

Niches écologiques et couplages de tableaux

Résumé

La fiche décrit une méthode d'ordination et deux méthodes de couplage basées sur la notion de marginalité des niches écologiques (Hausser, J., 1995, *Säugetiere der Schweiz. Mammifères de la Suisse. Mammiferi della Svizzera*. Birkhäuser Verlag, Basel, Boston, Berlin. 501 p.). Voisines de l'analyse des correspondances et de l'Analyse Canonique des correspondances, ces procédures s'en éloignent fortement en cas de facteurs écologiques limitant (pollution, salinité) ou favorisant (météorologie) l'abondance d'un groupe important de taxons. Le module Niche permet d'exécuter ces propositions.

Plan

| | |
|---|----|
| 1 — Introduction..... | 2 |
| 2 — La méthode OMI | 4 |
| 2.1 — Principe | 4 |
| 2.2 — Marginalité et tolérance..... | 6 |
| 2.3 — Propriétés de l'axe de marginalité..... | 8 |
| 2.4 — Marginalité multispécifique..... | 9 |
| 3 — Météorologie et piégeages lumineux..... | 10 |
| 4 — Une nouvelle méthode d'ordination | 17 |
| 4.1 — Gradients d'avifaune | 17 |
| 4.2 — Gradients Amont-Aval..... | 20 |
| 5 — Analyse Canonique de la marginalité | 23 |
| 6 — Conclusion | 29 |
| Références | 31 |

D. Chessel & C. Gimaret

1 — Introduction

Soit \mathbf{X} un tableau associant des stations (lignes) et des variables environnementales (colonnes) et un tableau floro-faunistique \mathbf{Y} associant les mêmes stations (lignes) et les mesures de l'abondance des taxons (colonnes). Les méthodes statistiques utiles dans cette situation sont désormais dominées par l'idée que la courbe de réponse d'une espèce à un gradient environnemental est une courbe en cloche dont le centre est l'optimum écologique du taxon et dont l'étalement est caractéristique de la tolérance du taxon pour cette variable de l'environnement. L'omniprésence de ce modèle conduit à concevoir l'organisation des communautés dominée par les problèmes de séparation de niche, qu'on peut résumer par le schéma de principe (figure 1) de l'analyse canonique des correspondances ¹.

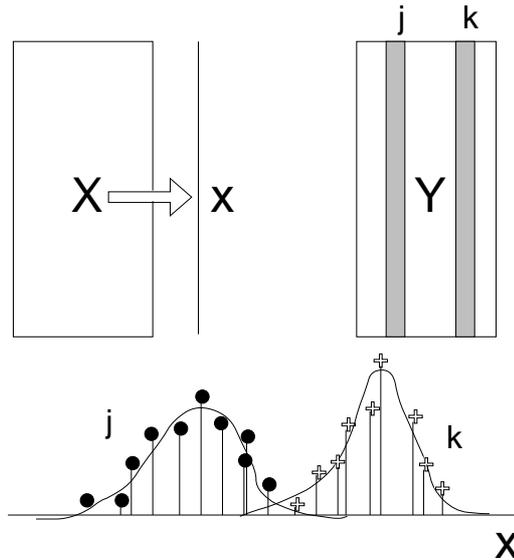


Figure 1 : Principe de l'analyse d'un couple de tableaux écologiques basée sur une AFC.

L'ACC donne une combinaison linéaire \mathbf{x} des variables de \mathbf{X} qui optimise la variance des positions moyennes des espèces. Fondamentalement l'analyse des correspondances de ce point de vue est la meilleure méthode d'ordination d'un tableau floro-faunistique car elle procède, outre d'une définition mathématique aisément utilisable, d'un modèle conceptuel de l'organisation des communautés ². Cet accord entre une procédure et un modèle biologique est l'argument premier des défenseurs de l'AFC en écologie ^{3 4 5}.

L'analyse de co-inertie⁶ des tableaux \mathbf{X} et \mathbf{Y} , dès que \mathbf{Y} a été traité par une AFC, respecte ce point de vue. Les deux analyses n'en sont pas pour autant équivalentes. La première impose à la variable \mathbf{x} d'être de variance unité tandis que la seconde ne le fait pas. Elles ont alors des champs d'applications séparés qu'on peut résumer par la figure 2.

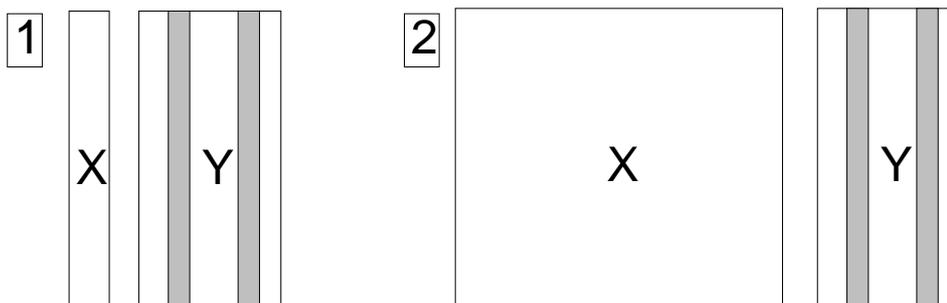


Figure 2 : 1 — Cas typique de l'utilisation de l'ACC. 2 — Cas typique de l'utilisation de l'analyse de co-inertie.

Quand il y a peu de variables de milieu et de nombreux relevés l'ACC s'impose. Quand les colonnes de \mathbf{X} sont en grand nombre, en particulier quand il s'agit de modalités de variables qualitatives, l'ACO est seule possible. Entre les deux chacune a ses qualités et ses défauts. L'ACC ne tient aucun compte des corrélations entre variables de \mathbf{X} et l'ACO peut leur donner trop d'importance.

L'important ici est que, pour arriver à l'objectif commun (séparer les niches), les deux analyses devront faire les mêmes opérations et supporter les mêmes conséquences. Dans les deux cas le tableau \mathbf{Y} devient un tableau de profils de distribution par colonnes. L'abondance d'un taxon ne se conçoit que relative. On note n le nombre de relevés, t le nombre de taxa et on calcule :

$$y_{i.} = \sum_{j=1}^t y_{ij} \quad y_{.j} = \sum_{i=1}^n y_{ij} \quad y_{..} = \sum_{j=1}^t y_{.j}$$

$$f_{i.} = \frac{y_{i.}}{y_{..}} \quad f_{.j} = \frac{y_{.j}}{y_{..}} \quad f_{ij} = \frac{y_{ij}}{y_{.j}}$$

Le relevé prend un poids $f_{i.}$, le taxon prend un poids $f_{.j}$ et le tableau \mathbf{Y} devient le tableau des profils de distribution par taxon \mathbf{F} de terme général f_{ij} . Que la position moyenne du taxon j porte le poids $f_{.j}$ est parfaitement acceptable : une espèce aura une position d'autant plus assurée qu'on l'aura trouvée plus souvent. Ce qui est plus contraignant, c'est que le poids $f_{i.}$ soit assigné au relevé i . Tant qu'on ne parle que du tableau \mathbf{Y} , c'est normal : un relevé est d'autant plus important qu'on y a rencontré plus d'espèces. Si la dimension totale d'un relevé n'est pas contrôlée c'est même indispensable d'en tenir compte.

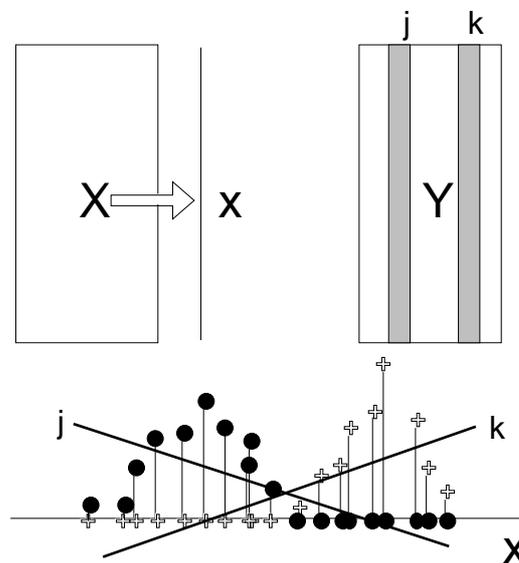


Figure 3 : Principe de l'analyse d'un couple de tableaux écologiques basée sur une ACP

Mais cela oblige à définir la moyenne et la variance des variables de milieu à l'aide de ces poids. Par exemple, on calculera la moyenne d'une variable indicatrice de la pollution sans tenir compte des points les plus pollués, ceci parce qu'ils ne contiennent plus d'organismes vivants ! L'importance d'une mesure de milieu ne peut se mesurer par l'abondance des organismes qu'on trouve au point de mesure. Elle vaut pour elle-même et ce n'est pas parce que l'altitude est trop grande ou que la salinité est trop forte pour les taxa étudiés que la moyenne de ces variables utilise certains points de

préférence à d'autres pour exécuter des opérations aussi simples et élémentaires que des calculs de moyennes, variances ou corrélation.

Si on refuse cette opération on utilisera pour Y une ACP centrée dans la co-inertie ou on préférera l'Analyse en Composantes Principales sur Variables Instrumentales (ACPVI) à l'ACC qui est une AFC sur variables instrumentales ⁷. C'est pourquoi le programme CANOCO ⁸ contient aussi l'analyse des redondances autre dénomination de l'ACPVI.

Dans les deux cas on perdra le modèle de départ au profit du principe de la figure 3. Au lieu de maximiser la variance des positions moyennes des espèces on maximisera la somme des carrés de covariances (ou de corrélations) entre la variable x et les abondances de chaque taxon. On comprend aisément qu'on obtiendra souvent des résultats proches dans les deux modèles, mais qu'il existe des cas (facteurs limitants) où la seconde procédure s'impose (figure 4).

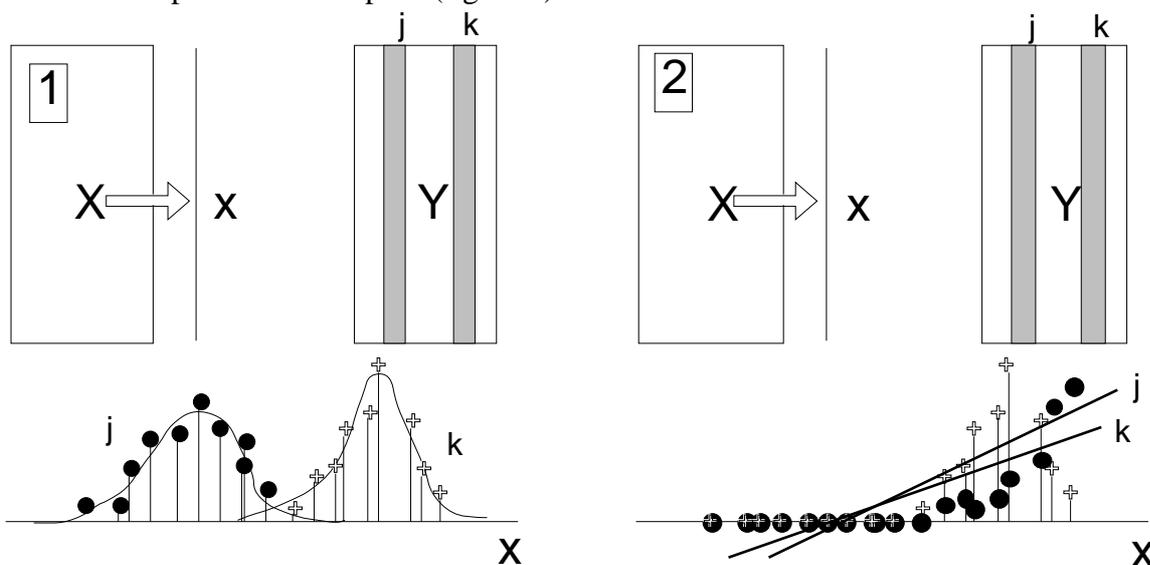


Figure 4 : 1 — Cas typique de l'utilisation de l'ACC (AFCVI) ou de la co-inertie sur AFC.
2 — Cas typique de l'utilisation de l'ACPVI ou de la co-inertie sur ACP.

Si on veut bien tenir compte de l'interaction entre méthodes statistiques et organisation propre des données, une certaine complexité se fait jour pour l'utilisateur. En particulier, si une méthode est meilleure dans un cas et une autre est meilleure dans un autre cas, cela demande de connaître le résultat avant l'analyse ce qui est, somme toute, inacceptable.

C'est pourquoi nous proposons ici une méthode simple qui introduit la notion de niche dans les deux configurations et élimine le problème des pondérations. On l'appellera méthode OMI pour Outlying Mean Index sur laquelle elle est basée. L'idée initiale vient de l'ouvrage de Hausser ⁹ qui s'est inspiré de la thèse de Perrin ¹⁰. Les idées de Hausser concerne essentiellement le problème de la niche d'une espèce et de son analyse dans le but de prédire son abondance. On transpose ici la question à un cortège multispécifique en réservant la question d'origine à un programme en cours de développement.

2 — La méthode OMI

2.1 — Principe

On conserve dans tout ce qui suit la notation X pour un tableau de variables mésologiques comportant n relevés et p colonnes. Ce tableau X est analysé par une

quelconque méthode à un tableau et peut comporter des variables quantitatives d'unités différentes (PCA : Correlation matrix PCA), des variables quantitatives homogènes (PCA: Covariance matrix PCA), des variables qualitatives (MCA : Multiple Correspondence Analysis)¹¹, des variables floues (MCA : Fuzzy Correspondence Analysis)¹² ou un mélange de variables qualitatives et quantitatives (MCA : Hill & Smith Analysis)¹³.

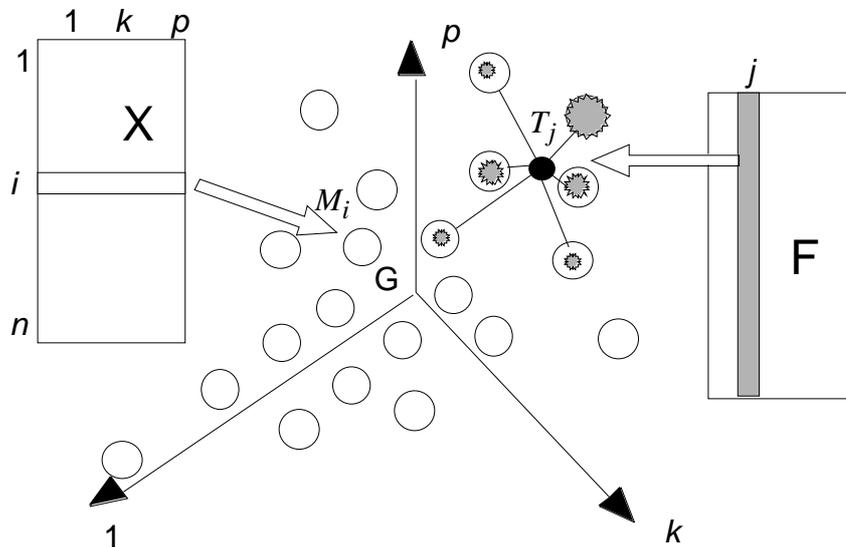


Figure 5 : Eléments de base dans une analyse OMI. Les lignes du tableau X définissent un nuage de points et chaque taxon (colonne de F est une pondération de ces points qui définit un centre de gravité (position moyenne du taxon dans l'espace).

