

ADE-4



Fiche thématique 5.D

Traits biologiques : variables ou K-tableau ?

Résumé

La fiche donne des indications sur la nature des tableaux de traits biologiques. A partir de plusieurs exemples on montre qu'il est possible de choisir entre deux types de stratégies, la première visant à transformer un ensemble de traits en paquets de variables quantitatives, la seconde gardant au tableau de traits le statut de *K*-tableau.

Plan

| | |
|--|----|
| 1 — INTRODUCTION..... | 2 |
| 2 — DONNEES DISTRIBUTIONNELLES | 5 |
| 2.1 Moyenne d'un trait pour un ensemble de taxa..... | 5 |
| 2.2 Variance d'un trait et diversité..... | 7 |
| 3 — L'ACP FLOUE..... | 9 |
| 4 — L'ACM FLOUE | 12 |
| 5 — STATIS : CORRELATION ENTRE VARIABLES FLOUES..... | 16 |
| 5.1 — STATIS et traits biologiques..... | 16 |
| 5.2 — STATIS et traits écologiques | 20 |
| 6 — ANALYSE FACTORIELLE MULTIPLE | 21 |
| 7 — DE K A $K + 1$ TABLEAUX..... | 28 |
| 8 — CONCLUSION..... | 36 |
| REFERENCES | 37 |

D. Chessel & S. Dolédec

1 — Introduction

Soit un tableau espèces-trait biologiques enregistrées par variables floues ¹. Le code des variables et des modalités est :

| | | |
|---|---|---------------------------|
| Potential size | 1 | ≤ 5 mm |
| | 2 | > 5-10 mm |
| | 3 | > 10-20 mm |
| | 4 | > 20-40 mm |
| | 5 | > 40 mm |
| Potential no. of reproductive cycles per year | 1 | < 1 cycle per year |
| | 2 | = 1 cycle per year |
| | 3 | > 1 cycle per year |
| Longevity of adults | 1 | < 1 day |
| | 2 | > 1-10 days |
| | 3 | > 10-30 days |
| | 4 | > 30-365 days |
| | 5 | > 365 days |
| Reproductive technique | 1 | hermaphroditism |
| | 2 | male et female |
| Parental care | 1 | ovoviviparity |
| | 2 | isolated eggs |
| | 3 | cemented aquatic eggs |
| | 4 | eggs in vegetation |
| | 5 | cemented terrestrial eggs |
| Dissemination potential of larvae | 1 | ≤ 10 m |
| | 2 | > 10-100 m |
| | 3 | > 100-1000 m |
| | 4 | > 1000 m |
| Attachment to substrate | 1 | swimmers |
| | 2 | crawlers |
| | 3 | burrowers |
| | 4 | temporary attachment |
| | 5 | permanent attachment |
| Body flexibility | 1 | high (> 45°) |
| | 2 | low (> 10-45°) |
| | 3 | none (< 10°) |
| Body form | 1 | streamlined |
| | 2 | flattened |
| | 3 | cylindrical |
| | 4 | spherical |
| Feeding habits | 1 | shredders |
| | 2 | scrapers |
| | 3 | deposit-feeders |
| | 4 | filter-feeders |
| | 5 | piercers |
| Food | 1 | detritus ≤ 1mm |
| | 2 | detritus > 1mm |
| | 3 | living plants ≤ 1mm |
| | 4 | living plants > 1mm |
| | 5 | living animals ≤ 1mm |
| | 6 | living animals > 1-10mm |
| | 7 | living animals > 10mm |
| Respiration | 1 | aquatic (plastron) |
| | 2 | aerial (spiracle) |
| | 3 | aquatic (spiracle) |
| | 4 | aquatic (gills) |
| | 5 | aquatic (tegument) |

Il y a 12 variables comptant respectivement 5, 3, 5, 2, 5, 4, 5, 3, 4, 5, 7 et 5 modalités.

| TraitCla | |
|----------|--------|
| 1 | 5.0000 |
| 2 | 3.0000 |
| 3 | 3.0000 |
| 4 | 2.0000 |
| 5 | 5.0000 |
| 6 | 4.0000 |
| 7 | 5.0000 |
| 8 | 3.0000 |
| 9 | 4.0000 |
| 10 | 5.0000 |
| 11 | 7.0000 |
| 12 | 5.0000 |

Le code des espèces est :

| | Espèce | | Espèce |
|----|--------------------------|----|--------------------------|
| 1 | Pisidium casertanum | 43 | Haliphus ruficollis |
| 2 | Pisidium henslowanum | 44 | Haliphus lineatocollis |
| 3 | Pisidium hibernicum | 45 | Haliphus oblicuus |
| 4 | Pisidium milium | 46 | Helophorus brevipalpis |
| 5 | Pisidium nitidum | 47 | Enochrus sp |
| 6 | Pisidium obtusale | 48 | Baetis pentaplebedes |
| 7 | Pisidium personatum | 49 | Baetis gr. fuscatus |
| 8 | Pisidium pseudosphaerium | 50 | Baetis vernus |
| 9 | Pisidium subtruncatum | 51 | Centroptilum luteolum |
| 10 | Sphaerium corneum | 52 | Cloeon simile |
| 11 | Sphaerium lacustre | 53 | Cloeon dipterum |
| 12 | Bithynia tentaculata | 54 | Caenis robusta |
| 13 | Valvata cristata | 55 | Caenis luctuosa |
| 14 | Acroloxus lacustris | 56 | Caenis horaria |
| 15 | Galba corvus | 57 | Ephemerella ignita |
| 16 | Lymnaea stagnalis | 58 | Leptophlebia vespertina |
| 17 | Radix ovata | 59 | Siphonurus lacustris |
| 18 | Physa acuta | 60 | Aeshna sp. |
| 19 | Physa fontinalis | 61 | Anax imperator |
| 20 | Anisus vortex | 62 | Chalcolestes viridis |
| 21 | Armiger crista | 63 | Coenagrion gr. puella |
| 22 | Bathyomphalus contortus | 64 | Enallagma cyathigerum |
| 23 | Ferrissia wautieri | 65 | Erythromma viridulum |
| 24 | Gyraulus albus | 66 | Ischnura elegans |
| 25 | Hippeutis complanatus | 67 | Sympetrum gr. striolatum |
| 26 | Planorbis carinatus | 68 | Platycnemis pennipes |
| 27 | Gammarus pulex | 69 | Silo nigricornis |
| 28 | Gammarus fossarum | 70 | Hydroptila martini |
| 29 | Asellus aquaticus | 71 | Oxyethira tristella |
| 30 | Proasellus meridianus | 72 | Lepidostoma hirtum |
| 31 | Dryops sp. | 73 | Athripsodes aterrimus |
| 32 | Agabus undulatus | 74 | Athripsodes albifrons |
| 33 | Agabus bipustulatus | 75 | Mystacides azurea |
| 34 | Hydroporus palustris | 76 | Oecetis furva |
| 35 | Hygrotus inaequalis | 77 | Anabolia nervosa |
| 36 | Hygrotus versicolor | 78 | Chaetopteryx villosa |
| 37 | Hyphydrus ovatus | 79 | Limnephilus gr. lunatus |
| 38 | Potamonectes depressus | 80 | Limnephilus rhombicus |
| 39 | Rhantus pulverosus | 81 | Cyrnus flavidus |
| 40 | Elmis aenea | 82 | Plectrocnemia conspersa |
| 41 | Esollus parallelepipedus | 83 | Sericostoma galeatum |
| 42 | Oulimnius tuberculatus | | |

2 — Données distributionnelles

Un trait biologique saisi par des notes d'association du taxon avec chacune des modalités du trait a_1, a_2, \dots, a_m est en fait une distribution de fréquence ou pondération des modalités associée au taxon. On calcule $a_t = \sum_{j=1,m} a_j$ et $p_j = a_j / a_t$.

Lorsqu'il y a plusieurs taxons et plusieurs traits, on notera :

- T , le nombre total de taxa ;
- t , le numéro d'ordre d'un taxon $1 \leq t \leq T$;
- V , le nombre total de traits biologiques ;
- v , le numéro d'ordre d'un trait biologique $1 \leq v \leq V$;
- M_v , le nombre total de modalités du trait v ;
- m_v , le numéro d'ordre d'une modalité du trait $1 \leq m_v \leq M_v$;
- a_{tm_v} la note d'association du taxon t à la modalité m_v
- $a_{t \cdot v}$ la somme des notes d'association du taxon t aux modalités du trait v ;
- f_{tm_v} le poids (fréquence) de la modalité m_v pour le taxon t ;

$$f_{tm_v} = a_{tm_v} / a_{t \cdot v}$$

On a évidemment :

$$\sum_{m_v=1}^{m_v=M_v} f_{tm_v} = 1$$

- M le nombre total des modalités de tous les traits ($M = M_1 + M_2 + \dots + M_v$);
- m le numéro d'ordre d'une modalité dans la série de toutes les modalités de tous les traits $1 \leq m \leq M$;

Les tableaux taxa-traits ont donc T lignes (taxa) et M colonnes (modalités) et se décomposent en V sous-tableaux (traits) contenant respectivement m_1, m_2, \dots, m_v colonnes (modalités des traits). On est directement, avec la notion de traits biologiques dans la partie multi-tableaux de la statistique multivariée. C'est pourquoi il a été fait peu d'usage de ce type d'informations et que de nombreuses décisions sont à prendre en ce qui concerne leur manipulation. Qu'est-ce que la variabilité d'un trait, la corrélation entre deux traits, la redondance d'une famille de traits sont des questions à résoudre. Commençons par la moyenne.

2.1 Moyenne d'un trait pour un ensemble de taxa

On suppose que chaque taxon a lui-même un poids qu'on notera π_t ($1 \leq t \leq T$). Ce poids sera $1/T$ si on décide d'une pondération uniforme des taxa (ils ont tous la même importance *a priori*) ou ces poids peuvent provenir d'un tableau faunistique (ils auront un poids proportionnel à leur abondance).

Pour le trait t , la moyenne du trait est défini par la pondération moyenne :

$$f^*_{m_v} = \sum_{t=1}^T \pi_t f_{tm_v}$$

Ceci ne soulève aucune difficulté. Pour représenter les tableaux de traits, nous utiliserons la représentation graphique associée à la lecture des tableaux de traits.



```
*****
* Description of a fuzzy coded matrix *
*****
Input file: T
Row number: 83, column number: 53
Number of modalities for each variable: file TCl1a
Row number: 12
Missing data: 0
```

Description of categories:

Variable number 1 has 5 categories

```
[ 1] Category: 1   Freq.: 0.32
[ 2] Category: 2   Freq.: 0.33
[ 3] Category: 3   Freq.: 0.177
[ 4] Category: 4   Freq.: 0.134
[ 5] Category: 5   Freq.: 0.0392
```

...

Variable number 12 has 5 categories

```
-----
[ 49] Category: 1   Freq.: 0.111
[ 50] Category: 2   Freq.: 0.0793
[ 51] Category: 3   Freq.: 0.0181
[ 52] Category: 4   Freq.: 0.519
[ 53] Category: 5   Freq.: 0.272
-----
```

Output file: TF

Row number: 83, column number: 53

It contains the modified table with type [L1/.../Lv]

[0, 0, ..., 0] are preserved for missing data

Les poids de la colonne de droite sont les moyennes des profils par taxon pour la distribution uniforme. Les valeurs f_{tm_v} sont dans le fichier TF.

On représente ces valeurs par la nouvelle option Tables : Fuzzy Variables :



On obtient la figure 1.

