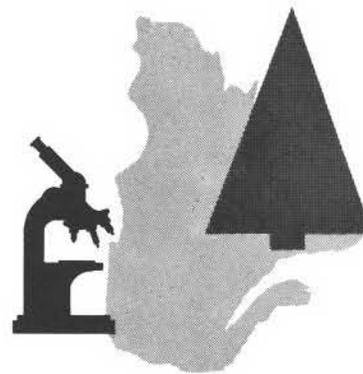




GOUVERNEMENT DU QUÉBEC  
MINISTÈRE DES TERRES ET FORÊTS  
SERVICE DE LA RECHERCHE



NOTE N° 4, 1975

L'ÉLABORATION DE TABLEAUX PHYTO-ÉCOLOGIQUES À L'AIDE DE CLASSIFICATIONS  
AUTOMATIQUES

Guy Waksman<sup>1</sup>

O.C.D. 182.5--015 (047) 714 L.C. QK 901

RÉSUMÉ

Dans cette étude, un tableau de végétation est élaboré par combinaison des résultats des classifications automatiques des relevés et des espèces. Ce type de classification entraîne quelques distorsions par rapport aux données initiales. L'écologue constate donc des anomalies dans les résultats et les corrige grâce à sa connaissance des milieux étudiés.

SUMMARY

*In this study, a vegetation table is created by combining results obtained by automatic classifications of samples and species. This type of classification leads to some distortions compared with initial data. The ecologist observes some anomalies in the results and corrects them according to his knowledge of the studied sites.*

1. INTRODUCTION

La recherche des associations végétales procède en trois étapes. La première est la répartition des relevés: c'est le problème de l'échantillonnage. La seconde est le contrôle de l'homogénéité (Guinochet, 1955; Godron, 1968). Le phyto-écologue note les espèces présentes sur une aire assez grande pour que "l'on ait des chances de trouver à peu près toutes les espèces de l'individu d'association considéré" (Guinochet, 1955). C'est l'aire minimum

<sup>1</sup> Guy Waksman, coopérant technique français

qui doit représenter la surface floristiquement homogène qu'est l'individu d'association. La notion d'homogénéité floristique "n'a rien à voir avec l'homogénéité physiologique conférée, par exemple, par la dominance d'une espèce" (Guinochet, 1955). Le forestier tient cependant naturellement compte de la dominance des espèces et établit les relevés en des endroits floristiquement homogènes et physiologiquement représentatifs de l'ensemble de l'individu d'association pour en mesurer la productivité.

Enfin, la troisième étape est l'exploitation des relevés. Celle-ci est le plus souvent entreprise à l'aide de tableaux de végétation.

## 2. LE TABLEAU DE VÉGÉTATION

Le tableau de végétation est un tableau rectangulaire dont chaque colonne représente un individu d'association et chaque ligne une espèce végétale. Chaque case du tableau contient soit une simple indication de présence-absence (1 ou 0), soit un coefficient d'abondance-dominance. Ce tableau peut contenir des notations décrivant le milieu où chacun des relevés a été établi: type de sol, type d'humus, altitude, recouvrement des différentes strates, âge du peuplement, etc. Un tel tableau est souvent appelé tableau phytosociologique ou encore tableau phyto-écologique.

Le tableau de végétation apparaît comme le résultat d'une étude phyto-écologique. C'est le document final sur lequel l'utilisateur éventuel s'appuiera. Mais le tableau est d'abord un outil de travail pour l'écologue. Celui-ci effectue une classification des relevés en reconnaissant des similitudes écologiques et floristiques entre relevés d'où il déduit l'existence d'associations végétales. De cette première classification découle une classification des espèces. En effet, après avoir regroupé les relevés (les colonnes), on regroupe les espèces (les lignes) qui sont caractéristiques de chaque association. Le tableau de végétation naît donc par approximations successives en permutant respectivement les colonnes puis les lignes.

Dans le cas où on dispose d'un grand nombre de relevés, l'élaboration d'un tableau devient délicate et fastidieuse. Aussi a-t-on envisagé de procéder à des classifications automatiques des relevés et des espèces puis d'en confronter les résultats dans un tableau de végétation.

## 3. MÉTHODES DE CLASSIFICATION

Une classification automatique comprend deux éléments: la définition d'une distance entre individus et celle d'un algorithme de classification. L'algorithme utilisé pour la classification des relevés et des espèces est le même. Mais les distances entre espèces et celles entre relevés sont calculées de façon tout à fait différente.

### 3.1 UNE DISTANCE ENTRE RELEVÉS: L'INDICE DE DICE

L'indice de Dice (1965, *in* Benzécri *et al.*, 1973) utilise les coefficients de présence-absence. Pour le calculer, on considère uniquement comme coïncidences entre deux relevés K et L les lignes du tableau phytologique où il y a un 1 dans chacune des deux colonnes K et L;  $N_{KL}$  est le nombre de ces coïncidences (c'est-à-dire le nombre d'espèces communes aux deux relevés). L'indice de Dice est égal à :

$$D(K,L) = \frac{2 N_{KL}}{2N_{KL} + U_{KL}}$$

où  $U_{KL}$  est le nombre de lignes pour lesquelles les colonnes K et L diffèrent.