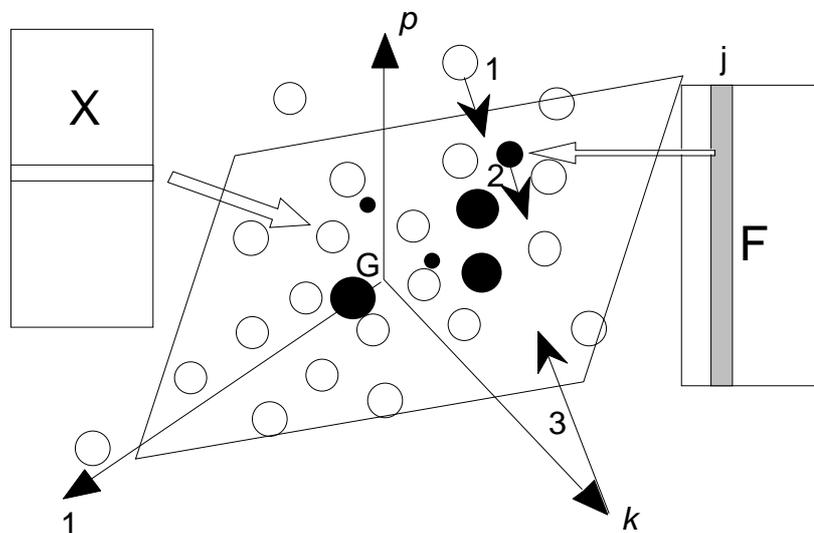


Figure 6 : Représentation simultanée dans une analyse OMI. 1 - Les lignes du tableau X (relevés) 2 - les colonnes du tableau F (position moyenne du taxon dans l'espace) 3 - les vecteurs de la base canonique (variables) sont positionnés par projection orthogonale sur un même plan.

L'analyse préliminaire de X se fait par un triplet (X, Q, D) ¹⁴ défini entièrement indépendamment de la suite.

$$D = \text{Diag}(p_1, \dots, p_i, \dots, p_n)$$

La pondération D peut être uniforme ou non mais ne dépend pas du tableau faunistique. Ce triplet définit un nuage de n points de \mathbb{R}^p dont l'analyse cherche les axes principaux, ce qui n'est pas ici l'objectif.

Le tableau faunistique \mathbf{Y} est transformé en tableau \mathbf{F} des profils colonnes. Chaque colonne définit une pondération des relevés qui engendre un centre de gravité. Notons T_j le centre de gravité de la colonne j du tableau faunistique (figure 5).

Le tableau des positions des centre de gravité est simplement dans $\mathbf{Z} = \mathbf{F}^t\mathbf{X}$, tableau à t lignes (taxons) et p colonnes variables. On conserve les poids des espèces dans la matrice diagonale $\mathbf{D}_t = \text{Diag}(f_1, \dots, f_j, \dots, f_t)$

Du point de vue statistique, l'analyse OMI est simplement l'analyse du triplet $(\mathbf{Z}, \mathbf{Q}, \mathbf{D}_t)$. L'utilisation des deux nuages de points, celui des n relevés initiaux et celui des t centres de gravités des taxons (figure 6) permet immédiatement de concevoir simplement la représentation simultanée des variables, des relevés et des espèces par projection dans un même espace sur un même plan.

2.2 — Marginalité et tolérance

L'extrême simplicité du principe de l'analyse est un argument décisif pour explorer sa validité. Quand l'analyse du tableau \mathbf{X} donne un nuage centré, ce qui est le cas de tous les exemples cités, le nuage des n points a son centre de gravité à l'origine qu'on appelle G .

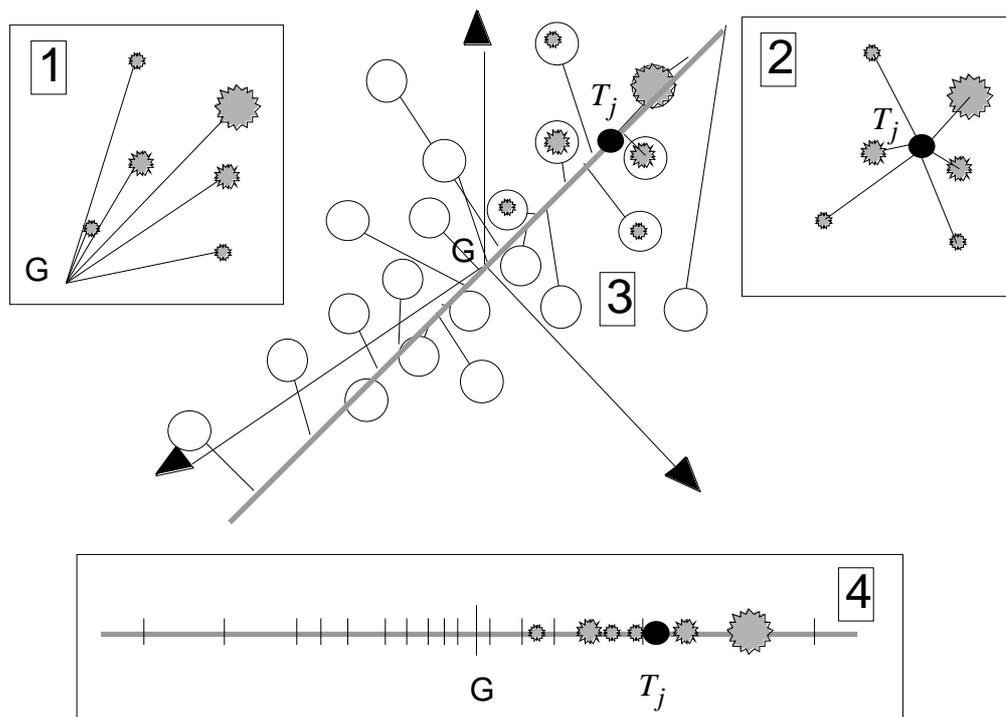


Figure 7 : Propriétés de l'axe de marginalité d'un taxon. 1 - Définition de l'inertie. 2 - définition du centre de gravité. 3 - Projection du nuage initial sur l'axe de marginalité. 4 - Définition de la marginalité.

L'inertie totale du nuage initial est :

$$I_T = \sum_{i=1}^n p_i \|M_i\|_Q^2$$

La quantité :

$$I_T(j) = \sum_{i=1}^n f_{ij} \|M_i\|_Q^2$$

est l'inertie du même nuage quand on utilise la pondération associée au taxon j . Les relevés où le taxon est absent ne jouent alors aucun rôle. La marginalité du taxon se définit simplement par :

$$m_a(j) = \|T_j\|_Q^2$$

Le terme de marginalité (Hausser op. cit. p. 5) a été traduit dans le programme par *Outlying Mean Index* (OMI). La projection du nuage de points sur l'axe de marginalité (la direction définie par le centre de gravité) est immédiate (figure 7). La marginalité s'associe directement à la tolérance sur cet axe. Si m_i est la projection du point M_i sur cet axe, la tolérance marginale est :

$$T_m(j) = \sum_{i=1}^n f_{ij} \|T_j - m_i\|_Q^2$$

Il apparaît immédiatement que les deux paramètres seront très liés, une marginalité forte entraînant nécessairement une tolérance marginale réduite.

On peut déduire des idées de Hausser qu'une opération simple achèvera de décrire la niche de l'espèce. La projection des points sur l'axe de marginalité s'accompagne de la projection sur le sous-espace orthogonal de dimension au plus égale à $p - 1$ (figure 8).

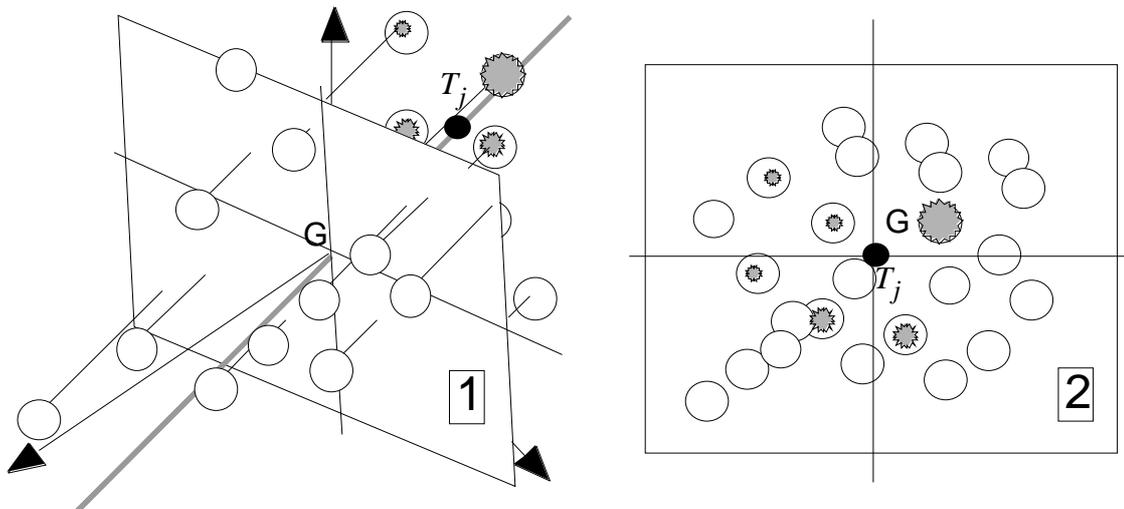


Figure 8 : Définition de la marginalité résiduelle d'un taxon. 1 - Projection du nuage initial sur l'orthogonal à l'axe de marginalité. 2 - Propriété de double centrage du nuage projeté.

La propriété barycentrique étant conservée par projection orthogonale, le nuage projeté est doublement centré et pour la pondération initiale et pour la distribution d'abondance du taxon. Après projection, l'inertie du nuage pour cette distribution peut s'appeler l'amplitude résiduelle ou tolérance résiduelle.

Une double application du théorème de Pythagore dans R^n et dans R^p montre que si cet indice est noté $T_r(j)$ on a :

$$I_T(j) = m_a(j) + T_m(j) + T_r(j)$$

Hausser s'est posé la question "Existe-t-il dans l'espace des directions pour lesquelles la tolérance résiduelle est particulièrement faible ?" et en a donné le principe de recherche.

Nous reviendrons plus tard sur cette question pour nous intéresser d'abord à l'aspect multispécifique de la marginalité.

2.3 — Propriétés de l'axe de marginalité

Très curieusement l'axe de marginalité renvoie aux deux modèles d'approche des couplages abordés dans l'introduction.

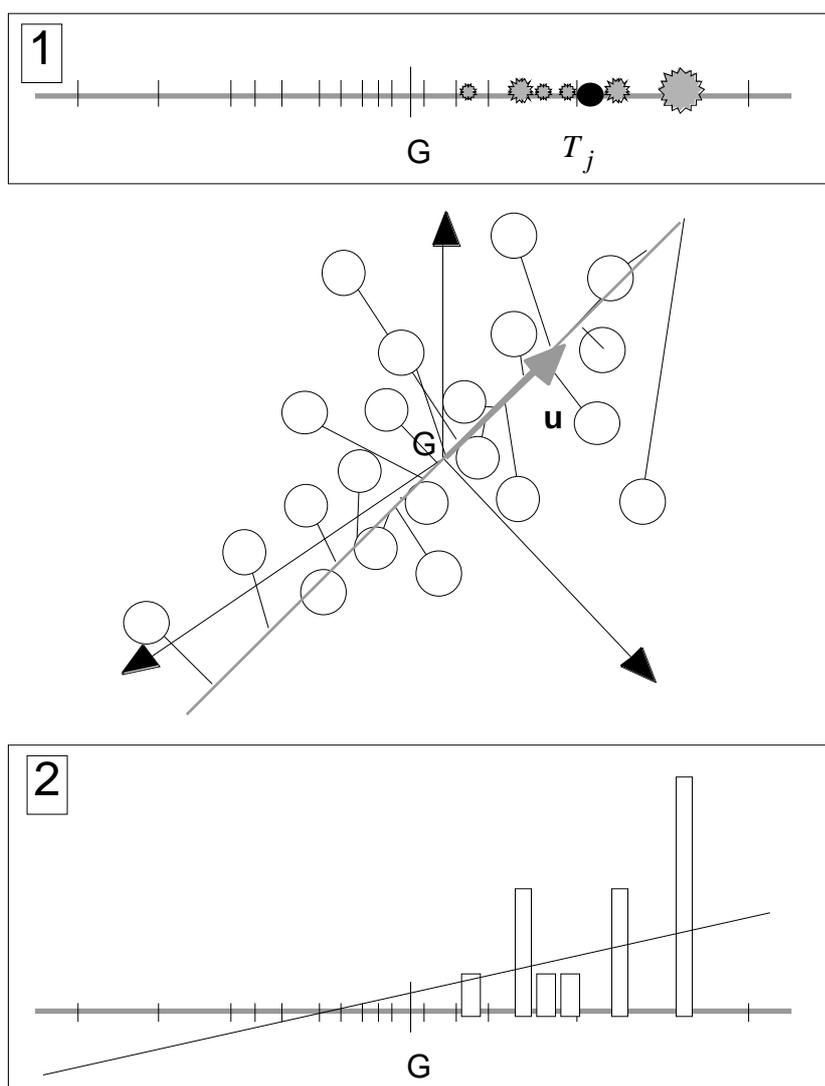


Figure 9 : Double propriété d'optimalité de l'axe de marginalité. 1 - l'abondance du taxon est une distribution de fréquence. 2 - L'abondance de l'espèce est une variable quantitative.

Considérons un vecteur \mathbf{u} de \mathbb{R}^p unitaire pour la norme \mathbf{Q} . On peut projeter dessus le nuage initial, ce qui donne le vecteur des coordonnées \mathbf{XQu} . On peut utiliser ensuite la pondération définie par l'abondance du taxon j et calculer la position du centre de gravité du nuage projeté pour cette pondération. On obtient $\mathbf{f}^t \mathbf{XQu}$ si on note \mathbf{f} le vecteur de coordonnées $(f_{1j}, \dots, f_{ij}, \dots, f_{nj})$.

Chercher le vecteur \mathbf{u} qui optimise la carré de cette quantité, revient à prendre le vecteur unitaire porté par l'axe de marginalité (figure 9) puisque la solution évidente est : $\mathbf{u} = \mathbf{X}^t \mathbf{f} / \|\mathbf{X}^t \mathbf{f}\|_Q$.