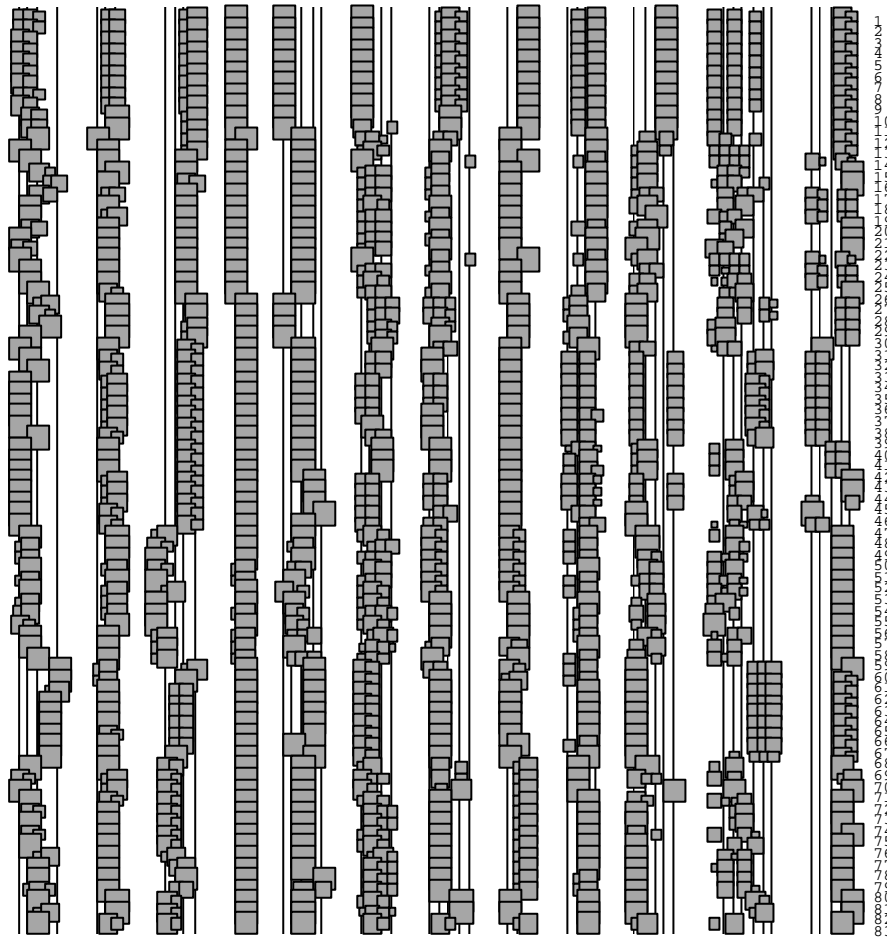


Figure 1 : Représentation d'un tableau de traits biologiques. Chaque ligne est une espèce et chaque groupe de colonnes est une variable.

2.2 Variance d'un trait et diversité

Mesurer la variance devient moins intuitif. C'est essentiel, si on veut passer de la mesure de la diversité taxonomique (qu'elle soit alpha-diversité dans une station ou beta-diversité entre stations) à la diversité biologique dans les mêmes conditions. Nous nous étions jusqu'à présent concentrés sur le contenu en traits biologiques et son déterminisme par la variabilité spatio-temporelle des écosystèmes. La notion de variabilité d'un trait avait été ignorée.

Mesurer le diversité d'un ensemble, c'est apprécier quantitativement la différence qui existe entre tous les éléments de cet ensemble. C'est clairement vrai pour une variable quantitative, dont on mesure la variabilité par la variance. Pour une pondération uniforme des éléments de la collection de valeurs on a :

$$\text{Var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - m(x))^2 = \frac{1}{n} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{j=1}^n x_j \right)^2$$

Or,

$$\frac{1}{n^2} \sum_{i,j=1}^n (x_i - x_j)^2 = \frac{1}{n^2} \sum_{i,j=1}^n (x_i^2 + x_j^2 - 2x_i x_j) = \frac{1}{n} \sum_{i=1}^n x_i^2 + \frac{1}{n} \sum_{j=1}^n x_j^2 - \frac{2}{n^2} \left(\sum_{i=1}^n x_i \right) \left(\sum_{j=1}^n x_j \right)$$

Donc :

$$2\text{Var}(x) = \frac{1}{n^2} \sum_{i,j=1}^n (x_i - x_j)^2$$

Curieusement, si on ne compte que les couples de deux points distincts, on donne le même sens à la variance estimée (dite sans biais) :

$$2\hat{\text{Var}}(x) = \frac{1}{n(n-1)} \sum_{\substack{i,j=1 \\ i \neq j}}^n (x_i - x_j)^2$$

Une variabilité est toujours la moyenne des carrés des distances entre objets de la collection. La diversité d'un ensemble dont on sait mesurer la différence entre deux éléments par un indice de distance et dans lequel chaque élément a un poids sera donc définie par :

$$\text{Div}(A) = \frac{1}{2} \sum_{x,y \in A} p_x p_y d^2(x,y)$$

Si A est une partie finie d'un espace euclidien muni d'un produit scalaire , on a alors une mesure euclidienne de la diversité par :

$$\text{Div}_\Omega(A) = \frac{1}{2} \sum_{x,y \in A} p_x p_y \|x - y\|_\Omega^2$$

Ceci n'est rien d'autre que l'inertie du nuage des points de A au sens de la norme . Nous avons discuté de la concordance des points de vue dans la comparaison entre l'indice CATANOVA de Light & Margolin ², et le procédé qu'avait utilisé Lebart ³ pour introduire les graphes de voisinages en analyse des données et la concordance entre l'indice de Simpson et l'inertie d'un tableau d'indicatrices ⁴ (Voir fiche 2.8).

Il y aura donc autant de mesure de diversité possible que de métrique dans l'espace utilisé. Ceci nous éloigne directement de l'AFC floue ¹ qui utilise la métrique du khi2 qui n'est

justifiée que par les propriétés qu'elle donne (double averaging) mais pas par les données elles-mêmes.

Pour mesurer la variabilité d'un ensemble de traits biologiques dans l'objectif de mesurer la diversité, il y aura peut-être intérêt à perdre certaines propriétés de l'AFC au profit de la simplicité de l'ACP sur données distributionnelles qui a déjà été connectée à la notion de diversité dans ⁵. Comparons donc la variabilité d'un trait biologique comme inertie de l'ACP du tableau constitué par les modalités de ce trait et comme AFC de ce même tableau. Le choix d'une métrique est ainsi l'élément de base qui entraîne de multiples conséquences aux niveaux de 2 tableaux, trois tableaux et K tableaux.

3 — L'ACP floue

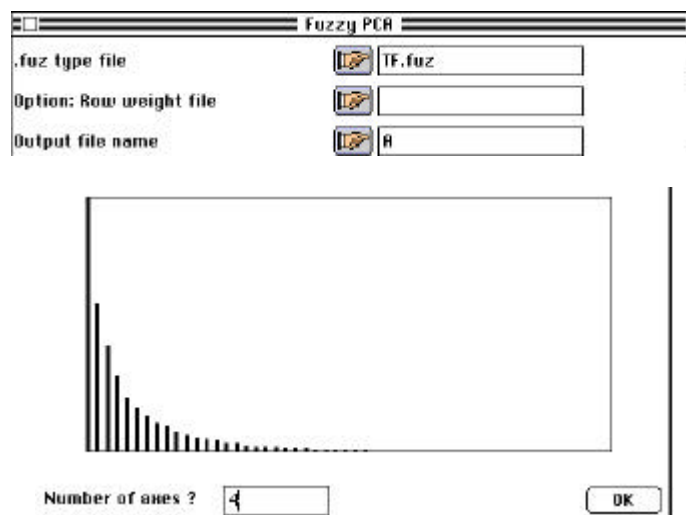
Dans cette stratégie, on considère que chaque trait définit une ACP centrée par modalité. Cette analyse utilise le triplet basé sur une pondération arbitraire des taxa :

$$\mathbf{D} = \text{Diag}(\pi_1, \pi_2, \dots, \pi_t)$$

et le tableau centré :

$$\mathbf{X} = [f_{tm_v} - f^*_{m_v}]_{\substack{1 \leq t \leq T, \\ 1 \leq v \leq V, \\ 1 \leq m_v \leq M_v}}$$

On peut hésiter sur la pondération des modalités. Pour des raisons mathématiques, on prendra la pondération uniforme. L'option nouvelle PCA : Fuzzy PCA permet cette opération :



Le triplet associé est constitué :

```
fp/FuzzyPCA: PCA on fuzzy table
Input file: TF.fuz for access to file TF
Row number: 83, column number: 53
Uniform row weights
```

```
File A.cppl contains the row weights
It has 83 rows and 1 column
```

File A.cppc contains the column weights $\text{Diag}(\text{Unif1}, \dots, \text{UnifV})$
 It has 53 rows and 1 column

File A contains the raw table
 It has 83 rows and 53 columns (categories)

File A.cpta contains the centred table
 It has 83 rows and 53 columns (categories)

File A.cpma contains
 ----- number of rows: 83
 ----- number of variables: 12
 ----- number of categories: 53
 ----- variable number of each category (vector of 53 values)

On utilise le triplet très simple :

$$\left(\mathbf{X} = \left[f_{tm_v} - f^*_{m_v} \right] \text{Diag} \left(\underbrace{\frac{1}{m_1}, \dots, \frac{1}{m_1}}_{m_1}, \dots, \underbrace{\frac{1}{m_V}, \dots, \frac{1}{m_V}}_{m_V} \right), \frac{1}{T} \mathbf{I}_T \right)$$

Cette ACP centrée est traitée normalement :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: A.cpta
--- Number of rows: 83, columns: 53
-----
Total inertia: 1.2316
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +3.4357E-01 +0.2790 +0.2790 | 02  +1.9903E-01 +0.1616 +0.4406 |
03  +1.4404E-01 +0.1170 +0.5575 | 04  +1.0170E-01 +0.0826 +0.6401 |
05  +7.2450E-02 +0.0588 +0.6989 | 06  +5.9631E-02 +0.0484 +0.7473 |
07  +4.7196E-02 +0.0383 +0.7857 | 08  +3.9520E-02 +0.0321 +0.8177 |
...

```

File A.cvpv contains the eigenvalues and relative inertia for each axis
 --- It has 53 rows and 2 columns

File A.cpcp contains the column scores
 --- It has 53 rows and 4 columns

```
File :A.cpcp
|Col.|  Mini  |  Maxi  |
|----|-----|-----|
|  1|-4.279e-01| 4.279e-01|
|  2|-4.007e-01| 2.734e-01|
|  3|-3.286e-01| 3.231e-01|
|  4|-2.472e-01| 2.126e-01|
|----|-----|-----|

```

File A.cpli contains the row scores
 --- It has 83 rows and 4 columns

```
File :A.cpli
|Col.|  Mini  |  Maxi  |
|----|-----|-----|
|  1|-1.173e+00| 5.832e-01|
|  2|-8.843e-01| 7.016e-01|
|  3|-6.799e-01| 6.984e-01|

```