Cet indice est discuté par G. Roux et M. Roux (*in* Benzécri *et al.*, 1973). D'après ces auteurs, l'indice de Dice semble donner les meilleurs résultats tout particulièrement lorsque, après un travail de réflexion préalable, sont éliminées des relevés les plantes appelées compagnes, abondantes et non caractéristiques. Pour notre part, nous n'éliminons aucune espèce. Il pourrait donc être utile d'essayer un indice utilisant les coefficients d'abondance-dominance.

### 3.2 UNE DISTANCE ENTRE LES ESPÈCES: LA DISTANCE DE L'ANALYSE FACTORIELLE DES CORRESPONDANCES.

Le calcul des distances entre espèces n'est pas effectué ici sur la base de leurs présences conjointes dans les relevés mais plutôt par comparaison des profils écologiques pour différentes variables. Rappelons que le profil écologique d'une espèce pour une variable donne le nombre de relevés effectués dans chacune des classes de la variable et où l'espèce a été trouvée.

Les données sont présentées, non sous la forme d'un tableau de végétation, mais sous la forme d'un tableau relatif à l'ensemble J des espèces et à l'ensemble I des états des différentes variables. Chaque élément du tableau représente alors le nombre de fois où une espèce j a été trouvée dans un état i des variables (Romane, 1972). Un tel tableau de données est analysable par l'analyse factorielle des correspondances. A cette méthode est associée une distance entre les espèces, distance qui peut être considérée comme une mesure de dissimilarité entre les profils des espèces.

Cette distance <sup>1</sup> est calculée pour toutes les espèces prises deux à deux. Pour les relevés comme pour les espèces, la matrice triangulaire des distances (ou des dissimilarités) est représentée schématiquement sous la forme d'un arbre (ou dendrogramme) construit à l'aide de la méthode "Arbre de portée minimum". (Gower et Ross, 1969).

### 3.3 UN ALGORITHME DE CLASSIFICATION

La matrice triangulaire étant calculée, les individus les plus proches sont regroupés et considérés comme un individu unique. Les distances à ce nouvel individu sont recalculées selon un critère choisi arbitrairement et particulier à chaque algorithme. Pour des raisons techniques, le critère adopté ici est celui de l'"ultramétrie inférieure maxima" <sup>2</sup>. Si a et a' sont deux individus regroupés, la distance de n'importe quel individu b au nouvel individu (a, a') est égal au minimum des distances de b à a et b à a'. L'agrégation de deux parties est effectuée suivant le saut minimum: il suffit d'un seul lien (d'une seule paire d'individus proches) pour qu'on décide de réunir deux parties (Benzécri *et al.*, 1973).

---

<sup>1</sup> Cette distance peut s'exprimer ainsi:

$$d_j^2(I) (j, j') = \sum_{i=1}^n P_i \cdot \left[ \frac{P_{j/i}}{P \cdot j} - \frac{P_{j'/i}}{P \cdot j'} \right]^2$$

où  $d_j(I) (j, j')$  = distance entre deux éléments j et j' de l'ensemble J en correspondance avec l'ensemble I;

$P_{ij}$  = probabilité d'apparition de l'élément (i, j); estimée à partir des fréquences.

$P_i = \frac{\sum_{j \in J} P_{ij}}{j \in J}$  = probabilité d'avoir l'élément i quel que soit j;

$P \cdot j = \frac{\sum_{i \in I} P_{ij}}{i \in I}$  = probabilité d'avoir l'élément j quel que soit i;

$P_{i/j} = \frac{P_{ij}}{P \cdot j}$  = probabilité d'avoir l'élément i, j étant fixé;

$P_{j/i} = \frac{P_{ij}}{P_i}$  = probabilité d'avoir l'élément j, i étant fixé;

<sup>2</sup> Soient a et a' deux parties d'un ensemble I, un point i de a, un point i' de a', d(a, a') la distance entre les deux parties, et b élément non encore regroupé.