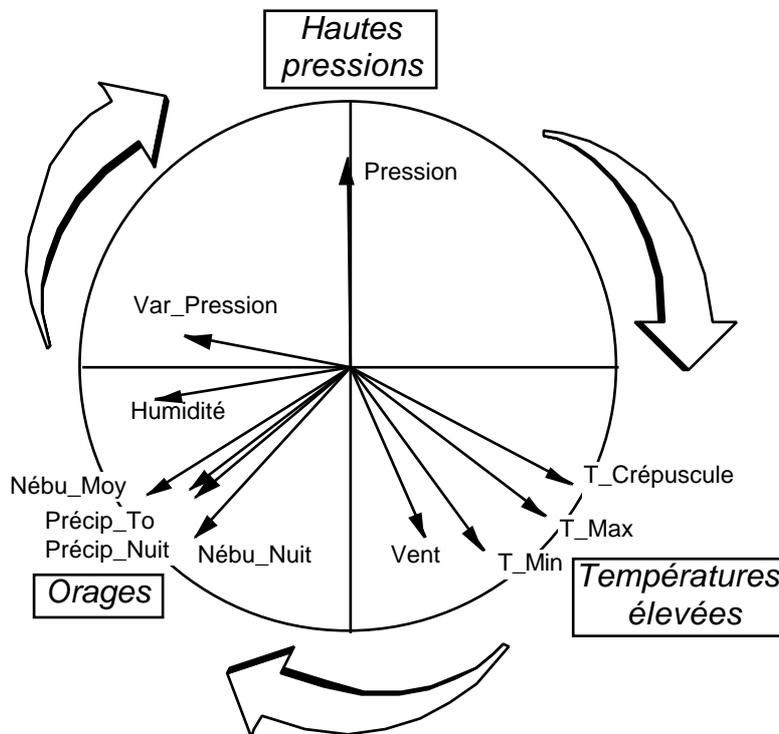


Figure 10 : Cercle des corrélations en ACP normée du tableau des variables météorologiques.

L'axe de marginalité est la direction dans laquelle la position moyenne de l'espèce est la plus éloignée de l'origine.

On peut d'autre part considérer que l'abondance du taxon est une variable quantitative à prédire. La variable $\mathbf{XQ}\mathbf{u}$ est alors considérée comme une combinaison linéaire des variables de milieu qui peut servir de prédicteur de l'abondance. Si on utilise la régression PLS au lieu de la régression ordinaire et si on s'en tient à la première composante PLS¹⁵ on optimisera la covariance entre la coordonnée $\mathbf{XQ}\mathbf{u}$ et la variable \mathbf{f} centrée soit $(\mathbf{XQ}\mathbf{u}\mathbf{f} - \text{moy}(\mathbf{f})\mathbf{1}_n)_D$. Quand \mathbf{D} est la pondération uniforme on obtient le même résultat.

Les deux points de vue ne sont donc pas aussi éloignés qu'on pourrait le croire. Tester par un test de permutation la première composante PLS de la prédiction de l'abondance du taxon ou la marginalité de la niche donne exactement le même résultat.

2.4 — Marginalité multispécifique

Il est donc assuré qu'on peut décrire en première approximation l'influence des variables environnementales sur le taxon par l'indice de marginalité. L'équiprobabilité des $n!$ permutations de la distribution d'abondance sur l'ensemble des relevés donnera un niveau de signification de cette marginalité (logique des tests de permutations¹⁶).

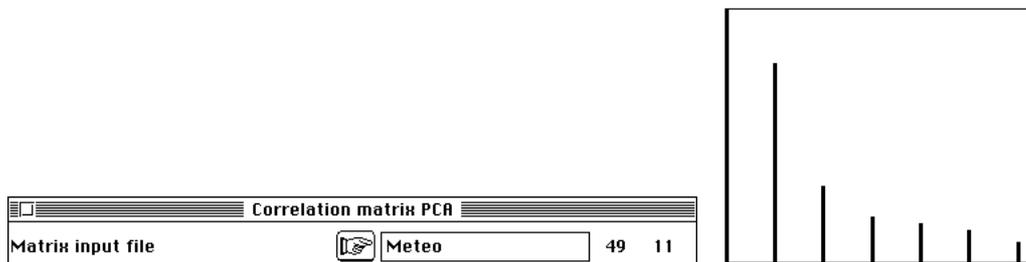
Au niveau multispécifique la marginalité moyenne des taxons (pour la pondération issue du tableau faunistique) n'est rien d'autre que l'inertie totale de l'analyse OMI, analyse du triplet $(\mathbf{Z}, \mathbf{Q}, \mathbf{D}_t)$. En effet :

$$\text{Trace}(\mathbf{Z}^t \mathbf{D}_t \mathbf{Z} \mathbf{Q}) = \text{Trace}(\mathbf{X}^t \mathbf{F} \mathbf{D}_t \mathbf{F}^t \mathbf{X} \mathbf{Q}) = \sum_{j=1}^t f_{.j} m_a(j)$$

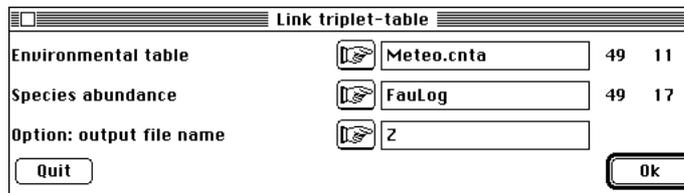
On aura donc un test de signification par permutations des lignes du tableau faunistique dans la logique décrite et étudiée par ¹⁷. On notera que ce test conserve toutes les associations inter-spécifiques et indique sans ambiguïté si le milieu influe sur la liste faunistique globalement.

3 — Météorologie et piégeages lumineux

Un premier exemple de la méthode OMI concerne les données de P. Usseglio¹⁸ éditées dans ¹⁹. Utiliser le dossier créé par la carte Light_trap de la pile ADE-4•Data. Faire l'ACP normée du tableau de variables météorologiques (PCA : Correlation matrix PCA) :



Garder deux facteurs, variables de synthèse qui décrivent l'évolution des paramètres météorologiques et interpréter la carte des variables (figure 10). Pour savoir si la météorologie influe sur les captures par piégeage lumineux, coupler (Niche : Link triplet-table) les deux tableaux Meteo (49-11) et le tableau faunistique (49-17) initialement transformé par $x \rightarrow \text{Log}(x + 1)$:



```
Environmental parameter : Meteo.cnta
---> Rows: 49 Col: 11
Faunistic array : FauLog
---> Rows: 49 Col: 17
File Z.omi contains the name of the input file
  1 ---> Meteo.cnta
  2 ---> FauLog
```

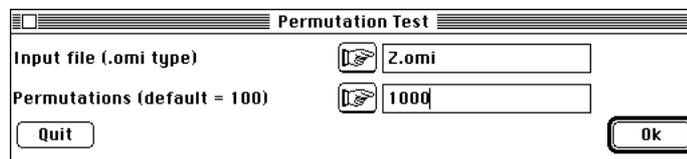
```
-----
Niche parameters
InerO = inertia
OMI = outlying mean index
T1 = outlying variance index
T2 = tolerance index
z* = 1000*z/InerO
```

| Esp | InerO | OMI | T1 | T2 | d2* | T1* | T2* |
|-----|-----------|-----------|-----------|-----------|-----|-----|-----|
| 1 | 6.434e+00 | 2.773e+00 | 1.021e+00 | 2.639e+00 | 431 | 159 | 410 |
| 2 | 1.191e+01 | 4.449e+00 | 2.388e+00 | 5.078e+00 | 373 | 200 | 426 |
| 3 | 1.057e+01 | 9.549e-02 | 2.539e+00 | 7.940e+00 | 9 | 240 | 751 |

Pour chaque espèce, on a un descriptif de la niche conformément aux définitions précédentes. Iner0 désigne $I_T(j)$, OMI est la marginalité $m_a(j)$, T1 la tolérance marginale $T_m(j)$ et T2 la tolérance résiduelle $T_r(j)$ avec $I_T(j) = m_a(j) + T_m(j) + T_r(j)$. Les mêmes paramètres sont exprimés en pourcentage de la somme. L'espèce 3 a un centre de gravité proche de l'origine et l'essentiel de son inertie est de la tolérance. L'espèce théorique uniformément répartie a une inertie égale à l'inertie totale soit 11, le nombre de variables de l'ACP normée.

| | | | | | | | |
|-----|-----------|-----------|-----------|-----------|-------|-------|-------|
| 4 | 7.626e+00 | 6.304e-01 | 7.349e-01 | 6.261e+00 | 83 | 96 | 821 |
| 5 | 1.047e+01 | 4.345e-01 | 3.924e+00 | 6.112e+00 | 41 | 375 | 584 |
| 6 | 7.431e+00 | 1.291e+00 | 1.551e+00 | 4.589e+00 | 174 | 209 | 618 |
| 7 | 1.436e+01 | 6.177e+00 | 4.759e+00 | 3.424e+00 | 430 | 331 | 238 |
| 8 | 1.124e+01 | 1.797e+00 | 2.765e+00 | 6.682e+00 | 160 | 246 | 594 |
| ... | | | | | | | |
| 16 | 1.225e+01 | 4.598e+00 | 3.518e+00 | 4.137e+00 | 375 | 287 | 338 |
| 17 | 9.392e+00 | 5.887e-01 | 2.523e+00 | 6.280e+00 | 63 | 269 | 669 |
| --- | ----- | ----- | ----- | ----- | ----- | ----- | ----- |

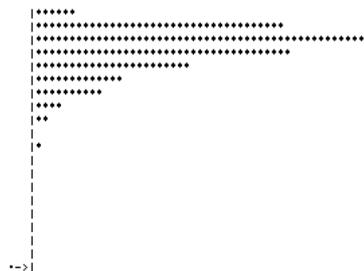
La signification statistique de la marginalité est testée par un test de randomisation (Niche : Permutation Test) :



Environmental parameter : Meteo.cnta
Faunistic array : FauLog

| Esp | Index | Eff(X>x0) | Probax10000 |
|-----|-----------|-----------|-------------|
| 1 | 2.773e+00 | 509 | 5090 |
| 2 | 4.449e+00 | 249 | 2490 |
| 3 | 9.549e-02 | 261 | 2610 |
| 4 | 6.304e-01 | 871 | 8710 |
| 5 | 4.345e-01 | 0 | 0 |
| 6 | 1.291e+00 | 3 | 30 |
| 7 | 6.177e+00 | 1 | 10 |
| 8 | 1.797e+00 | 69 | 690 |
| ... | | | |
| 15 | 2.900e+00 | 0 | 0 |
| 16 | 4.598e+00 | 13 | 130 |
| 17 | 5.887e-01 | 5 | 50 |

9 des 17 espèces ont une niche significativement excentrée. La météorologie a une action significative sur plus de la moitié des taxa. Il est donc prévisible que le test global sur la marginalité moyenne des espèces sera fortement significatif, ce qui est clair :



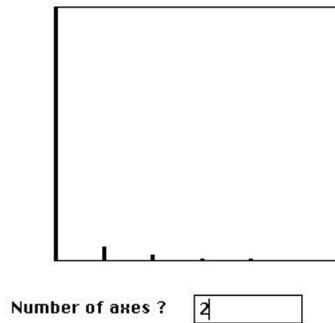
X = d2 = Total inertia of the set of gravity center
number of random matching: 1000 Observed: 1.396367

Histogramm: minimum = 0.152167, maximum = 1.396367
 number of simulation X<Obs: 1000 (frequency: 1.000000)
 number of simulation X>=Obs: 0 (frequency: 0.000000)

A partir du moment où la signification statistique est assurée un bilan global de la position des niches est utile (Niche : OMI Analysis) :



On garde deux facteurs pour la discussion, mais il est clair que la structure est unidimensionnelle:



Environmental data: Meteo.cnta
 Sites: 49 Variables: 11
 Species abundance: FauLog
 Sites: 49 Species: 17

L'analyse porte le titre na. Le tableau a autant de lignes que d'espèces et autant de colonnes que de variables :

Mean position of each species on each variable in the file Z.nata
 Row=species: 17 Col=variable: 11

Les poids des lignes dérivent de la marge du tableau faunistique :

Weights of species in file Z.napl
 Row=species: 17 Col=1

Les poids des colonnes sont ceux de l'ACP normée de départ (pondération unitaire) :

Weights of variables in file Z.napc (from input environmental triplet)
 Row=variable: 11 Col=1

Le triplet statistique subit le traitement habituel :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: Z.nata
--- Number of rows: 17, columns: 11
-----
Total inertia: 1.39637
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +1.2667E+00 +0.9071 +0.9071 | 02  +6.8793E-02 +0.0493 +0.9564 |
03  +3.3103E-02 +0.0237 +0.9801 | 04  +1.1567E-02 +0.0083 +0.9884 |
05  +7.2774E-03 +0.0052 +0.9936 | 06  +3.3927E-03 +0.0024 +0.9960 |
07  +2.3760E-03 +0.0017 +0.9977 | 08  +1.8243E-03 +0.0013 +0.9990 |
09  +6.8819E-04 +0.0005 +0.9995 | 10  +4.9675E-04 +0.0004 +0.9999 |
11  +1.9657E-04 +0.0001 +1.0000
```

File Z.napv contains the eigenvalues and relative inertia for each axis
 --- It has 11 rows and 2 columns

90% de la marginalité totale s'exprime sur un seul axe et le nuage des positions des taxons doit être simple. Sur les axes d'inertie sont projetés les relevés (-.nals) et les taxons (-.nali). Ces noms de fichiers se comprennent par référence à la théorie. Les centres des espèces (lignes du tableau traité -.nata) donnent les coordonnées de référence des lignes (-.nali) tandis que les relevés sont projetés en individus supplémentaires (-.nals) :

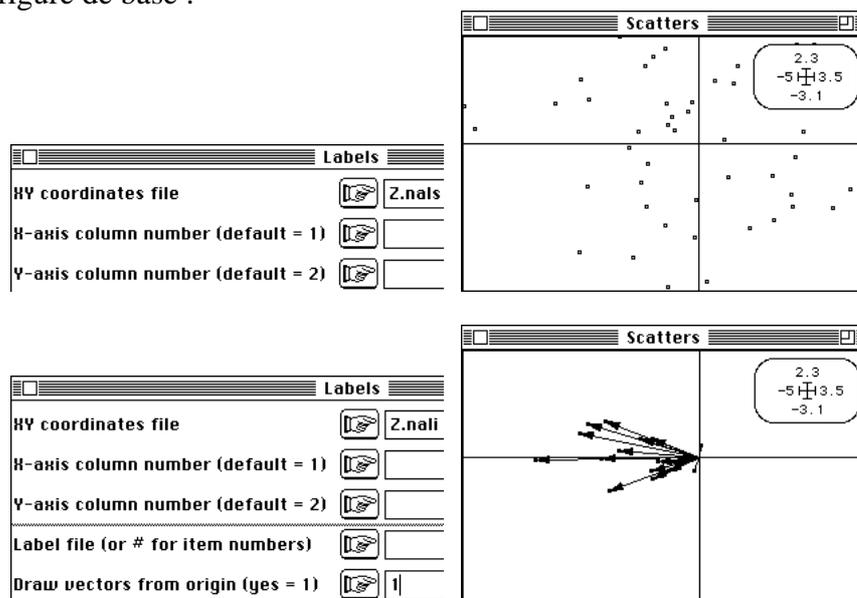
File Z.nali contains the row scores
--- It has 17 rows and 2 columns

