duquel il est le plus proche. Le programme fournit, comme indice de concordance des deux classifications, le « pourcentage des biens classés ». Le tableau 4 indique l'ordre dans lequel les variables ont été introduites d'un pas à l'autre, et le pourcentage de biens classés à chaque pas.

On notera que l'ordre dans lequel entrent les variables, n'est pas, *a priori*, l'ordre des pouvoirs discrimi-

TABLEAU 4
POURCENTAGE DE BOIS BIEN CLASSÉS
DANS LEUR GENRE APRÈS INTRODUCTION
DE CHAQUE VARIABLE

| Pas | Variable | Bien classés % |
|-----|----------|----------------|
| 1 | PR | 12,3 |
| 2 | VP | 45,7 |
| 3 | RS | 60,5 |
| 4 | PS | 60,5 |
| 5 | PC | 64,2 |
| 6 | PSE | 77,8 |
| 7 | PI | 82,7 |
| 8 | CS | 86,4 |
| 9 | VN | 86,4 |
| 10 | RN | 93,8 |
| 11 | VG | 92,6 |
| 12 | RH | 95,1 |
| 13 | VD | 95,1 |
| 14 | FLO | 95,1 |
| 15 | V | 97,5 |
| 16 | RCO | 97,5 |
| 17 | FLA | 98,8 |

nants individuels des variables (tableau 3) car le couple formé par 2 variables à fort pouvoir discriminant individuel mais corrélées, peut être moins discriminant que le couple formé par 2 variables à pouvoir discriminant plus faible mais non corrélées. En fait, dans notre cas, l'ordre d'introduction des variables est pratiquement l'ordre des pouvoirs discriminants individuels, car les variables retenues ne sont pas redondantes.

5.4. Les nuages de points de l'analyse discriminante.

Les figures 12 et 13 montrent la projection des individus dans les 2 premiers plans discriminants (plan engendré par les axes discriminants 1 et 2, puis 1 et 3).

Nous avons tracé les polygones de contour de tous les sous-nuages de genres dans le plan 1-2.

Du premier coup d'œil, on constate que, dans ce plan 1-2, les sous-nuages sont bien groupés entre-eux, et, dans l'ensemble, assez éloignés les uns des autres, sauf quelques-uns: *Tridesmostemon*, *Vitellaria* et *Achras*, d'une part, *Autranella*, *Gluema* et *Letestua*, d'autre part.

Si l'on examine la projection de ces seuls genres dans le plan 1-3, (fig. 13), on constate que les nuages *Autranella*, *Gluema* et *Letestua*, ne sont absolument plus voisins. De même, les nuages *Achras*, *Vitellaria*, *Tridesmostemon*, sont remarquablement séparés.

Ainsi, par l'analyse factorielle discriminante, nous obtenons une séparation claire de tous les genres de notre lot d'échantillons.

5.5. Identification d'un individu.

Dès lors que les genres forment des sous-nuages de points bien groupés, un bois inconnu, s'il appartient à l'un de ces genres, sera valablement déterminé par le sous-nuage dans lequel il viendra se placer.

Pour obtenir la position du point de cet inconnu, le calcul qui a donné les compositions des axes discriminants en fonction des 17 caractères retenus, donnera inversement les coordonnées du point sur les 3 axes, à partir des 17 valeurs de ses caractères individuels.

Il sera de la forme :

$$X_1 = \sum_0^{17} a_i(x_i - \bar{x}_i)$$

où :

- X_1 est la composante sur le 1^{er} axe de l'individu inconnu,
- a_i est le coefficient discriminant selon le 1^{er} axe pour le caractère anatomique i ,
- x_i est la valeur du caractère i pour l'individu inconnu,
- \bar{x}_i est la moyenne obtenue pour le caractère i , sur les 84 individus analysés.

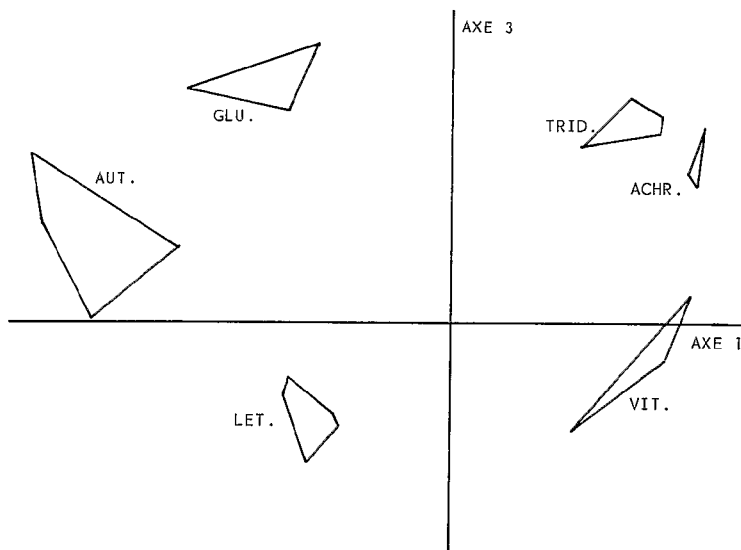


FIG. 13. — Analyse discriminante, plan 1-3. Les genres les plus proches dans le plan 1-2 apparaissent bien séparés dans ce plan, donc dans l'espace.

CONCLUSION

Nous sommes partis d'un lot d'échantillons qui ne différaient les uns des autres que par de légères variations de nombreux caractères, sans qu'il soit facile de fixer, pour chaque caractère, des limites entre les genres.

L'analyse en composantes principales a fait apparaître les 84 points-échantillons en « nuages » représentant le mieux possible, dans quelques plans, l'ensemble de leurs caractères, d'une façon absolument indépendante de leur identité.

A condition que les caractères soient judicieusement choisis et décrits, la méthode paraît riche en applications dans le domaine de la Systématique. Par exemple, un écart important ou, au contraire, la superposition de 2 groupes de points, seront des arguments plus convaincants apportés par l'anatomie du bois pour aider le systématicien à prendre position sur des genres

botaniques contestés, que leur différence sur un seul ou deux caractères.

Le point faible de la méthode est la difficulté de choix objectif des caractères à valeur systématique. Il n'est pas évident que la présence de contenus, résultant d'un métabolisme spécifique, ait moins de valeur fondamentale que le développement du parenchyme ou l'épaisseur des parois. Mais si cette présence a beaucoup de poids et qu'elle est cependant facultative, restant à l'état potentiel selon les conditions de moment ou de milieu, l'analyse en est perturbée.

Quant à l'analyse factorielle discriminante, son application, pour établir de nouveaux systèmes d'identification, paraît une expérience tout à fait intéressante à poursuivre dans des cas complexes comme l'ensemble des bois de *Sapotaceae*, ou plutôt, au niveau d'une division botanique donnant des difficultés d'identification.

RÉFÉRENCES

- (1) CAILLIEZ (F.) & PAGES (J. P.), 1976. — Introduction à l'analyse des données SMASH, 9, rue Duban, 75016 Paris, 640 p.
- (2) CAILLIEZ (F.) & GUENEAU (P.), 1972. — Analyse en composantes principales des propriétés technologiques des bois malgaches. C. T. F. T., Cahiers Scientifiques, n° 2, 51 p.
- (3) VITALIS-BRUN (A.), 1979. — Utilisation de l'anatomie et de la statistique dans la séparation des genres de Sapotacées tropicales à bois rouges et denses. Université Pierre-et-Marie-Curie (Paris 6), thèse pour l'obtention du titre de Docteur de l'Université, non publiée.