$$d(a, a') = \inf_{i \in a, i' \in a'} d(i, i')$$

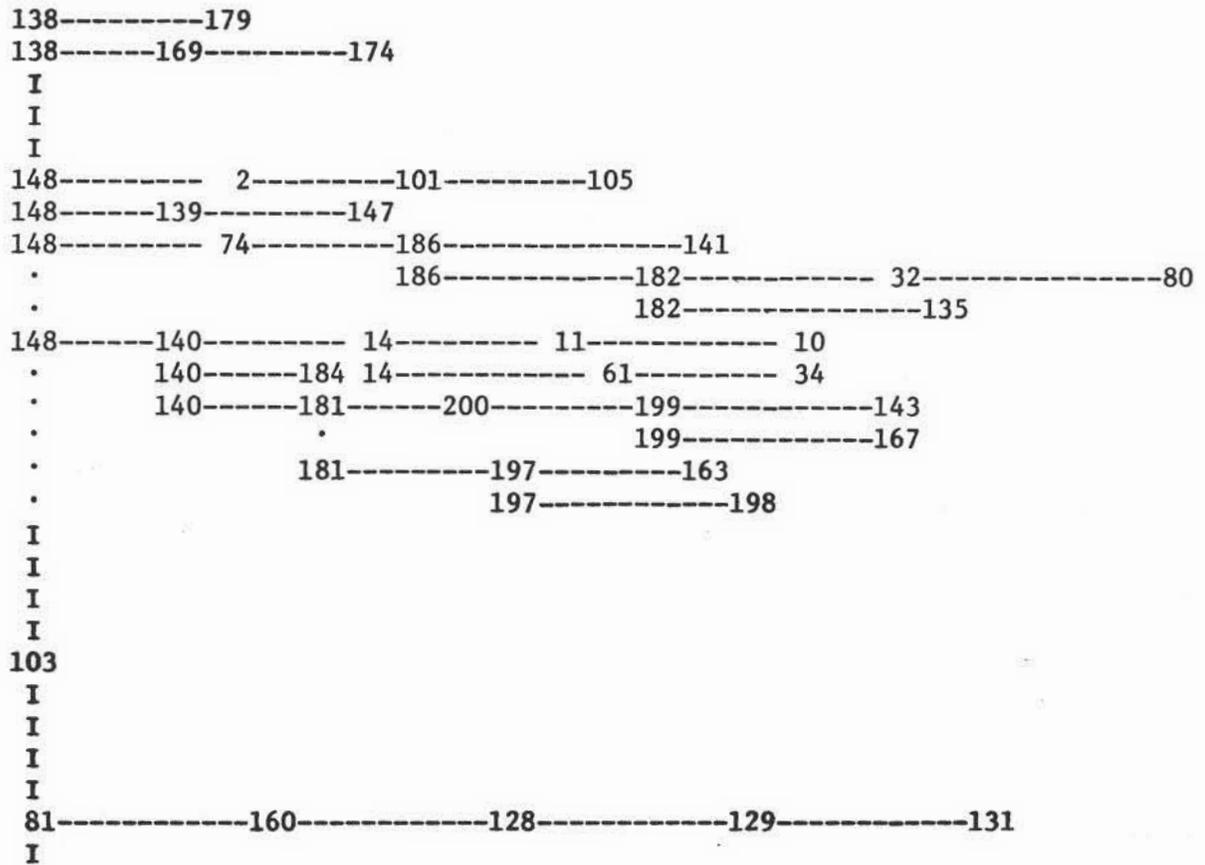
$$d[(a \cup a'), b] = \inf [d(a, b), d(a', b)]$$

Le résultat de la classification est un arbre qui illustre la façon dont les individus ont été successivement regroupés selon le critère retenu. On cherche une signification biologique aux groupes qui sont représentés par les "branches" qui partent de l'axe de l'arbre (figure 1). Ainsi peuvent être mis en évidence soit des groupes d'espèces, soit des groupes de relevés.

#### 4. CONCLUSION

Dans cette note, aucun résultat n'est présenté. Néanmoins, nous nous permettons d'indiquer que la combinaison des deux méthodes de classification est tout à fait fructueuse. La représentation des données par un arbre comporte quelques distorsions par rapport aux observations. La confrontation de la classification des espèces et de celle des relevés, élaborées indépendamment l'une de l'autre, au sein d'un même tableau phytosociologique, permet de constater certaines anomalies qui apparaissent clairement dans le tableau. Il suffit alors de permuter les quelques lignes et colonnes qui ont pu être mal classées.

FIGURE 1: EXEMPLE DE TRACÉ DE L'ARBRE DE PORTEE MINIMUM.



## BIBLIOGRAPHIE

- BENZECRI, J.P. *et al.* 1973. *L'analyse des données*. Tome 1. La taxinomie. Dunod. Paris.
- GODRON, M., 1968. *Quelques applications de la notion de fréquence en écologie végétale*. *Oecol. Plant.*, 3, 195-212.
- GOWER, J.C. et G.J.S. ROSS, 1969. *Minimum spanning trees and single linkage cluster analysis*. *App. Stat.*, 18, (1), 54-55.
- GUINOCHET, M., 1965. *Logique et dynamique du peuplement végétal*. Masson. Paris.
- ROMANE, F., 1972. *Quelques aspects de l'analyse multivariante des données phyto-écologiques*. *Inv. Pesq.*, 36, (1), p. 131-139.

Dépôt légal, Bibliothèque nationale du Québec.