```
File :Z.nali
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1  |-3.448e+00| 2.617e-02|
|  2  |-6.996e-01| 7.796e-01|
|----|-----|-----|
```

File Z.nals contains the coordinates of the sites
It has 49 rows and 2 columns

```
File :Z.nals
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1  |-4.947e+00| 3.496e+00|
|  2  |-3.040e+00| 2.287e+00|
|----|-----|-----|
```

D'où la figure de base :



Sur le plan d'inertie sont également projetés les axes de l'analyse du nuage des relevés (ACP normée) :

File Z.naax contains the coordinates of the sites
It has 2 rows and 2 columns

```
File :Z.naax
|Col.|   Mini   |   Maxi   |
|----|-----|-----|
|  1  |-8.680e-01| 4.613e-01|
|  2  | 2.822e-01| 6.014e-01|
|----|-----|-----|
```



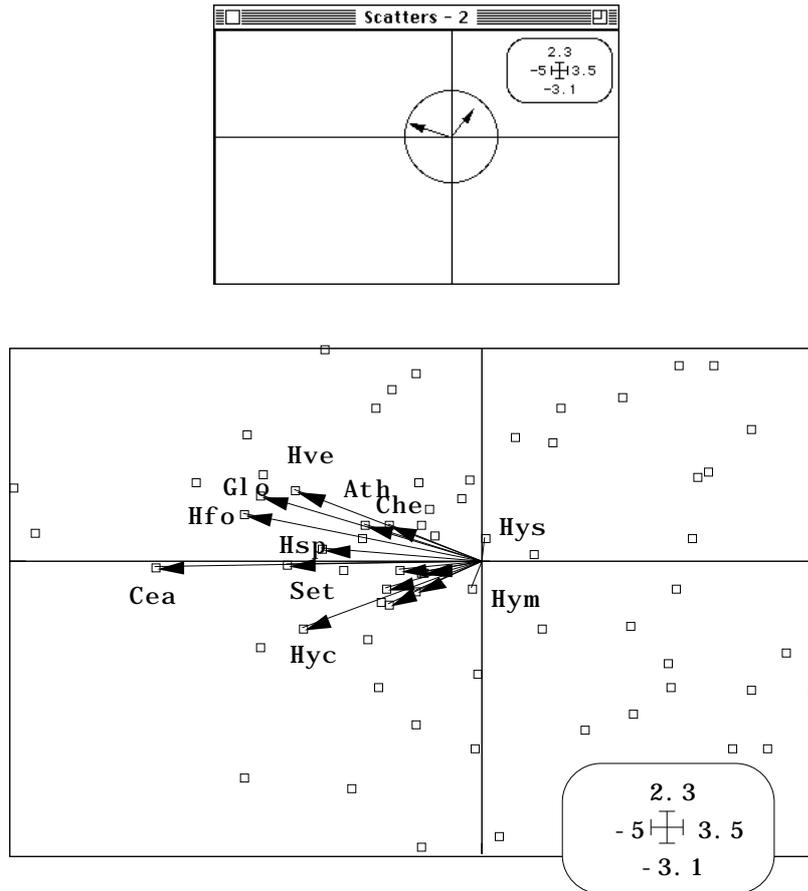


Figure 11 : Projections en analyse OMI. Les taxons (flèches) sont à la moyenne des relevés (carrés) pour leur distribution propre.

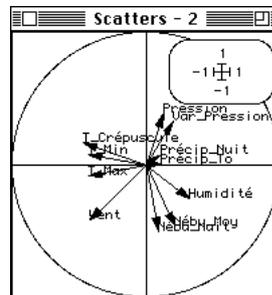
Sur le plan d'inertie sont également projetés les vecteurs de la base canonique :

File Z.nac1 contains the column scores with unit norm
It has 11 rows and 2 columns

File :Z.nac1

| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -4.651e-01 | 3.236e-01 |
| 2 | -5.011e-01 | 3.825e-01 |

| Correlation circle | |
|------------------------------------|-----------|
| XY coordinates file | Z.nac1 |
| X-axis column number (default = 1) | |
| Y-axis column number (default = 2) | |
| Label file (or # for item numbers) | Label_Var |



Enfin, sont données les coordonnées des variables de l'analyse OMI qui sont ici des covariances entre les scores des relevés et les variables initiales :

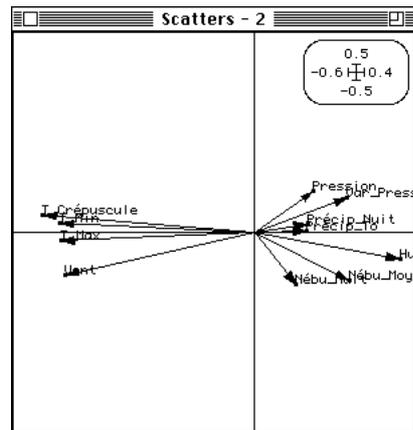
File Z.naco contains the column scores
--- It has 11 rows and 2 columns

File :Z.naco

| Col. | Mini | Maxi |
|------|------|------|
| | | |

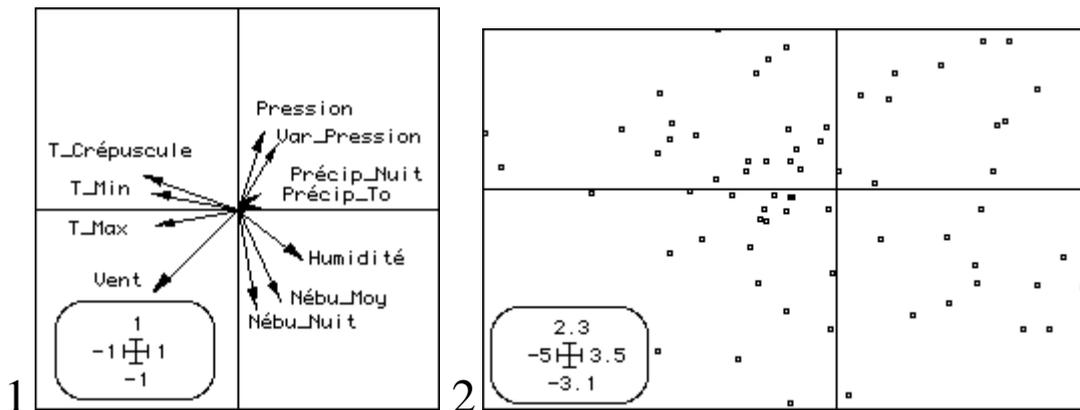
| | | |
|---|------------|-----------|
| 1 | -5.235e-01 | 3.642e-01 |
| 2 | -1.314e-01 | 1.003e-01 |

| Labels | |
|------------------------------------|-----------|
| XY coordinates file | Z.naco |
| X-axis column number (default = 1) | |
| Y-axis column number (default = 2) | |
| Label file (or # for item numbers) | Label_Var |
| Draw vectors from origin (yes = 1) | 1 |

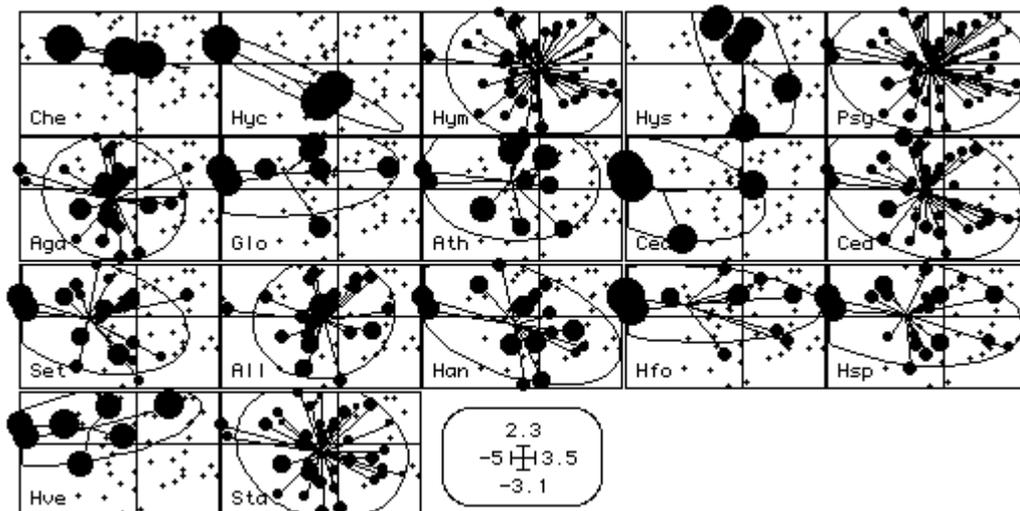


Utiliser ScatterDistri pour illustrer la notion d'averaging. Le fonctionnement de l'analyse peut s'écrire :

- 1 — On donne des poids aux espèces de somme de carrés unitaire (poids canonique).
- 2 — Chaque relevé est positionné par les scores (combinaisons des observations normalisées avec les poids canoniques).



3 — Chaque espèce prend une position moyenne :

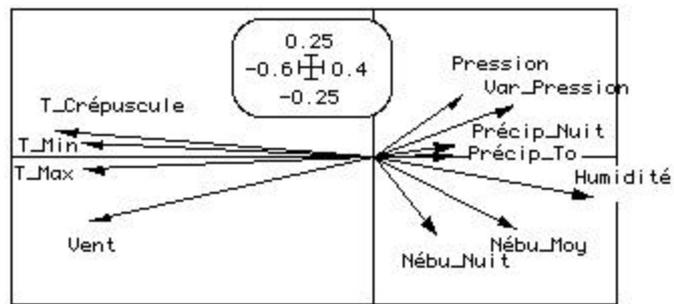
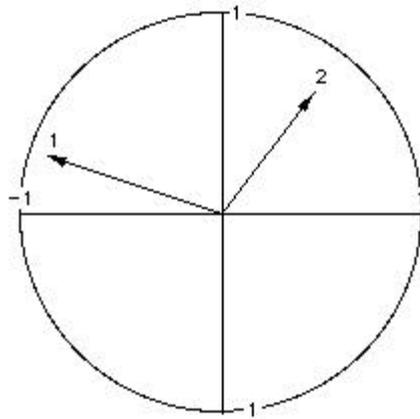
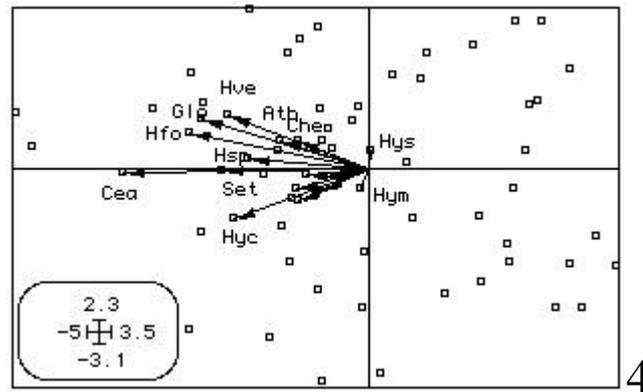


3

4 — On a ainsi maximisé la somme des carrés des distances à l'origine des positions moyennes des espèces.

5 — On peut rapprocher structure du nuage des positions moyennes des espèces et structure du nuage des points initiaux.

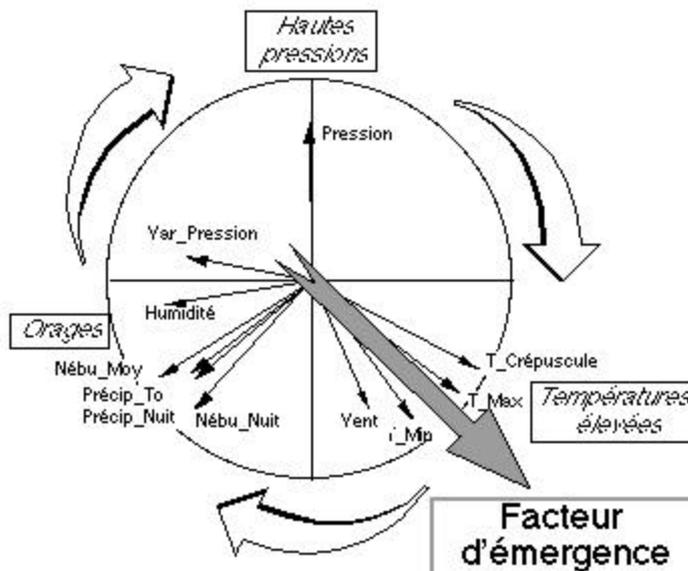
6 — On peut illustrer la position des variables par leurs covariances avec les scores. Par normalisation, on retourne en 1 (méthode aux valeurs propres).



5

6

On est conduit à l'interprétation sans ambiguïté qui se résume par :



Le fait décrit est fortement significatif mais le facteur d'émergence a un effet très variable suivant les espèces. Les espèces les plus proches de l'origine ne sont pas

concernées, les plus éloignées de l'origine sont les plus sensibles. On a ici un cas typique de facteur écologique favorisant une partie des taxons.

4 — Une nouvelle méthode d'ordination

Une question se pose alors clairement. Puisqu'il est possible de trouver une combinaison de variables de milieu qui optimise la marginalité moyenne des taxons, existe-t-il une méthode simple d'ordination qui utilise le même critère ?

La réponse est fort simple. Soit \mathbf{Y} un tableau floro-faunistique à n relevés (lignes) et t colonnes (taxons). Soit \mathbf{F} le tableau des profils colonnes associé (somme des colonnes unité) et \mathbf{D}_t la diagonale des poids des colonnes comme précédemment. Soit \mathbf{D}_n la matrice diagonale de la pondération uniforme des lignes et \mathbf{U}_{nt} la matrice à n lignes et t colonnes dont tous les éléments sont égaux à 1.

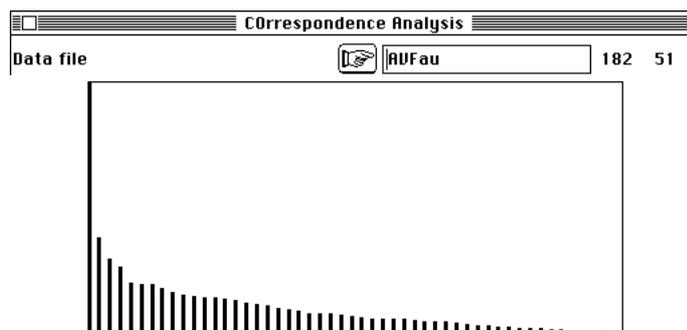
L'analyse du triplet $\mathbf{F} - \frac{1}{n} \mathbf{U}_{nt} \mathbf{D}_t \mathbf{D}_n$ donne des scores des relevés centrés et normés pour la pondération uniforme (composantes principales orthonormées) qui maximisent la marginalité moyenne des espèces. Il s'agit d'une ACP centrée des profils espèces pondérée comme une AFC pour les colonnes. Elle a la propriété d'averaging de l'AFC (espèces à la moyenne des relevés) mais l'origine est placée à la distribution uniforme et le point de référence est alors l'*espèce ubiquiste*, celle qui est uniformément présente dans tous les relevés. Elle a une propriété d'ACP centrée (optimisation de la somme des carrés des covariances avec les scores des relevés). Par contre, comme la covariance ordinaire de l'ACP centrée sur le tableau faunistique brut est égale à l'abondance de l'espèce multipliée par sa marginalité, elle limite fortement l'importance écrasante des espèces abondantes dans une ACP sur tableau faunistique.

La totalité des aides à l'interprétation des analyses d'inertie (décomposition de l'inertie des nuages, reconstitution des tableaux par le théorème d'Eckart-Young²⁰, Biplot lignes-colonnes^{21 22} sont immédiatement disponibles par les propriétés générales des schémas de dualité¹⁴. L'utilisation d'un modèle mathématique souhaitée par²³ (qui aurait pu permettre à Orloci²⁴ d'éviter d'appeler Analyse de la concentration —AOC—²⁵ ce qui est strictement une AFC) est ici centrale.