```
| 4|-6.724e-01| 6.297e-01|
|----|-----|-----|
```

Le listing donne en outre des informations spécifiques. La première de ces informations concerne la répartition de l'inertie totale (la variabilité biologique totale du groupe de taxa étudiés) entre les différents descripteurs. La variance des colonnes est sommée par blocs (traits) et rapportée au total (édition en 1 pour 10000) :

Total inertia: 1.2316e+00 - Number of axes: 4
Contribution of fuzzy variable to total inertia

| Num | Inertia | Ratio | |
|-----|------------|-------------|---------------|
| 1 | 1.1405e-01 | 926 | Size |
| 2 | 1.0537e-01 | 856 | Cycles |
| 3 | 8.3335e-02 | 677 | Longevity |
| 4 | 2.0379e-01 | 1655 | Reproduction |
| 5 | 1.1612e-01 | 943 | Parent_care |
| 6 | 6.0886e-02 | 494 | Dissemination |
| 7 | 6.7083e-02 | 545 | Attachement |
| 8 | 1.6714e-01 | 1357 | Flexibility |
| 9 | 1.0396e-01 | 844 | Body_form |
| 10 | 9.3533e-02 | 759 | Feeding |
| 11 | 4.0543e-02 | 329 | Food |
| 12 | 7.5791e-02 | 615 | Respiration |
| Sum | 1.2316e+00 | 10000 | |

16.55 % de la variabilité totale est associée à la variable 4 (Reproduction) et la variable la moins discriminante est la 11 (alimentation).



Contribution of fuzzy variable to eigenvalues (values)

| Num | Fac 1 | Fac 2 | Fac 3 | Fac 4 |
|-----|------------|------------|------------|------------|
| 1 | 3.0300e-03 | 2.8967e-03 | 7.5312e-03 | 1.7832e-02 |
| 2 | 2.8472e-04 | 1.4028e-02 | 7.0796e-02 | 4.2593e-03 |
| 3 | 8.1795e-03 | 2.0022e-02 | 5.8435e-03 | 1.0134e-02 |
| 4 | 1.8313e-01 | 8.9097e-03 | 1.9264e-03 | 1.0452e-03 |
| 5 | 1.4726e-02 | 2.5036e-02 | 2.6431e-03 | 2.3228e-02 |
| 6 | 1.4562e-02 | 4.1067e-03 | 4.6883e-04 | 1.3364e-02 |
| 7 | 8.8792e-03 | 3.8355e-03 | 2.0490e-02 | 2.0294e-03 |
| 8 | 2.6292e-02 | 8.0371e-02 | 1.3131e-02 | 3.6638e-03 |
| 9 | 5.5040e-02 | 1.1323e-02 | 6.2585e-03 | 2.2007e-04 |
| 10 | 2.5855e-02 | 7.0860e-03 | 4.6715e-03 | 1.8462e-02 |
| 11 | 2.4608e-03 | 1.8740e-03 | 1.6627e-03 | 4.9315e-03 |
| 12 | 1.1328e-03 | 1.9537e-02 | 8.6157e-03 | 2.5328e-03 |
| Sum | 3.4357e-01 | 1.9903e-01 | 1.4404e-01 | 1.0170e-01 |

Contribution of fuzzy variable to eigenvalues (in 1/10000)

| Num | Fac 1 | Fac 2 | Fac 3 | Fac 4 |
|-----|-------|-------|-------|-------|
| 1 | 109 | 179 | 644 | 2159 |
| 2 | 10 | 868 | 6053 | 516 |
| 3 | 293 | 1239 | 500 | 1227 |

| | | | | |
|-------|-------------|-------------|-------|-------------|
| 4 | 6565 | 551 | 165 | 127 |
| 5 | 528 | 1549 | 226 | 2813 |
| 6 | 522 | 254 | 40 | 1618 |
| 7 | 318 | 237 | 1752 | 246 |
| 8 | 942 | 4973 | 1123 | 444 |
| 9 | 1973 | 701 | 535 | 27 |
| 10 | 927 | 438 | 399 | 2236 |
| 11 | 88 | 116 | 142 | 597 |
| 12 | 41 | 1209 | 737 | 307 |
| ----- | ----- | ----- | ----- | ----- |
| Sum | 10000 | 10000 | 10000 | 10000 |

On a enfin la contribution de chaque variable à chaque facteur. La concentration de 66% de la valeur propre sur la variable floue 4 (17% d'inertie) souligne un fait très important. En calculant la covariance entre modalité on fait simultanément le calcul des covariances entre modalités d'une même variable et entre modalités de variables différentes. La matrice de covariances \mathbf{C} s'écrit ainsi :

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1V} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2V} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{V1} & \mathbf{C}_{V2} & \cdots & \mathbf{C}_{VV} \end{bmatrix}$$

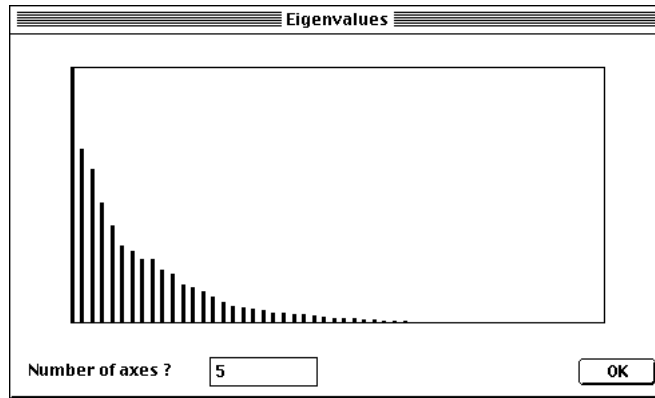
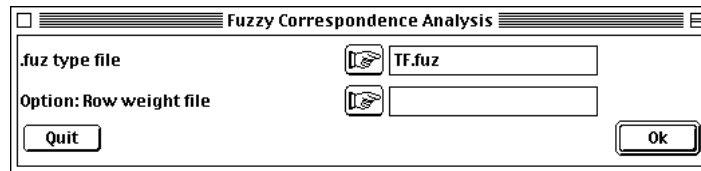
\mathbf{C}_{vv} est la matrice $m_v - m_v$ des covariances entre modalités de la même variable, alors que \mathbf{C}_{vw} est la matrice $m_v - m_w$ des covariances entre les m_v modalités de la variable v et les m_w modalités de la variable w . Il est clair que les covariances entre modalités d'une même variable sont essentiellement artefactuelles : elles expriment simplement soit que deux modalités mériteraient de n'en faire qu'une (covariance positive entre modalités utilisées ensemble) soit que deux modalités ne sont pas utilisées simultanément (covariance négative inhérente à ce type de variables).

Par contre les covariances entre deux modalités de deux variables contiennent de l'information biologique. C'est l'association entre diverses modalités de différentes variables qui définit la stratégie globale d'une ou plusieurs espèces. Ce mélange intime de deux types de covariation, l'une artefactuelle et l'autre significative laisse augurer de facteurs d'ACP consacrés soit à l'une, soit à l'autre soit encore au mélange des deux. L'artificiel l'a largement emporté dans l'analyse qui précède. Les covariances négatives entre modalité de la variable présentant le plus d'inertie font le facteur 1. Il n'est donc pas possible de voir dans un tableau de variables floues un simple tableau de modalités juxtaposées. C'est ici particulièrement net puisque la variable 4 est celle qui a le moins de modalité (2) et que chacune des deux est pratiquement utilisée seule (variable quasiment qualitative).

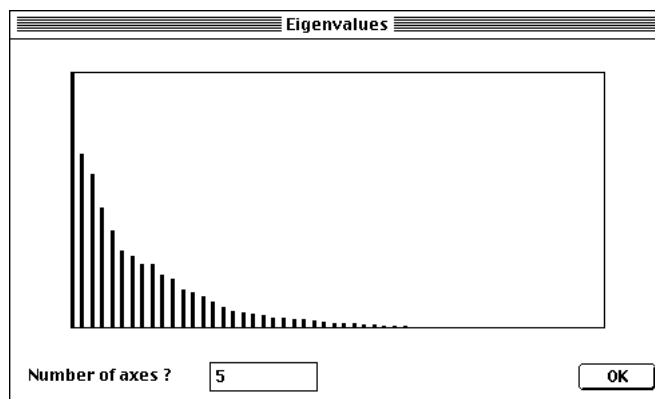
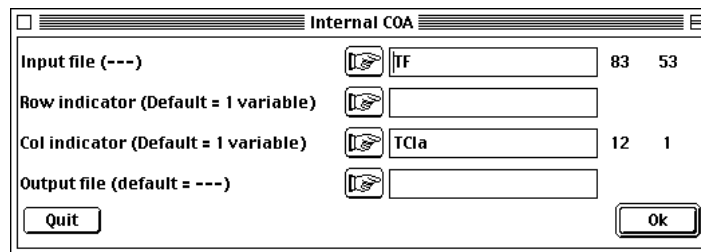
Un tableau de variables floues doit être considéré comme un K -tableau qu'on notera :

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_V]$$

4 — L'ACM floue



Comparons avec l'analyse des correspondances internes (voir ⁶) :



Les deux analyses ont exactement les mêmes valeurs propres, les mêmes coordonnées des lignes et les mêmes coordonnées des colonnes. Ces deux analyses sont strictement identiques. Ceci vient simplement du fait que le tableau flou contient des distributions de fréquences par bloc et que le recentrage par bloc explicite dans l'analyse des correspondances internes est implicite dans l'analyse des correspondances floues. Ce point est sans grande importance. Il permet simplement de refaire l'opération de décomposition des valeurs propres par sous-tableaux :

| | | | | | | | |
|-------|---|--------|--------|--------|--------|--------|--------|
| Bloc: | 1 | 0.0833 | 0.109 | 0.225 | 0.513 | 0.0926 | 0.267 |
| Bloc: | 2 | 0.0833 | 0.0305 | 0.028 | 0.0148 | 0.315 | 0.0245 |
| Bloc: | 3 | 0.0833 | 0.238 | 0.638 | 0.234 | 0.172 | 0.289 |
| Bloc: | 4 | 0.0833 | 0.515 | 0.0285 | 0.0556 | 0.112 | 0.0879 |

| | | | | | | | |
|-------|-------|--------|--------|--------|--------|--------|---------|
| Bloc: | 5 | 0.0833 | 0.71 | 0.408 | 0.571 | 0.14 | 0.372 |
| Bloc: | 6 | 0.0833 | 0.203 | 0.0641 | 0.114 | 0.0164 | 0.0595 |
| Bloc: | 7 | 0.0833 | 0.484 | 0.0238 | 0.105 | 0.314 | 0.0843 |
| Bloc: | 8 | 0.0833 | 0.422 | 0.424 | 0.0668 | 0.038 | 0.123 |
| Bloc: | 9 | 0.0833 | 0.511 | 0.143 | 0.0968 | 0.213 | 0.113 |
| Bloc: | 10 | 0.0833 | 0.816 | 0.271 | 0.353 | 0.295 | 0.127 |
| Bloc: | 11 | 0.0833 | 0.18 | 0.229 | 0.243 | 0.187 | 0.0751 |
| Bloc: | 12 | 0.0833 | 0.0453 | 0.428 | 0.195 | 0.112 | 0.00274 |
| ----- | ----- | ----- | 0.355 | 0.243 | 0.213 | 0.167 | 0.135 |

Eigenvalue percentage 1/10000

| | | | | | | | |
|-------|----|-------|-------------|-------------|-------------|-------------|-------------|
| Bloc: | 1 | ----- | 255 | 772 | 2004 | 461 | 1641 |
| Bloc: | 2 | ----- | 71 | 96 | 58 | 1571 | 151 |
| Bloc: | 3 | ----- | 558 | 2193 | 912 | 858 | 1778 |
| Bloc: | 4 | ----- | 1207 | 98 | 217 | 557 | 541 |
| Bloc: | 5 | ----- | 1665 | 1401 | 2229 | 696 | 2290 |
| Bloc: | 6 | ----- | 477 | 220 | 446 | 82 | 366 |
| Bloc: | 7 | ----- | 1136 | 82 | 411 | 1567 | 519 |
| Bloc: | 8 | ----- | 990 | 1458 | 261 | 189 | 759 |
| Bloc: | 9 | ----- | 1199 | 493 | 378 | 1060 | 697 |
| Bloc: | 10 | ----- | 1913 | 932 | 1376 | 1469 | 780 |
| Bloc: | 11 | ----- | 423 | 786 | 948 | 930 | 462 |
| Bloc: | 12 | ----- | 106 | 1469 | 761 | 559 | 17 |

La situation a radicalement changé. C'est parce que l'AFC floue est implicitement une analyse intra-classe et parce que l'analyse intra-classe est la plus simple des méthodes multi-tableaux qu'elle est bien préférable et a été utilisée.

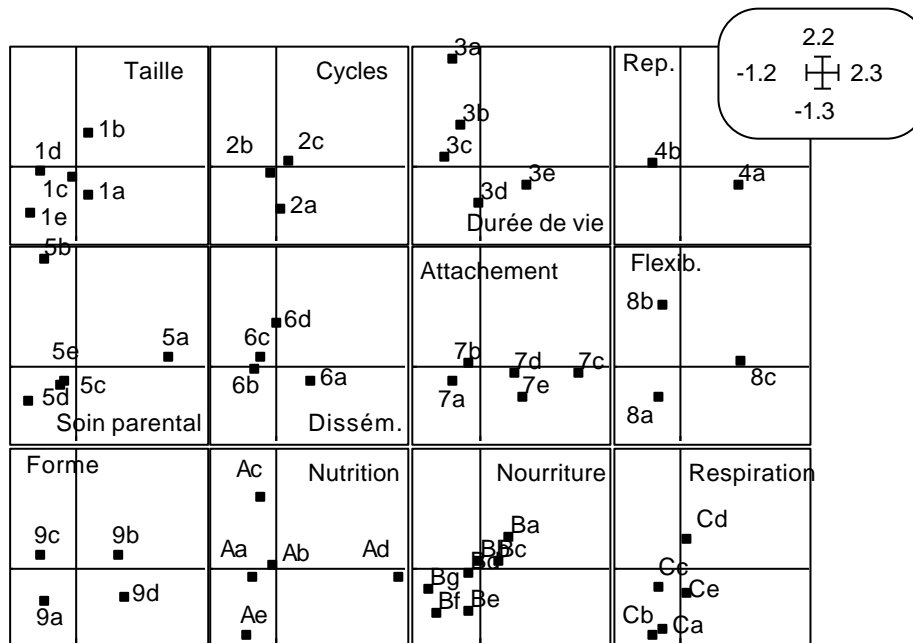
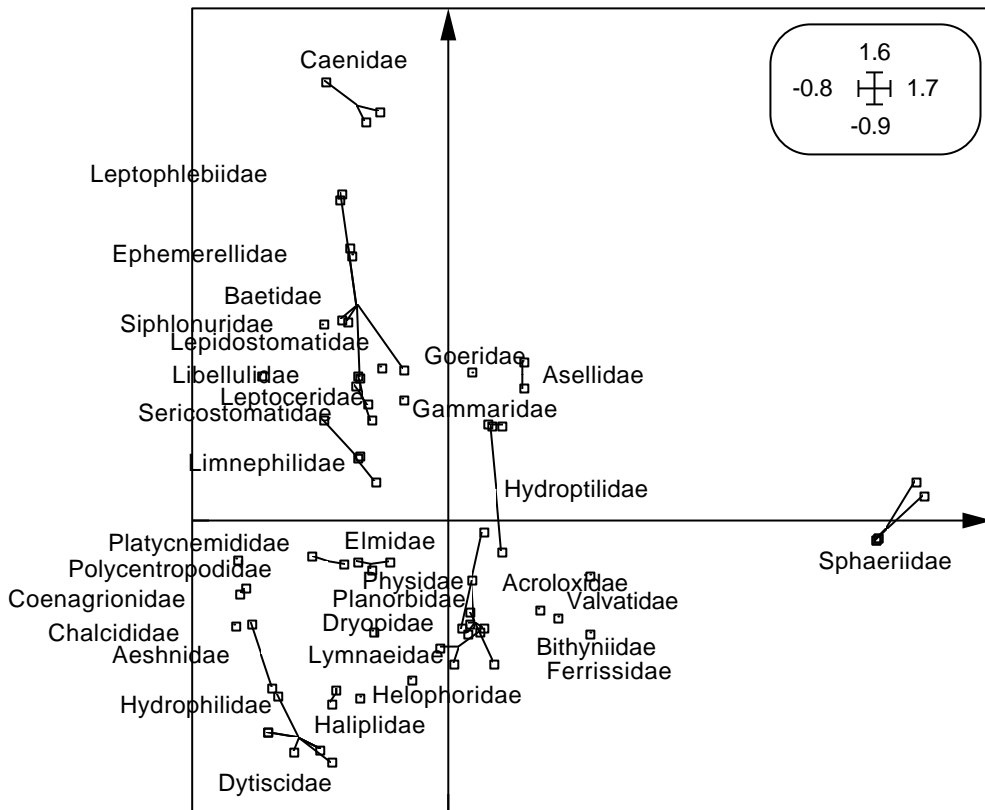


Figure 2 : Plan 1-2 de l'AFC floue. En haut cartes des lignes (espèces classées par famille), en bas cartes des colonnes (modalités réparties par variables).

Les principales propriétés de cette analyse sont (i) la symétrie lignes-colonnes, (ii) la métrique du Khi2 et (iii) la faible réduction de dimension.

(i) La symétrie lignes-colonnes est sensible sur la figure 2 ou l'analyse apparaît comme une AFC simultanée de chacun des tableaux-variables utilisant les mêmes scores-lignes.

Une espèce est vue comme une distribution de fréquence dans son usage d'une variable biologique et chaque modalité de chaque variable est vue comme une distribution de fréquence de son utilisation par les espèces. Le double averaging implicite conduit à la représentation simultanée des espèces et des modalités. Pour une description détaillée de la typologie des stratégies biologiques, cette analyse est indiscutablement utile.

(ii) Elle utilise cependant implicitement la métrique du Khi2 qui est contestable. Les carrés des différences entre deux profils sont pondérés par l'inverse de l'utilisation moyenne d'une modalité. La distance entre les taxons t_1 et t_2 s'écrit :

$$d_{t_1, t_2}^2 = \sum_{v=1}^V \sum_{m_v=1}^{M_v} \frac{(f_{t_1 m_v} - f_{t_2 m_v})^2}{f_{\cdot m_v}}$$

Une différence d'usage sur une modalité rare est ainsi fortement amplifiée, ce qui vu le type de mesure est loin de s'imposer. On préférerait simplement :

$$d_{t_1, t_2}^2 = \sum_{v=1}^V \sum_{m_v=1}^{M_v} (f_{t_1 m_v} - f_{t_2 m_v})^2$$

Mais l'analyse simple (ACP) associée à cette métrique est très défavorable.

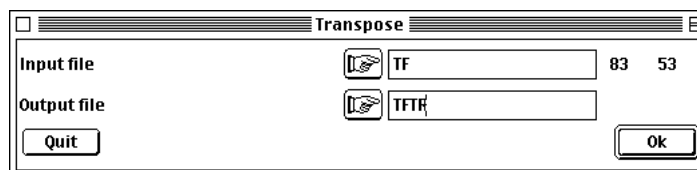
(iii) Le nombre de valeurs propres à prendre en compte est enfin élevé. Il est difficile de dire si cela vient vraiment des données et que les combinaisons stratégiques sont effectivement variées et complexes ou si cela vient du fait que les AFC multiples sont de mauvaises analyses d'inertie dans l'espace des modalités. L'AFC multiple est d'abord une analyse canonique (on y reviendra en détail) mais cette propriété est perdue dans la variante sur variables floues.

Ceci pose la question des méthodes K -tableaux dans l'usage des variables floues. L'analyse qui suit vise à savoir si cette introduction qui est relativement difficile au plan théorique en vaut la peine.

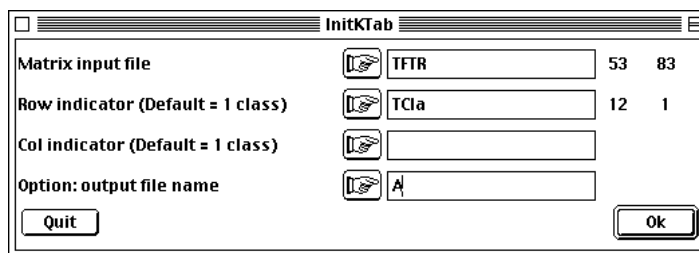
5 — STATIS : corrélation entre variables floues

5.1 — STATIS et traits biologiques

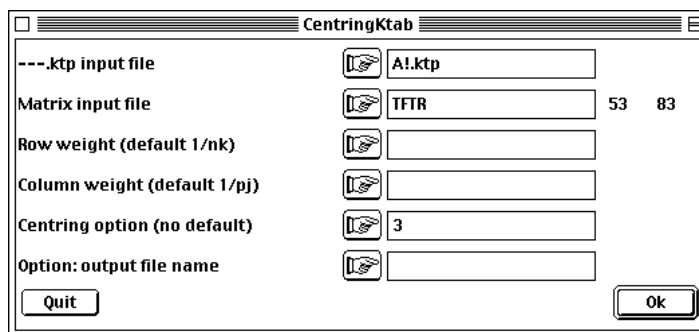
On travaillera systématiquement sur la version la plus simple d'une variable floue vue comme tableau, c'est-à-dire celle des profils par espèces simplement centrés. Il suffit de transposer le tableau flou :



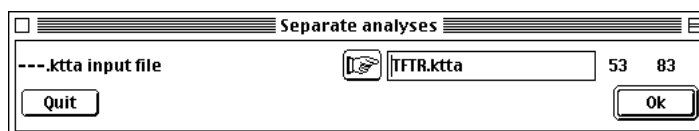
Les modalités (traits) sont en lignes et les espèces en colonnes. En faire un K -tableau (KTabUtil : InitKTab) :



Le centrer par KTabUtil : CentringKtab :



Chaque tableau a une inertie calculée par KTA-MFA : Separate analyses :



```

Row bloc: 1 - Col bloc: 1 - Total inertia: 1.31845
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +5.0588E-01 +0.3837 +0.3837  |02  +4.4992E-01 +0.3413 +0.7249  |
-----
Row bloc: 2 - Col bloc: 1 - Total inertia: 1.16524
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +8.7876E-01 +0.7541 +0.7541  |02  +2.4841E-01 +0.2132 +0.9673  |
-----
Row bloc: 3 - Col bloc: 1 - Total inertia: 1.24268
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +5.7713E-01 +0.4644 +0.4644  |02  +3.2389E-01 +0.2606 +0.7251  |
-----
Row bloc: 4 - Col bloc: 1 - Total inertia: 1.86253
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +1.8459E+00 +0.9911 +0.9911  |02  +1.6628E-02 +0.0089 +1.0000  |
-----
Row bloc: 5 - Col bloc: 1 - Total inertia: 1.07243
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +5.6811E-01 +0.5297 +0.5297  |02  +2.9113E-01 +0.2715 +0.8012  |
-----
Row bloc: 6 - Col bloc: 1 - Total inertia: 1.16809
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.9010E-01 +0.5908 +0.5908  |02  +3.4497E-01 +0.2953 +0.8861  |
-----
Row bloc: 7 - Col bloc: 1 - Total inertia: 0.921759
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +4.0215E-01 +0.4363 +0.4363  |02  +3.6346E-01 +0.3943 +0.8306  |
-----
Row bloc: 8 - Col bloc: 1 - Total inertia: 1.45585
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |

```

```

01  +1.0177E+00 +0.6990 +0.6990 | 02  +3.3870E-01 +0.2326 +0.9317 |
-----
Row bloc: 9 - Col bloc: 1 - Total inertia: 1.38184
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +8.5065E-01 +0.6156 +0.6156  | 02  +2.8404E-01 +0.2056 +0.8211  |
-----
Row bloc: 10 - Col bloc: 1 - Total inertia: 1.32414
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +5.8645E-01 +0.4429 +0.4429  | 02  +3.4790E-01 +0.2627 +0.7056  |
-----
Row bloc: 11 - Col bloc: 1 - Total inertia: 1.52417
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.1868E-01 +0.4059 +0.4059  | 02  +4.1539E-01 +0.2725 +0.6784  |
-----
Row bloc: 12 - Col bloc: 1 - Total inertia: 1.21243
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.4544E-01 +0.5324 +0.5324  | 02  +3.9088E-01 +0.3224 +0.8548  |
-----

```

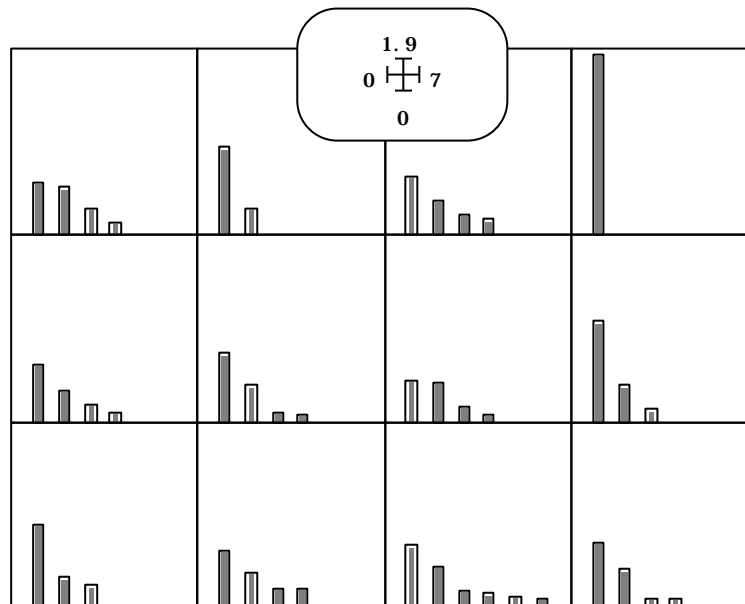
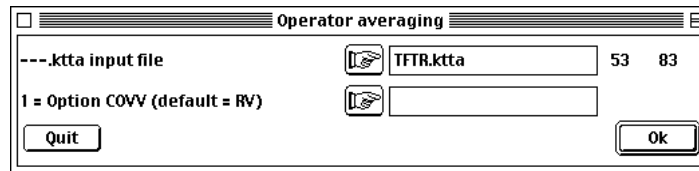


Figure 3: Valeurs propres des ACP centrées de chacun des traits biologiques.

Il est important de noter que dans ce K -tableau, chaque trait définit une analyse du type PCA : After row % transformation PCA. De plus chaque modalité a le même poids et chaque tableau a donc un poids unité (somme des poids des modalités d'un trait).

On voit immédiatement que les covariances entre les modalités d'un trait (dont une partie est artificielle puisque la somme des valeurs de départ vaut 1 par individus, ce qui implique que la somme des variables centrées est nulle) ne définit pas une structure (graphe des valeurs propres dans la figure 2). Seul le trait 4 (reproduction sexuée ou non) avec deux modalités presque disjonctive a une plus forte inertie entièrement concentrée dans l'unique valeur propre possible. Ceci est une conséquence directe du contenu mathématique et ne donne pas d'indication concrète. STATIS:Operator averaging permet de mesurer la corrélation entre ces ordinations :



Option operator averaging
 Input file TFTR
 Number of rows: 53, columns: 83