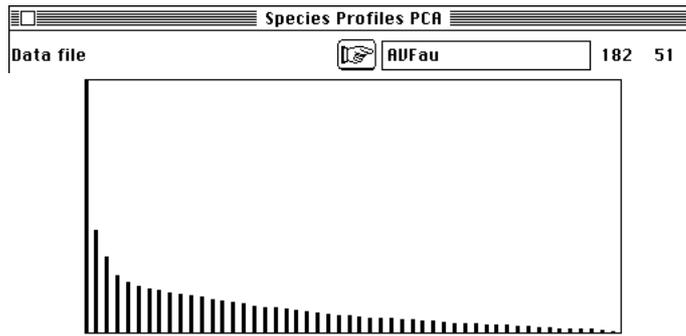
Il est intéressant de voir l'effet de cette simple modification du triplet sur quelques exemples bien étudiés et de confronter AFC, ACP et cette version qu'on appellera ACP des profils spécifiques. La méthode est disponible dans le module Niche.

4.1 — Gradients d'avifaune

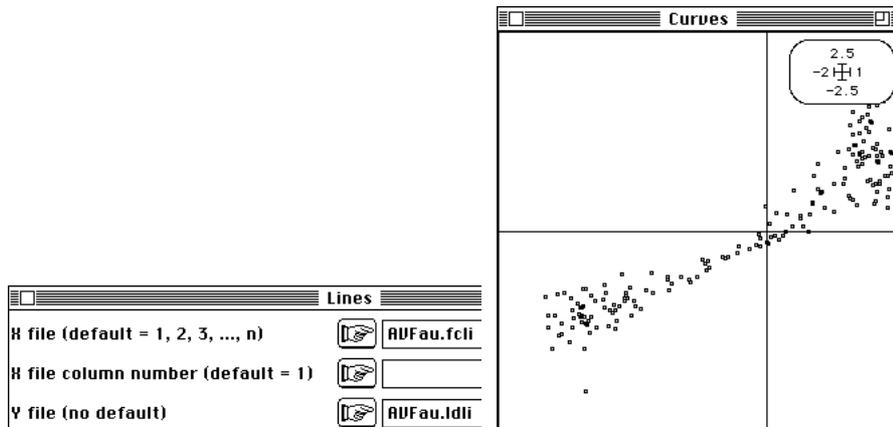
On utilise les données de Prodon et Lebreton²⁶ qui proposent 186 relevés sur 51 espèces d'Oiseaux. L'AFC présente une première valeur propre de grande importance :



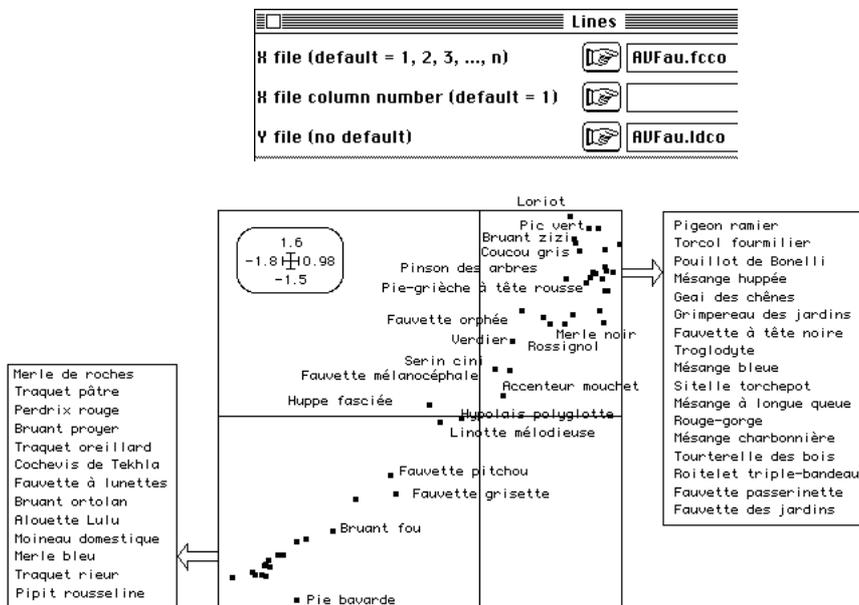
L'ACP sur profils fait de même :



Les deux ordinations ont des facteurs 1 voisins sans être identiques :



Les interprétations en terme d'espèces sont très voisines :



AFC et ACP sur profils fonctionnent de la même manière. Seule la contrainte de centrage pour la pondération moyenne (AFC) et pour la pondération uniforme (ACP) introduit une nuance sensible. On observera une sensibilité moindre à l'effet Guttman dans la seconde que dans la première (figure 11).

Un autre exemple est proposé par les données de Blondel et Frochot ²⁷. 302 relevés répartis en 12 catégories (2 régions et 6 strates) portent sur 60 espèces. La figure 12 donne les plans 1-2 des AFC et ACP sur profils que nous appellerons désormais PACP.

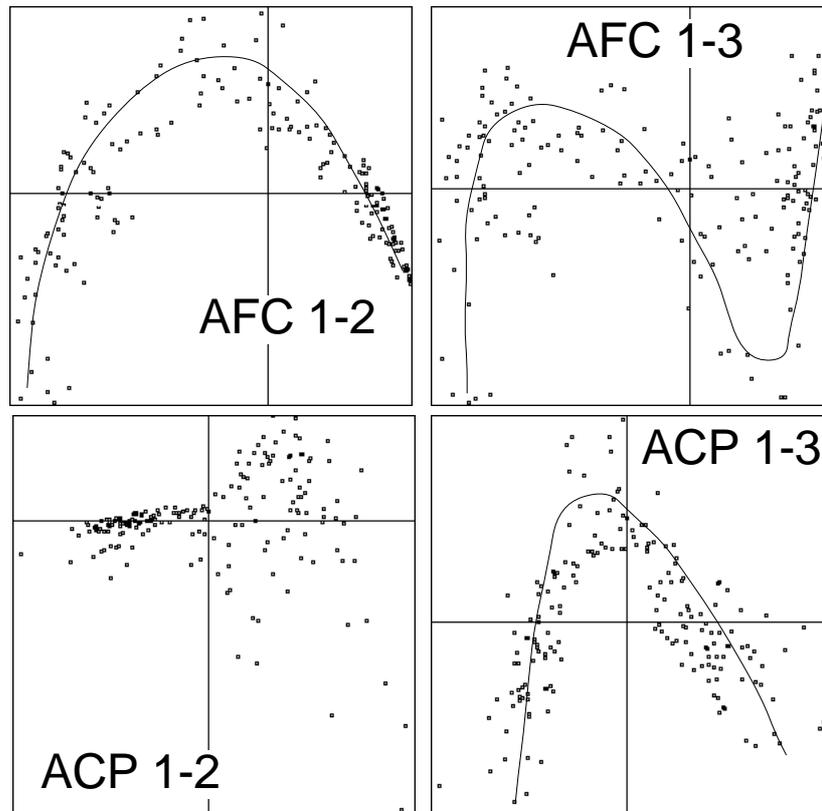


Figure 11 : Plans factoriels des relevés en AFC et ACP sur profils sur un gradient d'avifaune.

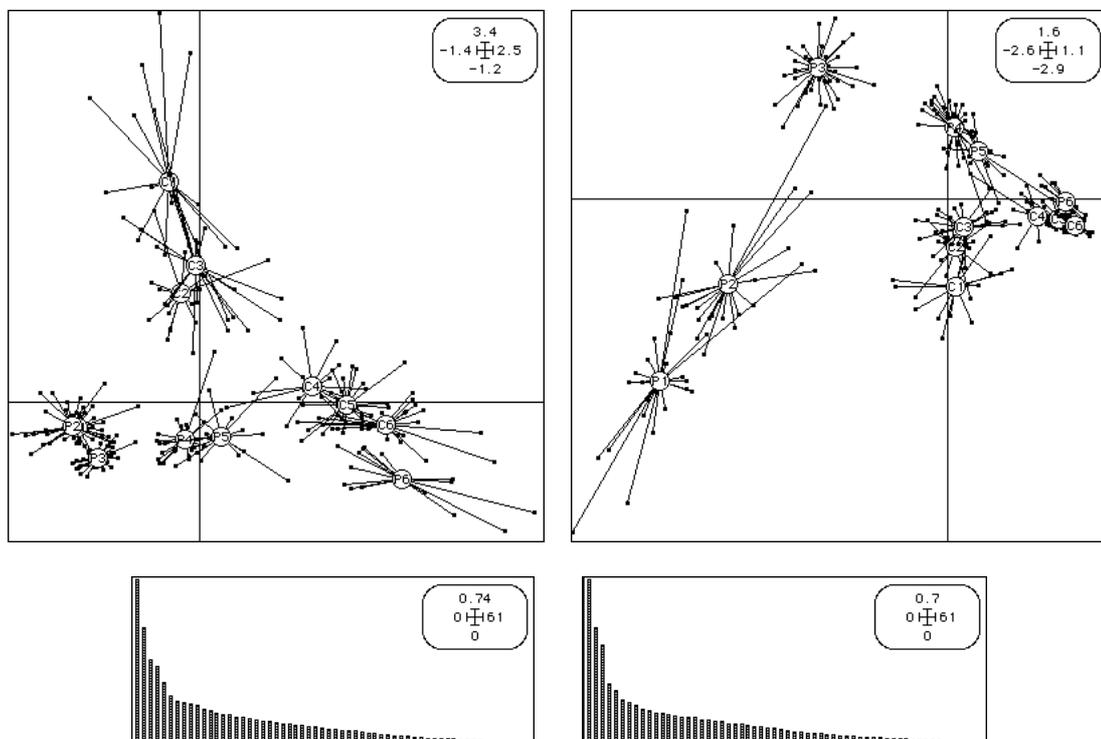
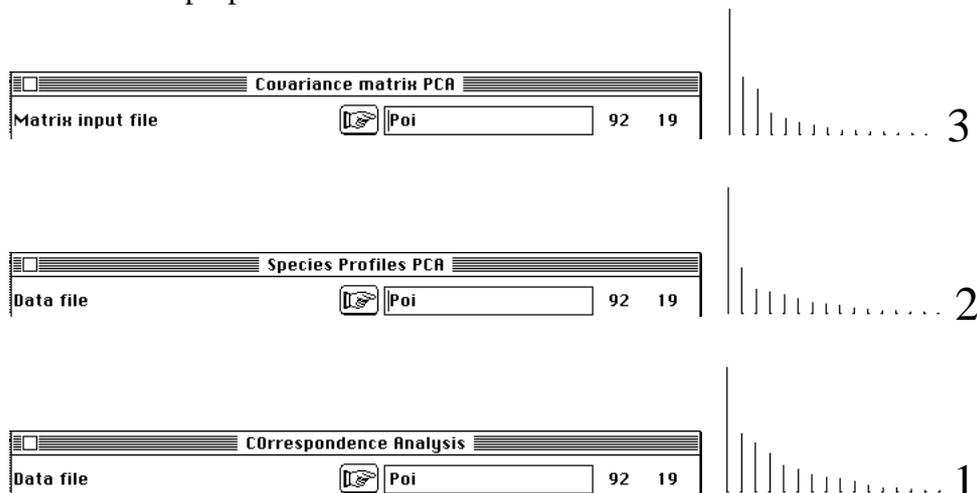


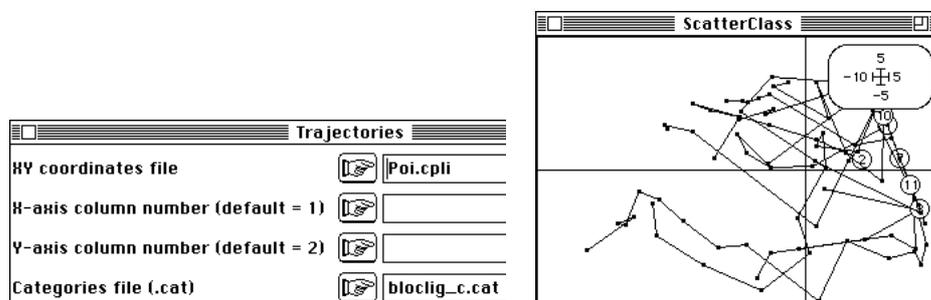
Figure 12 : Plans factoriels 1-2 des relevés en AFC (à droite) et PACP à gauche pour 302 relevés d'avifaune répartis en 12 classes. Le double centrage en AFC est une contrainte forte.

4.2 — Gradients Amont-Aval

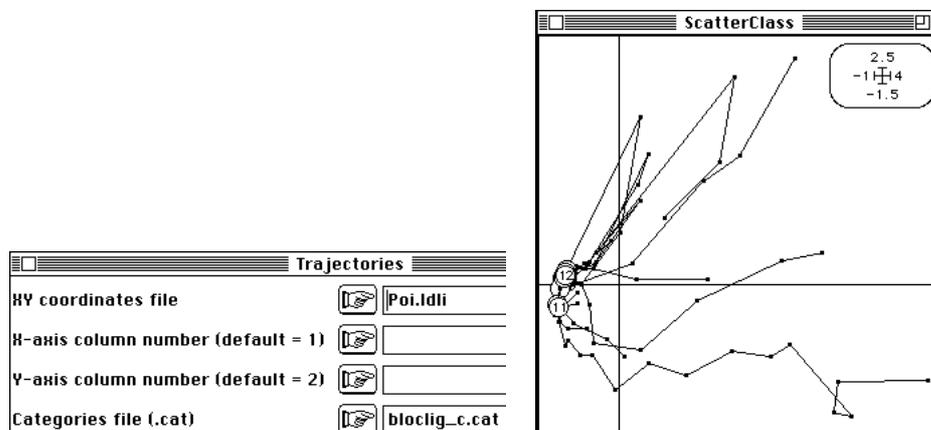
Si l'ordination des tableaux d'avifaune justifie en général une AFC ²⁸, l'usage de l'ACP classique en hydrobiologie reste très répandu. Les données de Verneaux ²⁹ forment un ensemble de référence en ichtyologie. 92 relevés répartis sur 12 rivières portent sur 19 taxons. La représentation des stations par rivières permet ici de tester les méthodes dans leur restitution de la structure faunistique longitudinale omniprésente dans ce type de données ³⁰. Les trois analyses ne donnent pas les mêmes indications pour les valeurs propres :



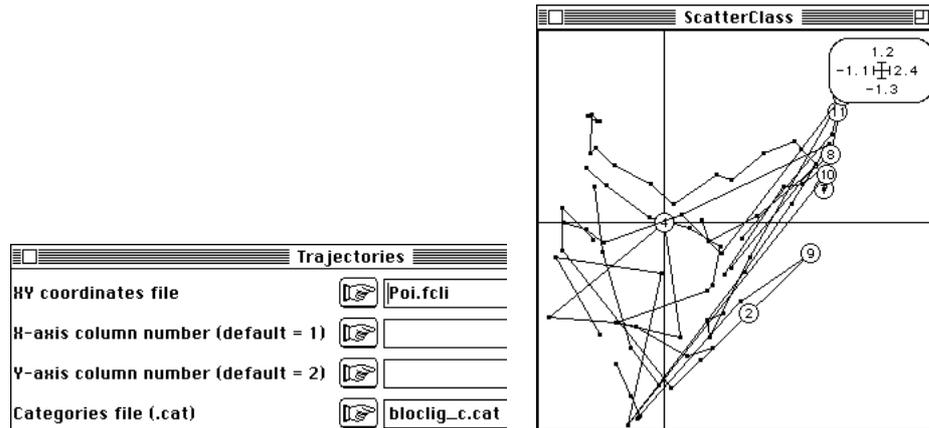
Si on compare les trois plans 1-2, les différences sont sensibles. L'ACP donne une image d'évolution parallèle de la faune le long des rivières :



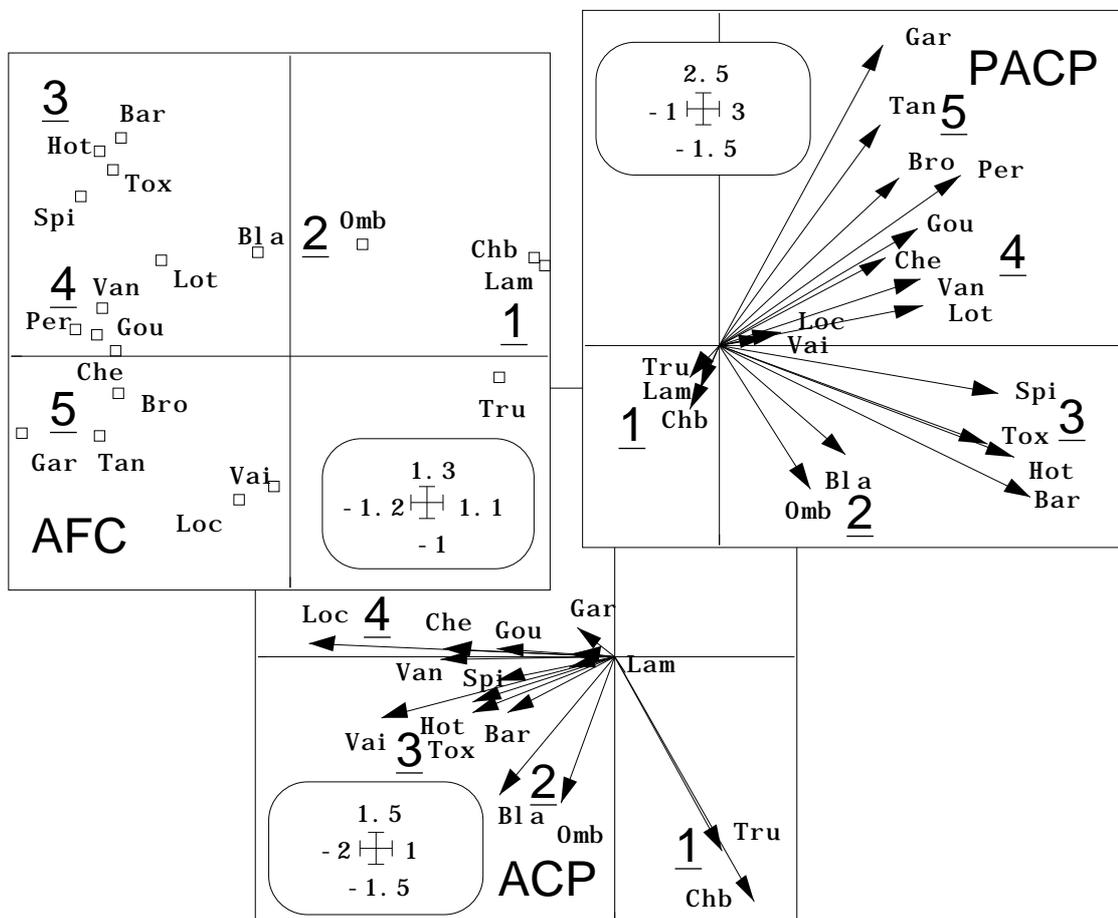
La PACP donne une image d'évolution divergente d'amont en aval :



L'AFC donne encore une image différente (ci-dessous).



Les cartes des espèces présentent également, en première lecture, des différences sensibles :



Les mêmes associations sont pourtant repérées : seule la position du couple Vairon-Loche est très instable. L'effet taille (augmentation simultanée de l'abondance de plusieurs espèces est saisi de multiple manière. Les contraintes sur les centrages sont également très différentes.

La superposition espèces-stations est particulièrement efficace en PACP (figure 13). On y voit clairement que le haut Doubs passe directement de la zone à Truite à la zone à Brème comme le Drugeon et la Clauge tandis que le Doubs moyen et la Loue présentent la succession Truite-Ombre-Barbeau. Le groupe Vairon-Loche a la distribution la plus large.

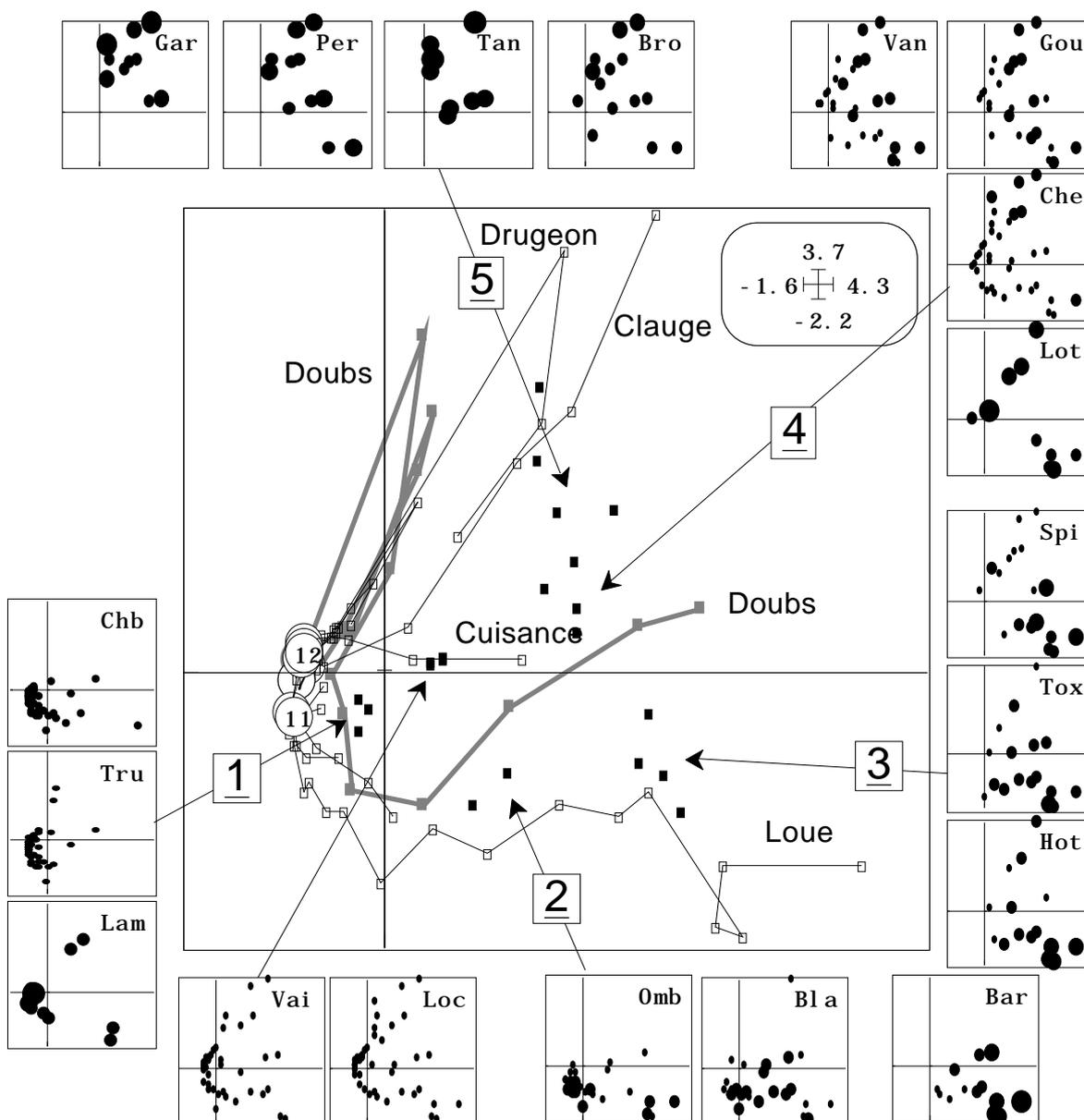


Figure 13 : Superposition stations-taxons par averaging dans un tableau ichthyologique et description des amplitudes d'habitat par PACP On utilise les fichiers .ldl1 (positions des stations par trajectoires), .ldco (position des espèces par averaging) et le fichier de départ pour illustrer les dispersions (module ScatterDiscri).

Il y a donc deux rivières de référence (Loue et Clauge) mais confusion des zones aval pour récupérer le compromis de STATIS (fiche thématique 6-2). Des groupes d'espèces peuvent donc avoir des fonctions typologiques identiques (opposition amont-aval) en étant totalement disjointes. La complémentarité de l'approche globale et de l'approche K-tableaux est ici exemplaire.

On retient alors que l'interprétation des structures faunistiques par position de niches spécifiques est plus efficace dès que la carte relevés a une interprétation en elle-même. Dans ce cas le couple AFC-PACP qui la supporte l'emporte sans conteste sur la lecture par covariances. Restent alors des différences de contraintes de fonctionnement qui peuvent avoir un rôle non négligeable.

5 — Analyse Canonique de la marginalité

Nous revenons au couplage de tableaux écologiques sur la base de l'optimisation de la marginalité. La méthode OMI est une méthode de couplage basée sur le schéma $(F^t X, Q, D_t)$ où F est le tableau des profils (colonnes = taxons), X le tableau de milieu, Q la norme utilisée dans l'analyse préliminaire du triplet (X, Q, D_n) et D_t les poids des taxons.

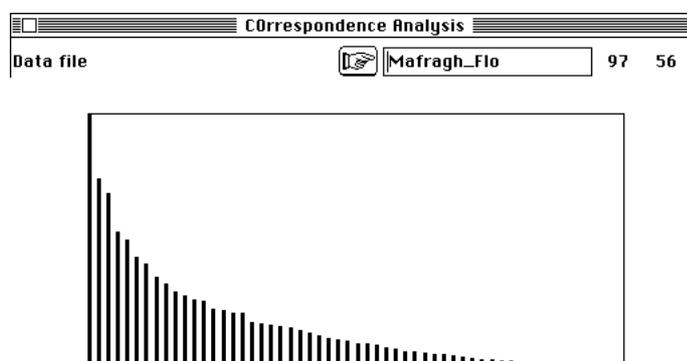
Ceci nous a amené à définir l'analyse du triplet $F - \frac{1}{n} U_{nt}, D_t, D_n$ que nous avons appelé PACP, méthode intermédiaire entre AFC et ACP. Il existe donc une analyse de co-inertie entre la PACP du tableau floro-faunistique et l'analyse de départ puisque les triplets (X, Q, D_n) et $F - \frac{1}{n} U_{nt}, D_t, D_n$ sont correctement appariés par les lignes

(relevés). Cette analyse est basée sur le triplet $F - \frac{1}{n} U_{nt} \quad D_n X, Q, D_t$. Comme D_n est la pondération uniforme, on retrouve au coefficient n près l'analyse OMI de départ. L'analyse OMI est donc une analyse de co-inertie.

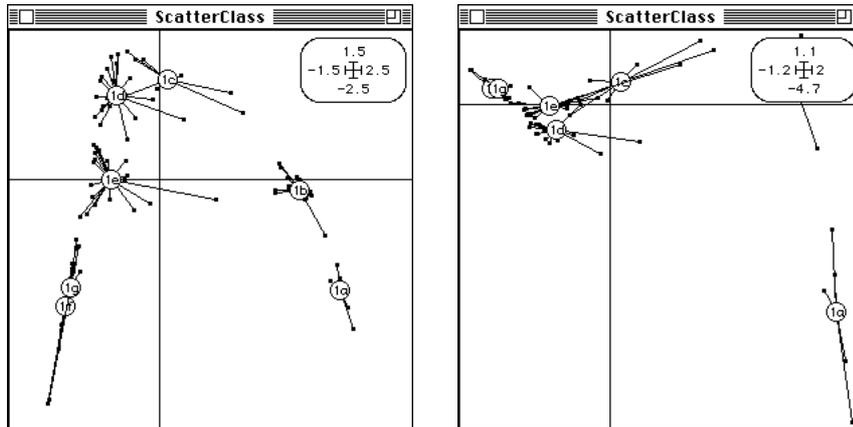