```
----- Correlation matrix -----
[ 1] 1000
[ 2]  18 1000
[ 3] 145 147 1000
[ 4]  33   6 164 1000
[ 5]  79 101 236 100 1000
[ 6]  70  17  99 181 294 1000
[ 7]  42 280 137 100 183 198 1000
[ 8]  63  16 199  97 216 213 162 1000
[ 9]  42  56 315  694 164 146 134 128 1000
[10] 158  76 159 343 315 317 291 193 277 1000
[11] 184  39 120  83 108 132 139 135  94 348 1000
[12] 125  14 262  34 128  26  92 239 128  70 138 1000
-----
```

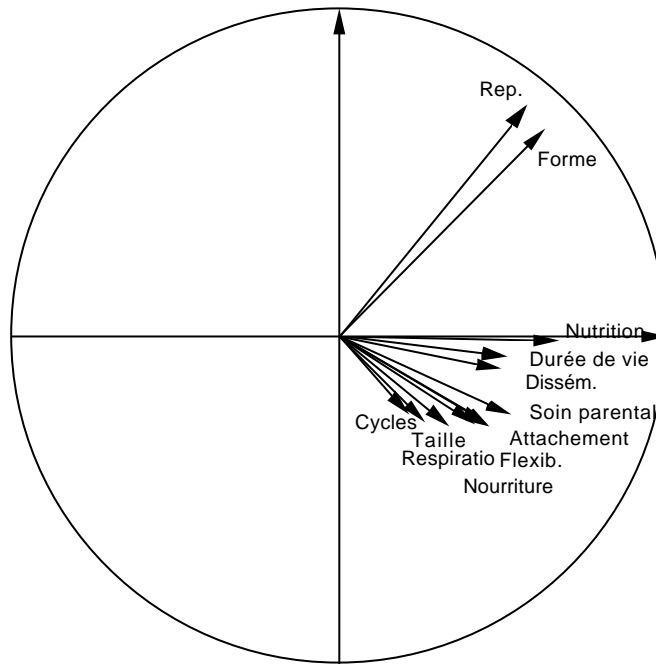
Seule la corrélation entre 4 (Reproduction) et 9 (Forme) est sensiblement élevée. Ceci plaide pour une complexité élevée de l'ensemble des traits. On cherche alors le compromis :

File TFTR.oa+RV contains the cosines between operators
 It has 12 rows and 12 columns
 Compromise with the cosines between operators

| Num. | Eigenval. | R.Iner. | R.Sum | Num. | Eigenval. | R.Iner. | R.Sum |
|------|-------------|---------|---------|------|-------------|---------|---------|
| 01 | +2.8091E+00 | +0.2341 | +0.2341 | 02 | +1.4001E+00 | +0.1167 | +0.3508 |
| 03 | +1.2016E+00 | +0.1001 | +0.4509 | 04 | +1.1463E+00 | +0.0955 | +0.5464 |
| 05 | +1.0494E+00 | +0.0874 | +0.6339 | 06 | +8.7149E-01 | +0.0726 | +0.7065 |
| 07 | +7.6326E-01 | +0.0636 | +0.7701 | 08 | +6.9271E-01 | +0.0577 | +0.8278 |
| 09 | +6.5214E-01 | +0.0543 | +0.8822 | 10 | +6.3062E-01 | +0.0526 | +0.9347 |
| 11 | +5.0728E-01 | +0.0423 | +0.9770 | 12 | +2.7598E-01 | +0.0230 | +1.0000 |

File TFTR.oa+EV contains the eigenvalues of the interstructure analysis
 It has 12 rows and 1 columns

File TFTR.oa+IS contains the table scores of the interstructure analysis
 It has 12 rows and 12 columns



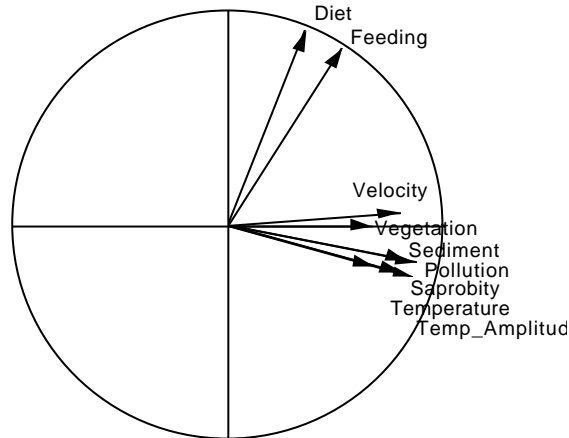
Cette figure est fort explicite. Il n'existe pas de corrélation forte entre les diverses ordination faites entre les espèces par les traits biologiques. La présence de redondance entre traits biologiques est faible. Ceci explique l'échec (qui devient relatif) de l'ACP et la complexité du graphe des valeurs propres de l'AFC. S'il n'y a pas de compromis, la suite est de faible intérêt et sera ignorée. En cherchant un compromis à analyser STATIS est ici invalidé dans sa fonction mais son principe est fort utile en montrant que l'objectif qu'il poursuit n'est pas judicieux, ce qu'on retrouve dans :

| Number | Rows | Weights | NS norm2 | Cos2 |
|--------|------|-----------|-----------|--------------|
| 1 | 5 | 1.568e-01 | 4.129e-03 | 0.003 |
| 2 | 3 | 1.294e-01 | 9.797e-03 | 0.001 |
| 3 | 5 | 3.091e-01 | 2.274e-03 | 0.103 |
| 4 | 2 | 3.443e-01 | 4.153e-02 | 0.716 |
| 5 | 5 | 3.156e-01 | 5.216e-03 | 0.113 |
| 6 | 4 | 2.948e-01 | 1.978e-03 | 0.112 |
| 7 | 5 | 2.770e-01 | 1.992e-03 | 0.045 |
| 8 | 3 | 2.713e-01 | 1.650e-02 | 0.195 |
| 9 | 4 | 3.753e-01 | 5.757e-03 | 0.543 |
| 10 | 5 | 4.052e-01 | 3.005e-03 | 0.241 |
| 11 | 7 | 2.497e-01 | 4.412e-04 | 0.017 |
| 12 | 5 | 2.008e-01 | 2.730e-03 | 0.030 |

Les traits biologiques sont faiblement corrélés. Ils sont encore multivariés mais les composantes d'un trait sont des variables dont la corrélation est artificielle. Un besoin se fait alors jour. Peut réduire les dimensions du tableau de variables biologiques tout en conservant l'information qu'il contient ?

5.2 — STATIS et traits écologiques

Une autre question mérite un détour. La méthode STATIS est-elle en cause dans le résultat qui précède ou bien un tableau de même forme peut-il donner un résultat totalement différent ? On utilise pour ce faire le tableau de profils écologiques publié dans ⁷ (carte Coléoptères). Le tableau ES a 110 lignes-espèces et 32 colonnes-modalités. Il est simplement transposé en ESTR (32-110) et transformé en *K*-tableau par KTabUtil: InitKTab sur ES-Bloc. Il est enfin centré par variables (KTabUtil: CentringKtab) et envoyé dans STATIS: Operator averaging :



A partir d'une matrice de RV très différente :

```

----- Correlation matrix -----
Vegetation[1]          1000
Sediment[2]           430 1000
Velocity[3]           429 689 1000
Pollution[4]         516 763 653 1000
Saprobity[5]          436 668 553 782 1000
Diet[6]                200 154 289 162 135 1000
Feeding[7]             343 314 452 324 257 921 1000
Temperature[8]        490 506 443 484 511 96 222 1000
Temp_Amplitude[9]    558 700 636 774 715 114 258 618 1000
-----

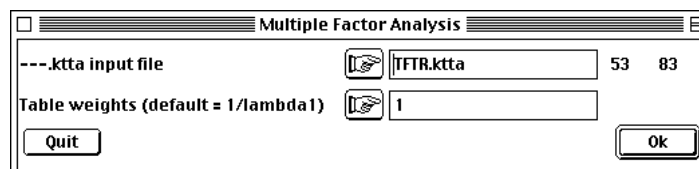
```

on obtient un résultat très différent. La fiche 2.5 montre que l'AFC floue fonctionne très bien sur ce tableau de traits écologiques. On voit ici que c'est d'abord parce qu'il s'agit d'un tableaux de traits corrélés. On notera que l'association des variables 6 et 7 définit le deuxième axe de l'AFC alors que le compromis sur les autres variables définit le premier.

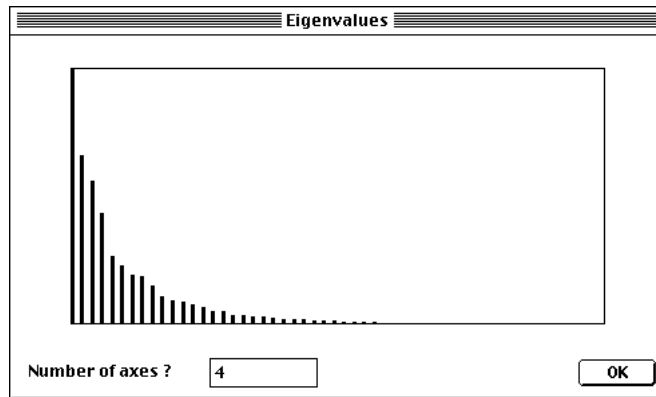
La vision de la corrélation inter-traits (RV de STATIS) sépare bien les deux cas. Entre ces deux tableaux, une distance considérable a été franchie par les biologistes qui les ont établis.

6 — Analyse factorielle multiple

L'analyse factorielle multiple compose un compromis par une analyse au mieux d'analyser un compromis constitué *a priori*. Elle peut donc proposer une solution (voir ⁸).



Multiple factor analysis
 Escofier & Pages 1984
 Input file TFTR
 Row number: 53, Column number: 83, Block number: 12



Selected weight option
 -> Total inertia from separate analysis

On décide de mettre l'inertie de chaque tableau à 1. Ceci signifie que la typologie induite par chaque trait a la même importance.

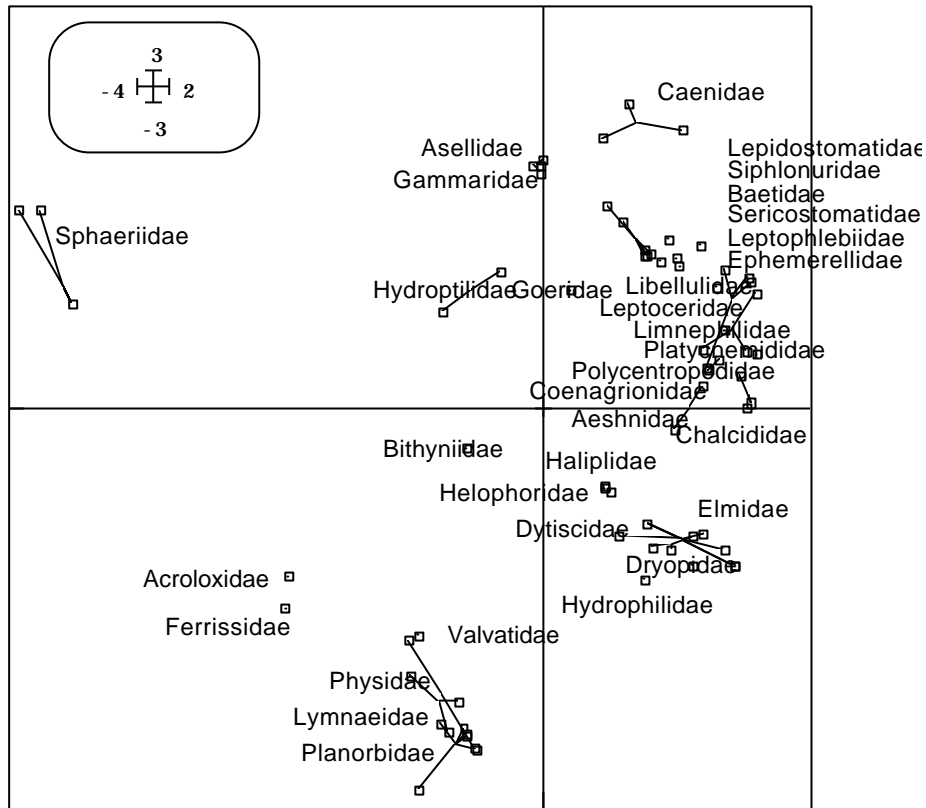
```
DiagoRC: General program for two diagonal inner product analysis
Input file: TFTR.++ta
--- Number of rows: 53, columns: 83
-----
Total inertia:      12
-----
```

On conserve clairement 4 facteurs. soit 59 % de la variabilité initiale.

```
File TFTR.++co contains the column scores
--- It has 83 rows and 4 columns
File :TFTR.++co
```

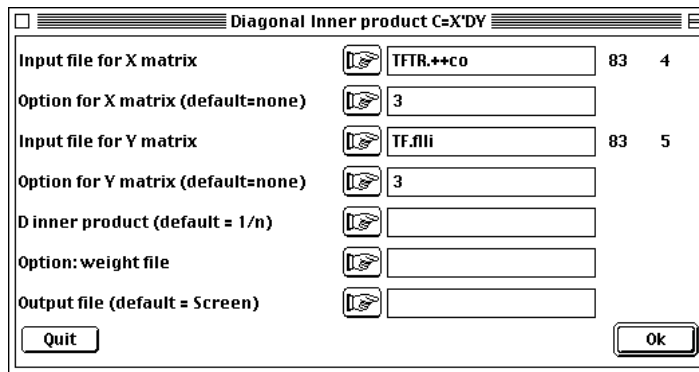
| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -3.920e+00 | 1.617e+00 |
| 2 | -2.879e+00 | 2.280e+00 |
| 3 | -2.775e+00 | 1.954e+00 |
| 4 | -2.296e+00 | 2.335e+00 |

Les scores sont de véritable coordonnées des espèces :



On obtient un résultat très voisin de celui de l'AFC, ce qui montre que, dans l'AFC floue, c'est la composante implicite K -tableau qui est favorable et non pas la composante proprement analyse des correspondances. Les scores de synthèse de l'AFMULT sont donc des candidats sérieux au statut de résumé concentré de l'information.

On notera encore les corrélations entre coordonnées proposées par l'AFMULT et coordonnées proposées par l'AFC floue :



```
X input file: TFTR.++co
--- Number of rows: 83, columns: 4
Y input file: TF.flli
--- Number of rows: 83, columns: 5
Diagonal inner product: uniform weight
XtDY output file: screen
--- Number of rows: 4, columns: 5
Input file: screen
--- Number of rows: 4, columns: 5
-----
```