Dans la PACP on cherche des scores des relevés centrés réduits pour optimiser la marginalité moyenne des espèces, dans l'analyse OMI on cherche des combinaisons de variables de milieu pour faire de même. On a donc intérêt à chercher les axes de marginalité ayant des bonnes propriétés d'inertie (forte variance des scores) et des bonnes propriétés de marginalité (éloignement des centres de gravité des espèces de l'origine).

On se retrouve exactement dans la discussion entre ACC et Co-Inertie. Il y a nécessairement une analyse qui cherche des combinaisons de variables de milieu centrées réduites qui optimise la marginalité, ce qui est la version ACC de l'analyse OMI. On peut appeler cette procédure Analyse canonique OMI. Elle est basée sur le triplet $F^t X, (X^t D_n X)^-, D_t$ où on introduit la norme de Mahalanobis (au sens large, voir fiche thématique 5-3).

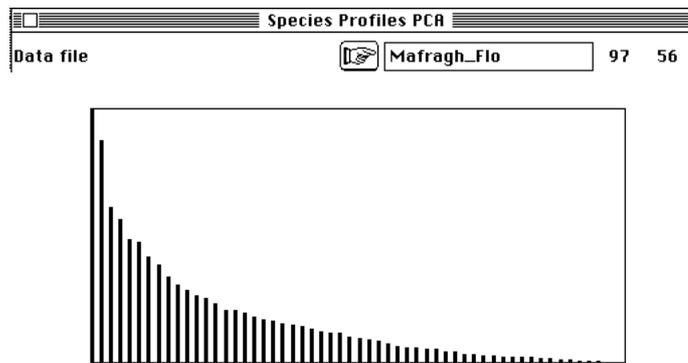
Cette analyse est programmée dans le module Niche comme les précédentes. Un exemple regroupe les trois méthodes. Les données de de Bélair ³¹ éditées dans ³² propose 97 relevés de végétation accompagnés de 97 relevés de milieu. Comparons une dernière fois AFC et PACP.



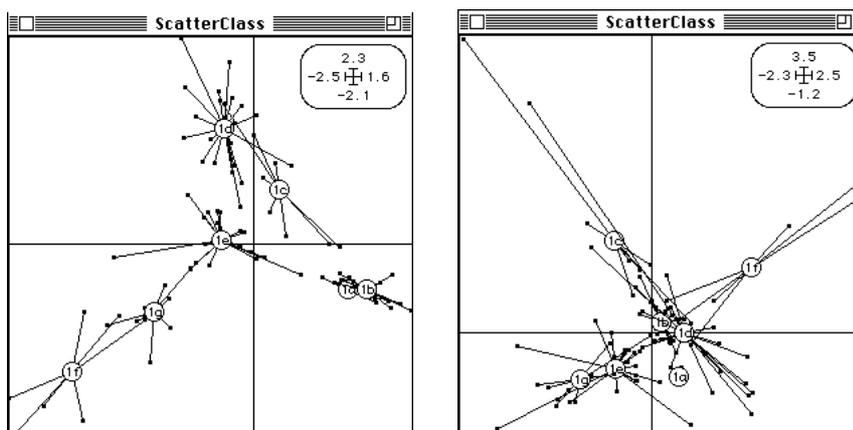
L'analyse déjà interprétée par les auteurs fournit 7 groupes de relevés (individus d'association ou groupe écologique). Cette partition est reportée sur la carte factorielle (plans 1-2 et 1-3) :



L'effet Guttman est très prononcé. Le groupe a est fortement identifié. Le même tableau donne :



On utilise les plans 1-2 (à gauche) et 3-4 (à droite). Les différences sont évidentes. Les groupes f et g indissociables en AFC sont bien isolés et le groupe a semble difficile à distinguer du groupe b.



Les groupes a et b sont des faciès de l'association *Scirpetum maritimi*. Les autres unités appartiennent à l'association *Suaedetum fruticosae* : c est dans la sous-association à *Salicornia fruticosa* et les unités d e f et g sont des groupes floristiques de la sous-association à *Inula crithmoides*. Il est curieux que les discriminations a contre b et f contre g soient exprimées de manière si différentes. Ce sont aux extrémités des gradients d'ordination que les contraintes numériques sont les plus fortes.

Les cartes des coordonnées factorielles sont tracées dans la figure 14. Le tableau de milieu (variables édaphiques essentiellement) fait l'objet d'une ACP normée donnant un facteur 1 marqué (figure 15).

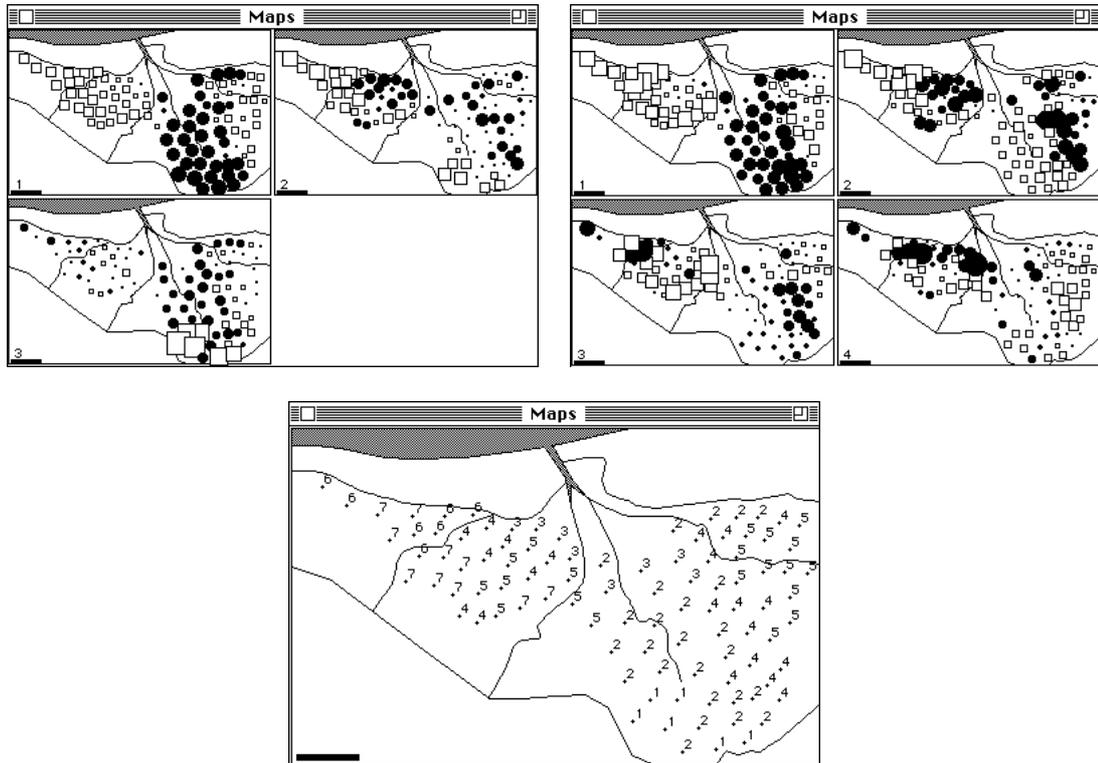
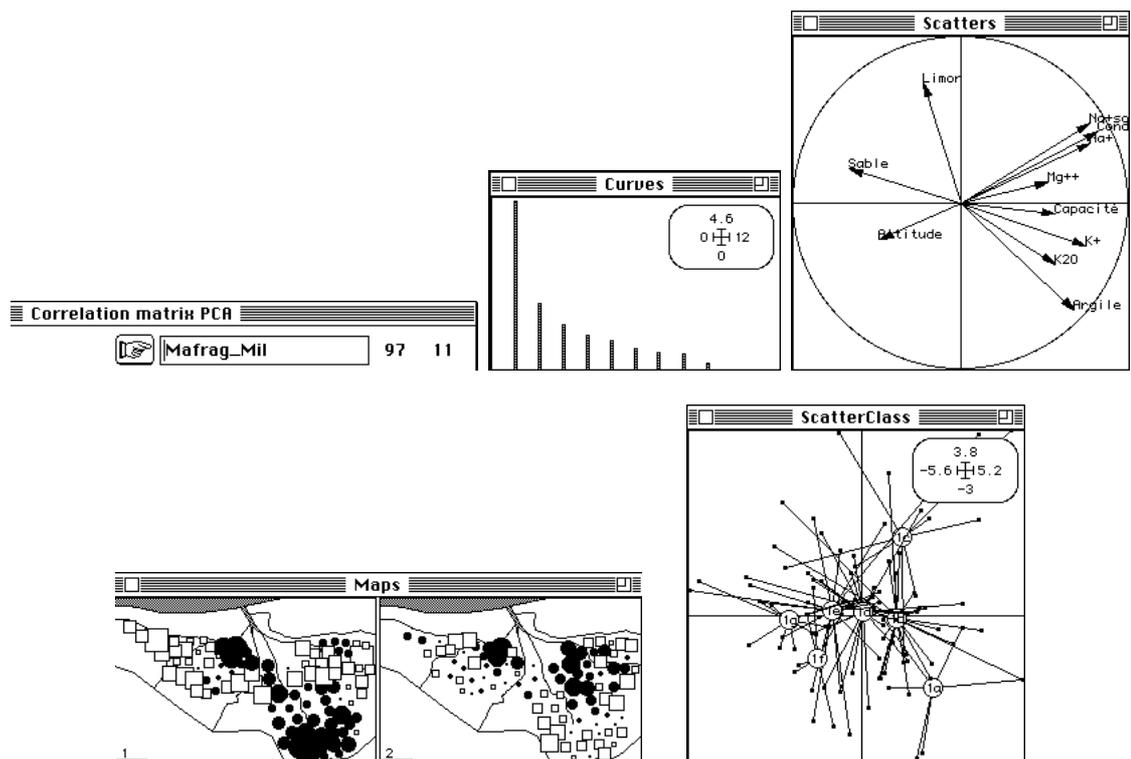
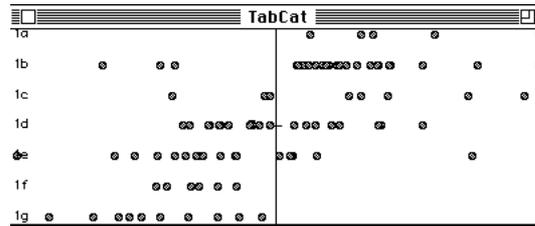


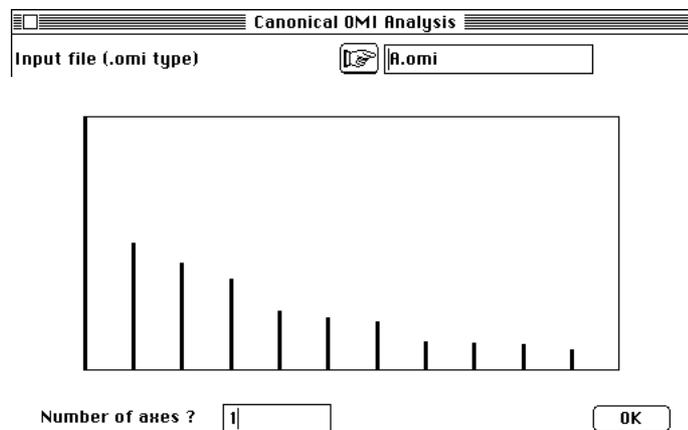
Figure 14 : Cartographie de coordonnées factorielles (à gauche 3 facteurs d'AFC, à droite 4 facteurs de PCAP, en dessous la partition retenue par les auteurs avec 1 pour a, 2 pour b, ...). Des cartes de codes numériques orthogonaux même en nombre réduit sont difficiles à synthétiser.



Les espèces s'ordonnent sur un gradient de salinité (croissant avec l'altitude décroissante). Le tableau de milieu valide le modèle du gradient sans retrouver trace de l'approche phyto-sociologique :



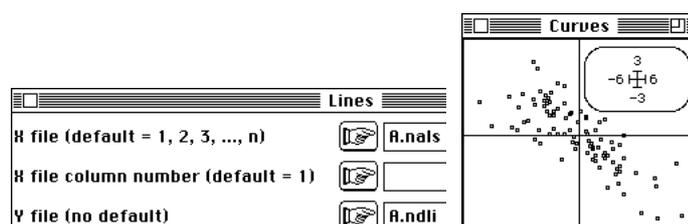
En effet, sur le gradient, les unités végétales s'ordonnent simplement sans rupture. On peut supposer que la corrélation initiale entre variables de milieu impose ce point de vue. On peut donc décorréler les variables. C'est l'objet soit de l'ACC (version AFC) soit de la discrimination OMI (version PACP).



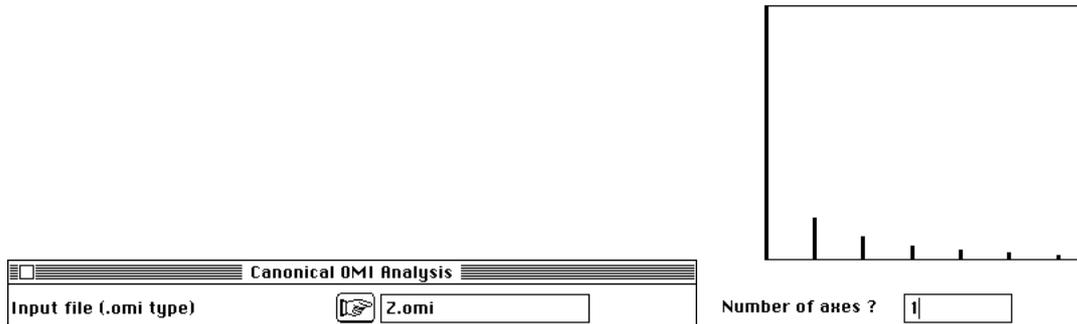
On obtient un codage numérique des relevés de variance unité. Ce code numérique est une combinaison de variables normalisées calculée à l'aide des poids canoniques (à gauche) et présentant une certaine corrélation avec les variables (à droite) :

| | | |
|---------|------------------|----------|
| - 2. 47 | Argil e | - 0. 310 |
| - 1. 88 | Li mon | 0. 107 |
| - 2. 05 | Sabl e | 0. 229 |
| 0. 02 | K20 | - 0. 434 |
| 0. 11 | Mg++ | - 0. 297 |
| 0. 08 | Na+ | 0. 600 |
| - 0. 78 | K+ | - 0. 798 |
| - 0. 58 | Conductivité | - 0. 608 |
| 0. 10 | Capacité de rét. | - 0. 263 |
| 0. 30 | Na+sol | - 0. 552 |
| 0. 32 | Altitude | 0. 642 |

Comme d'habitude une incohérence numérique se fait jour et on retrouve tous les problèmes liés aux stratégies de l'inversion de norme (régression multiple contre régression PLS, analyse discriminante contre inter-classe, ACPVI contre co-inertie, ...). Nous préférons donc la version simple en notant que les deux approches ne sont pas contradictoires :



On retrouvera la même difficulté dans l'exemple des piégeages lumineux (paragraphe 3, page 10) :

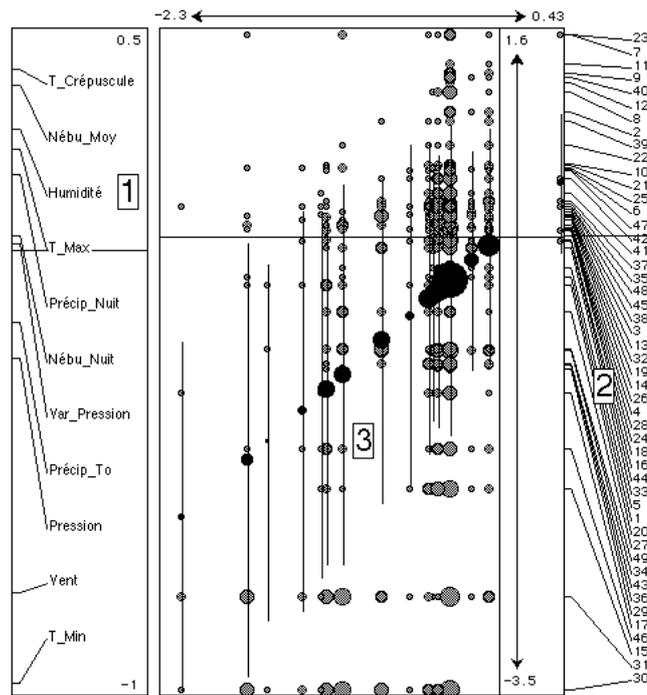


A gauche noter les poids canoniques et à droite les corrélations code-variables :

| | | |
|-----------|---------------|-----------|
| 0. 2277 | T_Max | - 0. 7002 |
| 0. 4097 | T_Crépuscul e | - 0. 7667 |
| - 0. 9758 | T_Min | - 0. 7188 |
| - 0. 7696 | Vent | - 0. 6902 |
| - 0. 2423 | Pressi on | 0. 2073 |
| 0. 0155 | Var_Pressi on | 0. 3260 |
| 0. 2749 | Humi di té | 0. 5390 |
| 0. 0333 | Nébu_Nui t | 0. 1458 |
| 0. 1716 | Préci p_Nui t | 0. 1889 |
| 0. 3737 | Nébu_Moy | 0. 3483 |
| - 0. 1618 | Préci p_Tot | 0. 1821 |

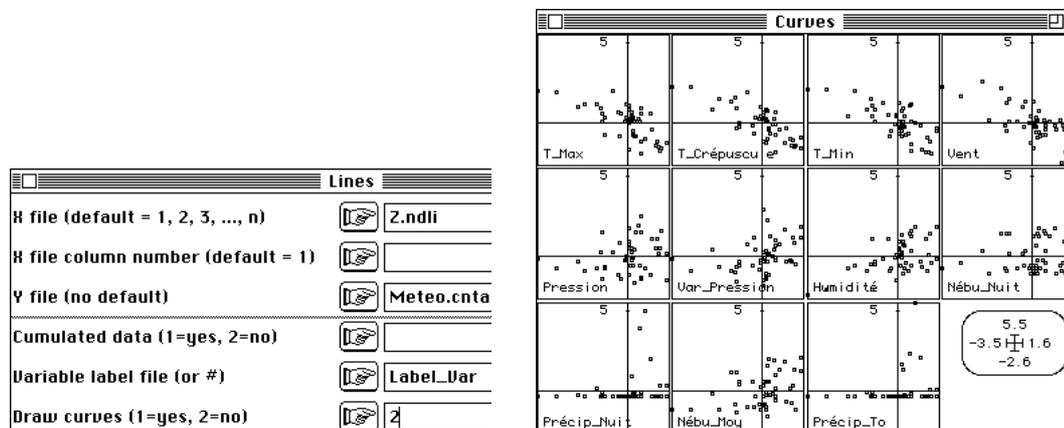
Si on accepte la difficulté, vue la forte cohérence sur deux variables, on peut illustrer le principe de cette démarche :

- 1 — On donne des poids aux variables (poids canoniques).
- 2 — Chaque relevé est positionné par un score (combinaisons des observations normalisées avec les poids canoniques) de variance unité.
- 3 — Chaque espèce prend une position moyenne sur ce score.



4 — On a ainsi maximisé la somme des carrés des distances à l'origine des positions moyennes des espèces. La quasi totalité des positions moyennes des espèces étant déportées du même côté (facteur favorable à l'émergence), on a un principe très proche de celui exprimé pour la CCA³³ qui serait ici, par définition, totalement prise en défaut.

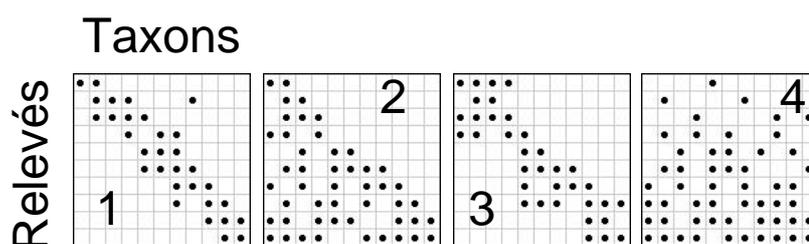
5 — On peut interpréter les scores par leur liaison avec les variables initiales.



6 — Conclusion

L'introduction d'une variante des méthodes de base n'est pas sans soulever un problème sérieux. Faut-il encore perturber l'utilisateur occasionnel par de nouveaux boutons ? La question renvoie à la fonction de la statistique, ici la statistique euclidienne destinée à voir des structures multidimensionnelles. Cet objectif n'est pas simple et l'énorme littérature biométrique sur l'ordination en témoigne. Faire des citations relève du tirage aléatoire.

C'est d'abord une question de modèles conceptuels. Il n'y a aucune méthode qui repère avec une même précision des organisations des types 1 (ordination des niches sur un gradient), 2 (ordination sur un facteur limitant), 3 (partition dans le modèle ohyto-sociologique) ou 4 (facteur d'abondance pur) :



Il y a une très forte contradiction entre la connaissance déjà ancienne de la généralité du modèle ACP et de la plasticité potentielle des paramètres d'une part, de la domination de deux programmes (ACP sur matrice de covariances et AFC) d'autre part. Les décompositions de l'inertie dans les ACP centrées par espèces (mode R) et par relevés (mode Q) avec ou sans normalisation sont explicites dans ³⁴ et l'effet du centrage discuté dans ³⁵. Neuf transformations préalables sont étudiées dans ³⁶

- aucune (*Unmodified data*)
- standardisation par espèces : l'abondance est exprimée en pourcentage de la valeur maximum observée pour une espèce (*Species maximum = 100*)
- fréquence par relevés : l'abondance est exprimée en pourcentage de la valeur totale observée pour un site (*Stand total = 100*)

- fréquence par relevés après standardisation par espèces (*Stand total = 100*)
after species maximum = 100)
- normalisation sans centrage par espèces (division par la longueur d'après Orloci 1967³⁷)
- centrage et normalisation par sites (*Standard deviation*)
- passage en présence/absence
- double centrage additif (*Deviation from expected value*)

Pour chacune de ces transformations, la matrice des produits scalaires est diagonalisée et le rapport des deux premières valeurs propres édité sur les cartes factorielles. Pour M.P. Austin, la notion de la généralité de l'ACP est clairement acquise et l'ordination inventée par les écologues (Méthode de Bray et Curtis³⁸) est abandonnée au profit de l'ACP dans³⁹.

Douze transformations initiales sont introduites dans la théorie générale de la diagonalisation des matrices $\mathbf{X}^t\mathbf{X}$ et $\mathbf{X}\mathbf{X}^t$ dans⁴⁰.

- aucune (*original data*)
- centrage par espèces (*centering by species mean*)
- centrage par relevé (*centering by site mean*)
- double centrage additif (*double centering by both species and site*)
- profils par espèces (*standardization by species total*)
- profils par sites (*standardization by site total*)
- normalisation sans centrage par espèces (*standardization by species norm*)
- normalisation sans centrage par sites (*standardization by site norm*)
- double normalisation sans centrage ($x_{ik} / \sqrt{\sum_i x_{ik}^2} \sqrt{\sum_k x_{ik}^2}$)
- centrage et normalisation par sites (*standardization by standard deviation*)
- centrage et normalisation par espèces (*standardization by standard deviation*)
- double centrage normalisé alternatif jusqu'à convergence.

En référence à⁴¹ on a donc un usage intensif du schéma de dualité à deux produits scalaires canoniques dans l'article de Noy-Meir (1973)⁴⁰. D'autres transformations préalables sont encore citées dont $x_{ik} / \sqrt{\sum_i x_{ik}^2} \sqrt{\sum_k x_{ik}^2}$ (*double standardization by site total and species total*) en référence à⁴² car il s'agit bien d'AFC. Le terme *standardization* est utilisé en général pour division par une quantité. On a donc passage en fréquences (profils) par *standardization* par la marge, passage en longueur unité par *standardization* par la norme et passage en variance unité par *standardization* par l'écart-type. Au risque de répéter une évidence, on notera encore que parler d'analyse de données à la française est une ineptie.

Le terme de dualité (*The R/Q problem: duality*) est introduit en écologie dans l'article fondamental de Noy-Meir et Whittaker 1977⁴³.

After twenty-five years of development of continuous multivariate techniques in ecology, some of the early optimistic promises, as well as some of the septical criticisms, seem to have been overstated. No single method has emerged as a solution to all problems of describing and explaining patterns of compositional variation in natural communities. However, several ordination techniques have proved to be helpful tools in the analysis community data for various scientific and practical purposes.

La généralisation des méthodes de couplages, des ordinations à K tableaux, des analyses sous contraintes d'échelles relance le débat. L'ACP centrée par espèce sur les profils espèces (PACP), l'analyse de co-inertie qui lui est associée (Analyse OMI) et l'analyse canonique (au sens Canonical Correspondence Analysis) qui en dérive basée sur la marginalité des niches peuvent faire partie de ces outils utiles.

Références

¹ Ter Braak, C.J.F. (1986) Canonical correspondence analysis : a new eigenvector technique for multivariate direct gradient analysis. *Ecology* : 69, 69-77.

² Whittaker, R.H. (1973) *Ordination and classification of communities*. Handbook of vegetation science. Part V. Dr. W. Junk b.v., The Hague. 1-737.

³ Hill, M.O. (1974) Correspondence analysis : A neglected multivariate method. *Journal of the Royal Statistical Society, C* : 23, 340-354.

Hill, M.O. (1973) Reciprocal averaging : an eigenvector method of ordination. *Journal of Ecology* : 61, 237-249.