```

[ 1] -9.6947e-01  1.0086e-01  5.3009e-02 -3.3721e-03  1.6103e-01
[ 2]  1.4079e-01  7.4811e-01  4.3848e-01  3.8995e-01  1.8124e-01
[ 3]  6.3073e-02  3.4543e-01  1.1914e-01 -8.4544e-01 -7.9511e-03
[ 4]  1.2883e-02  4.6168e-01 -8.1533e-01 -2.7615e-02  4.8779e-02

```

Un autre point de convergence est proposée par :

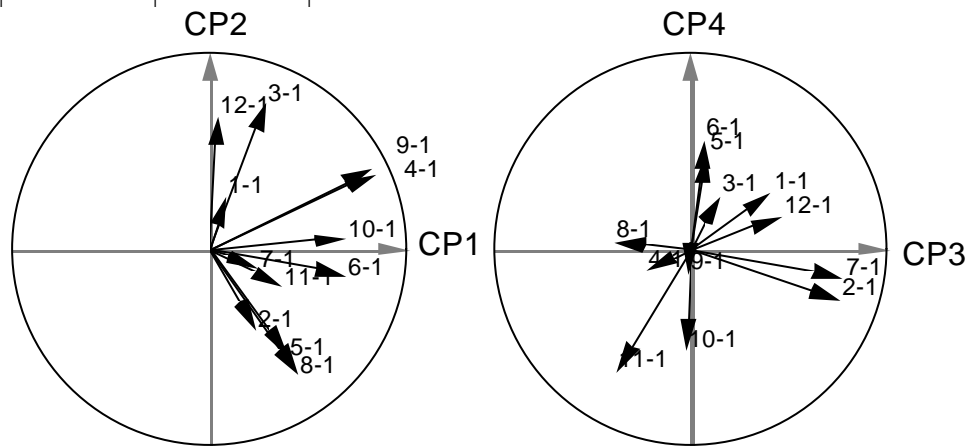
Projected inertia of variable groups

| | fac 1 | fac 2 | fac 3 | fac 4 |
|--------|------------------|-----------|-----------|-----------|
| Tab 1 | 2.981e-02 | 3.242e-02 | 6.698e-02 | 1.302e-01 |
| Tab 2 | 4.918e-02 | 1.509e-01 | 5.357e-01 | 6.332e-02 |
| Tab 3 | 1.038e-01 | 2.989e-01 | 4.341e-02 | 9.332e-02 |
| Tab 4 | 6.922e-01 | 1.416e-01 | 5.427e-02 | 1.241e-02 |
| Tab 5 | 1.933e-01 | 1.765e-01 | 6.277e-03 | 1.846e-01 |
| Tab 6 | 3.383e-01 | 2.719e-02 | 6.541e-03 | 2.173e-01 |
| Tab 7 | 1.978e-01 | 2.833e-02 | 3.455e-01 | 3.409e-02 |
| Tab 8 | 1.932e-01 | 3.338e-01 | 1.054e-01 | 1.394e-02 |
| Tab 9 | 4.590e-01 | 1.756e-01 | 5.207e-02 | 1.048e-02 |
| Tab 10 | 3.201e-01 | 5.112e-02 | 4.851e-02 | 2.282e-01 |
| Tab 11 | 7.158e-02 | 4.246e-02 | 8.284e-02 | 1.540e-01 |
| Tab 12 | 1.264e-02 | 2.969e-01 | 1.494e-01 | 1.870e-02 |
| Total | 2.661e+00 | 1.756e+00 | 1.497e+00 | 1.161e+00 |

Ces valeurs donnent le pourcentage de chaque inertie de départ (1) prise en compte par les scores de synthèse. 69% de l'inertie de la variable 4, 46% de celle de la variable 9, ... sont récupérées sur la première coordonnée.

File TFTR.++T4a contains scores of axes of separate analysis
 It has 48 rows and 4 columns
 It is to be used with --T4a.label and --T4a.cat files
 File :TFTR.++T4a

| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -7.166e-01 | 8.320e-01 |
| 2 | -6.339e-01 | 7.335e-01 |
| 3 | -3.829e-01 | 7.640e-01 |
| 4 | -6.129e-01 | 5.561e-01 |

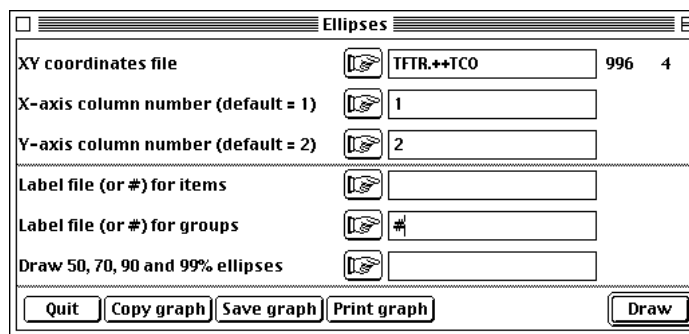


Cette figure est assez explicite. Les composantes principales de rang 1 de chacun des tableaux qui exprime l'essentiel de la variabilité de chaque trait ne forme pas des cônes de variables corrélées mais il faut quatre dimensions pour récupérer la position de l'ensemble. La non redondance des traits biologiques devient un fait majeur contenu dans le tableau de données.

File TFTR.++TCO contains columns scores issued from each table
 It has 996 rows and 4 columns
 It is to be used with --TCc.label and --TCc.cat files
 File :TFTR.++TCO

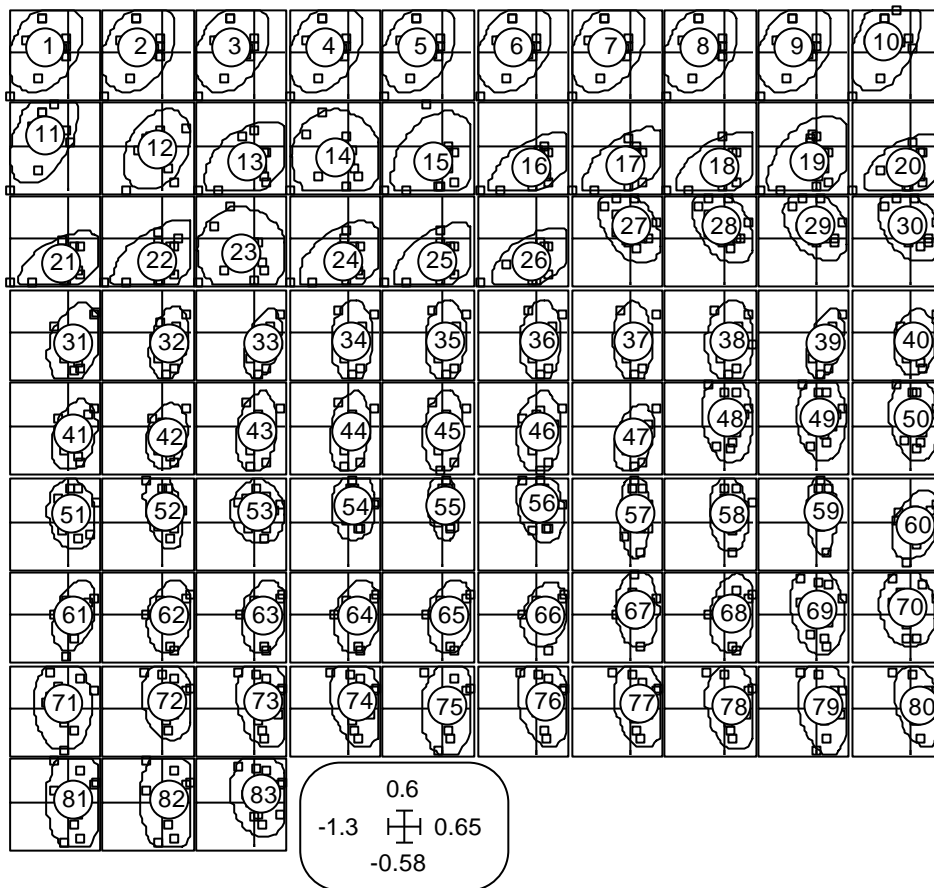
| Col. | Mini | Maxi |
|------|------------|-----------|
| 1 | -1.255e+00 | 6.402e-01 |
| 2 | -5.789e-01 | 5.973e-01 |
| 3 | -1.110e+00 | 7.311e-01 |
| 4 | -6.675e-01 | 6.295e-01 |

Chaque espèce est alors représentée par une position associée à chacun de ses traits :

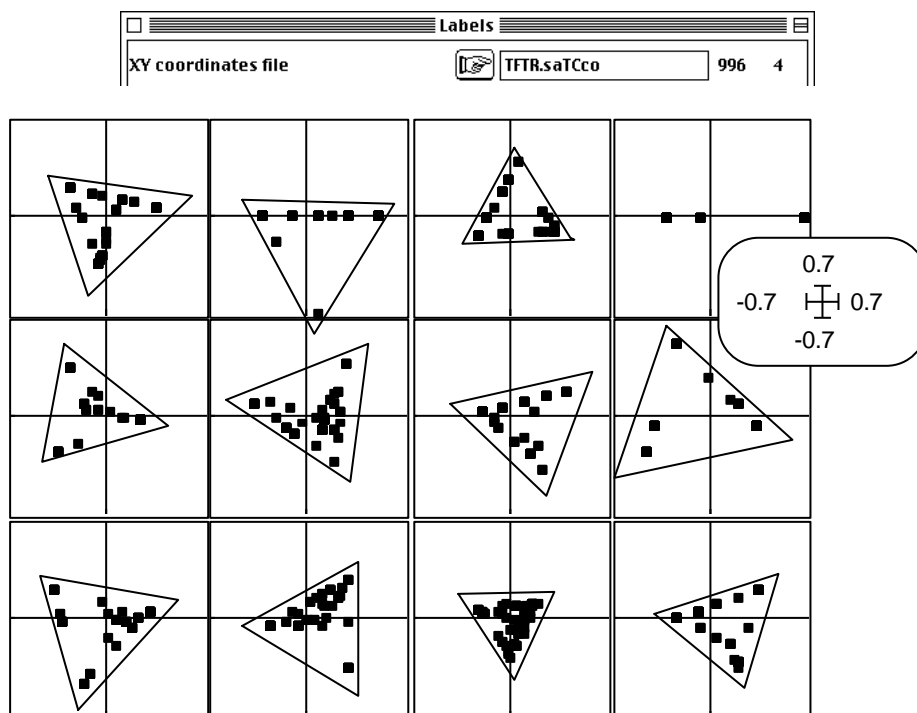


Le résultat est consigné ci-dessous.

L'ensemble souligne encore que, si des espèces peuvent être identique ou très voisines (l'essentiel de la variabilité se fait entre familles), deux espèces différentes ne le sont pas par une majorité de modalités. Des combinaisons de modalités peuvent coexister chez certaines espèces et être exclues chez d'autres, ce qui engendre une hétérogénéité bien répartie des positions d'un même taxon.

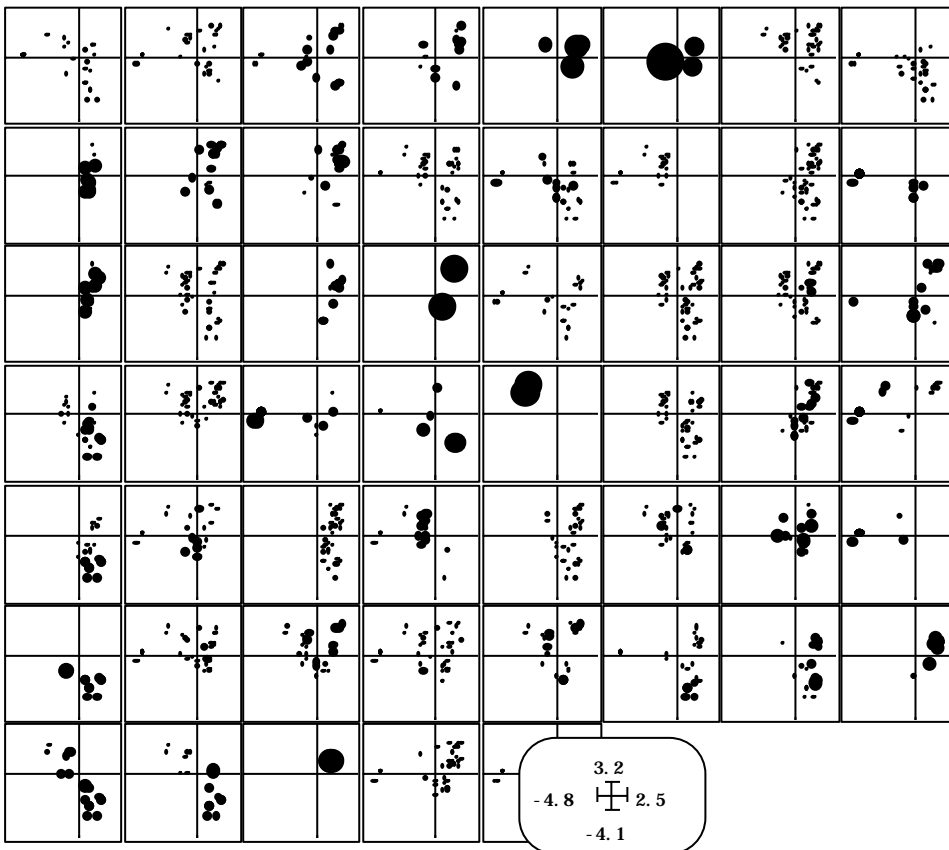
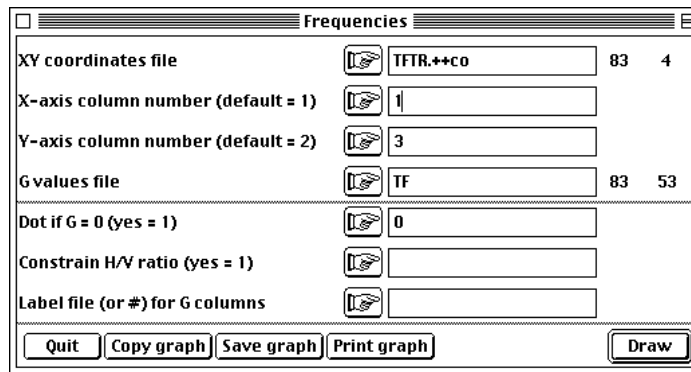


La complexité de la structure des stratégies biologiques est donc importante. Une variable biologique seule peut-elle être réduite à une variable quantitative : La réponse est clairement non. En effet :



Chaque analyse séparée (à l'exception de la variable 4) donne un nuage de point pratiquement inscrit dans un triangle. Ceci signifie qu'on peut en général trouver trois modalités et des taxons préférant l'une des trois. Il faut deux nombres au moins pour positionner une préférence entre trois catégories (représentation triangulaire dont l'archétype est argile-limon-sable). D'autre part ces analyses séparées sont peu cohérentes, ce qui veut dire qu'un choix dans un des triangles ne permet pas guère de prédire le choix fait dans un autre. De multiples combinaisons sont possibles.

Il est possible par l'AFC floue et l'AFMULT de réduire numériquement cette complexité avec 4 ou 5 variables numériques de synthèse. C'est une très bonne chose au plan statistique si on vise le couplage avec un tableau faunistique classique. Ce n'est pas une bonne chose au plan pratique comme en témoigne la figure obtenue par



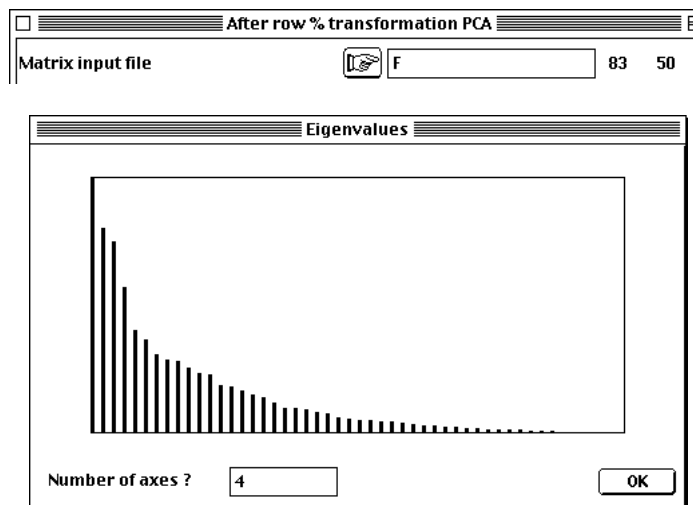
Si on veut utiliser ces variables de synthèse, il faudra les rapporter aux modalités d'origine pour les interpréter. On représente donc le mode d'utilisation de chaque modalité dans le plan de la typologie des espèces.

Autant la figure des analyses séparées est simple, autant celle-ci est complexe. La synthèse numérique est possible mais elle détruit l'information. Ce n'est pas totalement incompréhensible. Avec 12 sous-espace de \mathbf{R}^{83} qui totalise une dimension de 31 (43-12) il existe de bonnes variables rendant compte de l'ensemble des covariations mais ces variables sont alors liées aux données de base de manière très complexe.

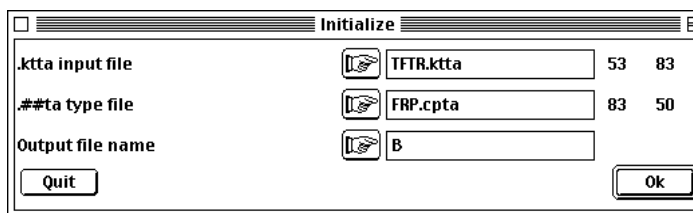
En bref, l'ensemble des variables biologique est essentiellement un K -tableau avec un faible compromis et une dimension irréductible de la plupart d'entre eux. En face d'un tableau faunistique, ce n'est pas un tableau qui se présente mais autant de tableaux qu'il y a de variables biologiques. Il est en effet probable que le lien se fera avec une ou plusieurs des variables mais non avec une synthèse du K -tableau. La question devient : pouvons nous repérer le ou lesquels de ces tableaux sont associés à la structure faunistique ?

7 — De K à $K + 1$ tableaux

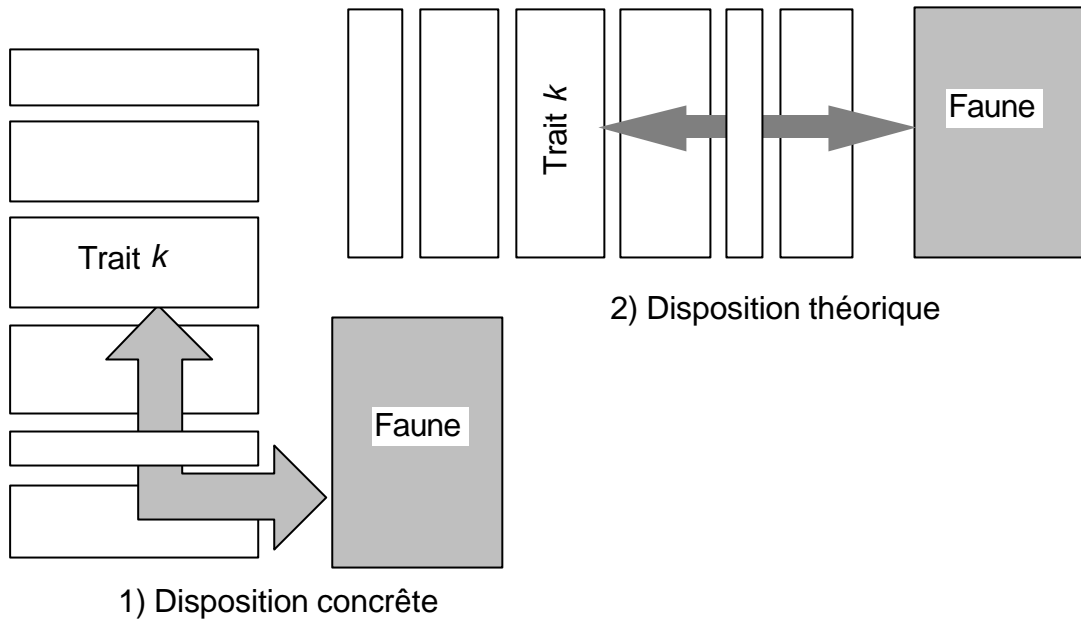
On fait dans ce sens un premier essai. Le tableau faunistique est traité par une ACP centrée dans laquelle chaque espèce est un profil de distribution inter-relevés :



On peut déjà être surpris par la présence de 4 axes utilisables. Le tableau faunistique est associé au K -tableau biologique ($K+1$: Initialize) :



Il se passe une chose fort simple à partir de la situation qu'on peut résumer par :