⁴ Ter Braak, C.J.F. (1985) Correspondence analysis of incidence and abundance data : properties in terms of a unimodal response model. *Biometrics* : 41, 859-873.

⁵ Thioulouse, J. & Chessel, D. (1992) A method for reciprocal scaling of species tolerance and sample diversity. *Ecology* : 73, 670-680.

⁶ Dolédec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.

⁷ Chessel, D., Lebreton, J.D. & Yoccoz, N. (1987) Propriétés de l'analyse canonique des correspondances. Une utilisation en hydrobiologie. *Revue de Statistique Appliquée* : 35, 4, 55-72.

⁸ Ter Braak, C.J.F. (1987) CANOCO - a FORTRAN program for Canonical community ordination by [partial][detrended][canonical] correspondence analysis and redundancy analysis. Software documentation. Version 2.1, TNO Institute of Applied Computer Science, Wageningen.

⁹ Hausser, J. (1995) *Säugetiere der Schweiz. Mammifères de la Suisse. Mammiferi della Svizzera*. Birkhäuser Verlag, Basel, Boston, Berlin. 501 p.

¹⁰ Perrin, N. (1984) *Contribution à l'écologie du genre Cepaea (Gastropoda) : approche descriptive et expérimentale de l'habitat et de la niche écologique*. Thèse Univ. Lausanne. 1-158.

¹¹ Tenenhaus, M. & Young, F.W. (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika* : 50, 1, 91-119.

¹² Chevenet, F., Dolédec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309.

¹³ Hill, M.O. & Smith, A.J.E. (1976) Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* : 25, 249-255.

¹⁴ Escoufier, Y. (1987) The duality diagram : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute, Serie G. Springer Verlag, Berlin. 139-156.

¹⁵ Chessel, D. & Mercier, P. (1993) Couplage de triplets statistiques et liaisons espèces-environnement. In : *Biométrie et Environnement*. Lebreton, J.D. & Asselain, B. (Eds.) Masson, Paris. 15-44.

¹⁶ Manly, B.F.J. (1991) *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London. 1-281.

¹⁷ Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1995) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* : 20, 643-656.

¹⁸ Usseglio-Polatera, P. & Auda, Y. (1987) Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie* : 23, 1, 65-79.

¹⁹ Rojas-Camousseight, F. (1985) *Etudes préliminaires sur l'utilisation des Trichoptères adultes comme descripteurs écologiques*. Thèse de doctorat, Université Lyon 1. 1-215.

²⁰ Eckart, C. & Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika* : 1, 211-218.

²¹ Gabriel, K.R. (1971) The biplot graphical display of matrices with application to principal component analysis. *Biometrika* : 58, 453-467.

²² Gabriel, K.R. & Zamir, A. (1979) Lower rank approximation of matrices with any choice of weights. *Technometrics* : 21, 459-498.

²³ Kenkel, N.C. & Orloci, L. (1986) Applying metric and nonmetric multidimensional scaling to ecological studies : some new results. *Ecology* : 67, 919-928.

²⁴ Orloci, L. (1981) Probing time series vegetation data for evidence of succession. *Vegetatio* : 46, 31-35.

²⁵ Feoli, E. & Orloci, L. (1979) Analysis of concentration and detection of underlying factors in structured tables. *Vegetatio* : 40, 49-54.

²⁶ Prodon, R. & Lebreton, J.D. (1981) Breeding avifauna of a Mediterranean succession : the holm oak and cork oak series in the eastern Pyrénées. 1 : Analysis and modelling of the structure gradient. *Oikos* : 37, 21-38.

²⁷ Blondel, J., Chessel, D. & Frochot, B. (1988) Niche expansion and density compensation of island birds in mediterranean habitats. A case study from comparison of two ecological successions. *Ecology* : 69, 6, 1899-1917.

²⁸ Lebreton, J.D. & Yoccoz, N. (1987) Multivariate analysis of bird count data. *Acta Oecologica, Oecologia Generalis* : 8, 2, 125-144.

²⁹ Verneaux, J. (1973) *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie*. Thèse d'état, Besançon. 1-257.

³⁰ Huet, M. (1959) Profiles and biology of western european streams as related to fish management. *Transactions of the American Fisheries Society* : 88, 3, 166-163.

³¹ Belair, G. de. (1981) *Biogéographie et aménagement : la plaine de La Mafragh (Annaba, Algérie)*. Thèse de 3^o cycle. Université Paul Valéry, Montpellier. 1-150.

³² Belair, G. de & Bencheikh-Lehocine, M. (1987) Composition et déterminisme de la végétation d'une plaine côtière marécageuse : La Mafragh (Annaba, Algérie). *Bulletin d'Ecologie* : 18, 4, 393-407.

³³ Ter Braak, C.J.F. (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio* : 69, 69-77.

³⁴ Orłóci, L. (1966) Geometric models in ecology. I. The theory and applications of some ordination methods. *Journal of Ecology* : 54, 193-215.

³⁵ Orłóci, L. (1967) Data centering : a review and evaluation with reference to component analysis. *Systematic Zoology* : 16, 208-212.

³⁶ Austin, M.P. & Greig-Smith, P. (1968) The application of quantitative methods to vegetation survey. II Some methodological problems of data from rain forest. *Journal of Ecology* : 56, 827-844.

³⁷ Orłóci, L. (1967) An agglomerative method for classification of plant communities. *Journal of Ecology* : 55, 193-206.

³⁸ Bray, J.R. & Curtis, J.T. (1957) An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* : 27, 4, 325-349.

³⁹ Austin, M.P. & Orłóci, L. (1966) Geometric models in ecology II An evaluation of some ordination techniques. *Journal of Ecology* : 54, 217-227.

⁴⁰ Noy-Meir, I. (1973) Data transformations in ecological ordination. I. Some advantages of non-centering. *Journal of Ecology* : 61, 329-341.

⁴¹ Gower, J.C. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* : 53, 325-338.

⁴² Benzecri, J.P. (1969) Statistical analysis as a tool to make patterns emerge from data. In : *Methodologies of pattern recognition*. Watanabe, S. (Ed.) Academic Press, New-York. 35-60.

⁴³ Noy-Meir, I. & Whittaker, R.H. (1977) Continuous multivariate methods in community analysis: some problems and developments. *Vegetatio* : 33, 79-98.