```

First Input File TFTR.ktta
---> rows: 53, columns: 83
Second Input File FRP.cpta
---> rows: 83, columns: 50

```

```

New ASCII file: B.k+1
It contains the input file names

```

```

New binary file: B.k+1ta. It contains the crossed K-table
It has 53 rows and 50 columns

```

Apparié par les espèces au tableau espèces-relevés, chaque trait définit une analyse de co-inertie. Il y a donc 12 analyses de co-inertie possibles. Ces 12 analyses de co-inertie commence donc par 12 analyses d'inertie des traits et une analyse d'inertie du tableau relevés. Elles sont directement exécutées dans l'option :

```

-----
| Xk - Separate inertia analysis |
-----

```

```

Table n°:  1| Lig:  5| Col:  83| Rank:  4|
Total inertia: 1.140e-01| Squared norm (H.-S.): 4.129e-03
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +4.6345E-02 +0.4064 +0.4064 |02  +3.7927E-02 +0.3326 +0.7389 |
03  +2.1933E-02 +0.1923 +0.9313 |04  +7.8395E-03 +0.0687 +1.0000 |
05  +0.0000E+00 +0.0000 +1.0000

```

```

-----
Table n°:  2| Lig:  3| Col:  83| Rank:  2|
Total inertia: 1.054e-01| Squared norm (H.-S.): 9.797e-03
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +9.8760E-02 +0.9372 +0.9372 |02  +6.6132E-03 +0.0628 +1.0000 |
03  +0.0000E+00 +0.0000 +1.0000

```

```

-----
Table n°:  3| Lig:  5| Col:  83| Rank:  4|
Total inertia: 8.334e-02| Squared norm (H.-S.): 2.274e-03
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +3.9518E-02 +0.4742 +0.4742 |02  +2.0295E-02 +0.2435 +0.7177 |
03  +1.5176E-02 +0.1821 +0.8999 |04  +8.3459E-03 +0.1001 +1.0000 |
05  +0.0000E+00 +0.0000 +1.0000

```

Table n°: 4 | Lig: 2 | Col: 83 | Rank: 1 |
Total inertia: 2.038e-01 | Squared norm (H.-S.): 4.153e-02
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +2.0379E-01 +1.0000 +1.0000 | 02 +0.0000E+00 +0.0000 +1.0000 |

Table n°: 5 | Lig: 5 | Col: 83 | Rank: 4 |
Total inertia: 1.161e-01 | Squared norm (H.-S.): 5.216e-03
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +6.2719E-02 +0.5401 +0.5401 | 02 +3.1581E-02 +0.2720 +0.8121 |
03 +1.5774E-02 +0.1358 +0.9480 | 04 +6.0414E-03 +0.0520 +1.0000 |
05 +0.0000E+00 +0.0000 +1.0000

Table n°: 6 | Lig: 4 | Col: 83 | Rank: 3 |
Total inertia: 6.089e-02 | Squared norm (H.-S.): 1.978e-03
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +4.1291E-02 +0.6782 +0.6782 | 02 +1.6158E-02 +0.2654 +0.9435 |
03 +3.4373E-03 +0.0565 +1.0000 | 04 +0.0000E+00 +0.0000 +1.0000 |

Table n°: 7 | Lig: 5 | Col: 83 | Rank: 4 |
Total inertia: 6.708e-02 | Squared norm (H.-S.): 1.992e-03
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +3.8854E-02 +0.5792 +0.5792 | 02 +2.0774E-02 +0.3097 +0.8889 |
03 +7.0905E-03 +0.1057 +0.9946 | 04 +3.6406E-04 +0.0054 +1.0000 |
05 +0.0000E+00 +0.0000 +1.0000

Table n°: 8 | Lig: 3 | Col: 83 | Rank: 2 |
Total inertia: 1.671e-01 | Squared norm (H.-S.): 1.650e-02
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +1.1913E-01 +0.7127 +0.7127 | 02 +4.8015E-02 +0.2873 +1.0000 |
03 +0.0000E+00 +0.0000 +1.0000

Table n°: 9 | Lig: 4 | Col: 83 | Rank: 3 |
Total inertia: 1.040e-01 | Squared norm (H.-S.): 5.757e-03
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +7.1612E-02 +0.6888 +0.6888 | 02 +2.3450E-02 +0.2256 +0.9144 |
03 +8.9005E-03 +0.0856 +1.0000 | 04 +0.0000E+00 +0.0000 +1.0000 |

Table n°: 10 | Lig: 5 | Col: 83 | Rank: 4 |
Total inertia: 9.353e-02 | Squared norm (H.-S.): 3.005e-03
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +4.4658E-02 +0.4775 +0.4775 | 02 +2.8159E-02 +0.3011 +0.7785 |
03 +1.1585E-02 +0.1239 +0.9024 | 04 +9.1318E-03 +0.0976 +1.0000 |
05 +0.0000E+00 +0.0000 +1.0000

Table n°: 11 | Lig: 7 | Col: 83 | Rank: 6 |
Total inertia: 4.054e-02 | Squared norm (H.-S.): 4.412e-04
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +1.5798E-02 +0.3897 +0.3897 | 02 +1.1262E-02 +0.2778 +0.6674 |
03 +6.9341E-03 +0.1710 +0.8385 | 04 +3.2009E-03 +0.0789 +0.9174 |
05 +2.3490E-03 +0.0579 +0.9753 | 06 +9.9964E-04 +0.0247 +1.0000 |
07 +0.0000E+00 +0.0000 +1.0000

Table n°: 12 | Lig: 5 | Col: 83 | Rank: 4 |
Total inertia: 7.579e-02 | Squared norm (H.-S.): 2.730e-03
Num. Eigenval. R.Iner. R.Sum | Num. Eigenval. R.Iner. R.Sum |
01 +4.6126E-02 +0.6086 +0.6086 | 02 +2.4229E-02 +0.3197 +0.9283 |
03 +3.2838E-03 +0.0433 +0.9716 | 04 +2.1521E-03 +0.0284 +1.0000 |
05 +0.0000E+00 +0.0000 +1.0000

Eigenvalues are stored by column into file TFTR.siavp
Row : 7 - Col (tables): 12

File TFTR.siatLlax contains normed row scores (principal axes)
in the separate inertia analyses
It has 53 rows and 4 columns

File TFTR.siatCcax contains normed column scores (principal components)
in the separate inertia analyses
It has 996 rows and 4 columns

File TFTR.siapa contains rank, total inertia and squared norm of Hilbert-Schmidt
for each separate analyses
It has 12 rows and 3 columns

L'analyse du tableau unique confronté au 12 traits donne :

```
-----  
| Y - Separate inertia analysis |  
-----  
Input file: FRP.cpta  
--- Number of rows: 83, columns: 50  
-----  
Total inertia: 9.101e-02  
-----  
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |  
01  +1.2175E-02 +0.1338 +0.1338 | 02  +9.7495E-03 +0.1071 +0.2409 |  
03  +9.1103E-03 +0.1001 +0.3410 | 04  +6.9752E-03 +0.0766 +0.4177 |  
05  +4.9292E-03 +0.0542 +0.4718 | 06  +4.4629E-03 +0.0490 +0.5209 |  
...  
File FRP.cvpv contains the eigenvalues and relative inertia for each axis  
--- It has 50 rows and 2 columns  
File FRP.cpcax contains the normed column scores (principal axes)  
--- It has 50 rows and 4 columns  
File FRP.cplax contains the normed row scores (principal component)  
--- It has 83 rows and 4 columns
```

Suivent alors 12 analyses de co-inertie séparées dont on conserve au plus 4 axes si c'est possible :

```
-----  
| Separate co-inertia analysis |  
-----  
Table n°: 1 | Lig: 5 | Col: 50 | Rank: 4 |  
Total co-inertia: 1.376e-04 | RV: 9.445e-02 |  
-----  
| Num | Eigenval. | Covaria. | Ratio | Sum ratio |  
-----  
| 1 | 6.376e-05 | 7.985e-03 | 4.633e-01 | 4.633e-01 |  
| 2 | 4.456e-05 | 6.676e-03 | 3.238e-01 | 7.871e-01 |  
| 3 | 1.840e-05 | 4.290e-03 | 1.337e-01 | 9.208e-01 |  
| 4 | 1.090e-05 | 3.301e-03 | 7.917e-02 | 1.000e+00 |  
-----  
Table n°: 2 | Lig: 3 | Col: 50 | Rank: 2 |  
Total co-inertia: 1.570e-04 | RV: 6.996e-02 |
```

```
-----
|Num|Eigenval.| Covaria.| Ratio |Sum ratio|
-----
| 1|1.546e-04|1.243e-02|9.845e-01|9.845e-01|
| 2|2.433e-06|1.560e-03|1.549e-02|1.000e+00|
-----
```

Table n°: 3 | Lig: 5 | Col: 50 | Rank: 4 |
Total co-inertia: 2.109e-04 | RV: **1.951e-01** |

```
-----
|Num|Eigenval.| Covaria.| Ratio |Sum ratio|
-----
| 1|1.656e-04|1.287e-02|7.852e-01|7.852e-01|
| 2|2.706e-05|5.202e-03|1.283e-01|9.135e-01|
| 3|1.343e-05|3.664e-03|6.366e-02|9.772e-01|
| 4|4.816e-06|2.195e-03|2.284e-02|1.000e+00|
-----
```

Table n°: 4 | Lig: 2 | Col: 50 | Rank: 1 |
Total co-inertia: 3.698e-04 | RV: 8.004e-02 |

```
-----
|Num|Eigenval.| Covaria.| Ratio |Sum ratio|
-----
| 1|3.698e-04|1.923e-02|1.000e+00|1.000e+00|
-----
```

Table n°: 5 | Lig: 5 | Col: 50 | Rank: 4 |
Total co-inertia: 1.250e-04 | RV: 7.631e-02 |

```
-----
| 1|7.097e-05|8.425e-03|5.679e-01|5.679e-01|
| 2|3.936e-05|6.274e-03|3.149e-01|8.828e-01|
...
-----
```

Table n°: 6 | Lig: 4 | Col: 50 | Rank: 3 |
Total co-inertia: 9.953e-05 | RV: 9.870e-02 |

```
-----
| 1|7.855e-05|8.863e-03|7.891e-01|7.891e-01|
| 2|1.828e-05|4.275e-03|1.836e-01|9.727e-01|
...
-----
```

Table n°: 7 | Lig: 5 | Col: 50 | Rank: 4 |
Total co-inertia: 1.195e-04 | RV: 1.181e-01 |

```
-----
| 1|6.778e-05|8.233e-03|5.673e-01|5.673e-01|
| 2|3.028e-05|5.503e-03|2.534e-01|8.208e-01|
...
-----
```

Table n°: 8 | Lig: 3 | Col: 50 | Rank: 2 |
Total co-inertia: 3.157e-04 | RV: 1.084e-01 |

```
-----
| 1|2.676e-04|1.636e-02|8.479e-01|8.479e-01|
| 2|4.802e-05|6.930e-03|1.521e-01|1.000e+00|
-----
```

Table n°: 9 | Lig: 4 | Col: 50 | Rank: 3 |
Total co-inertia: 1.913e-04 | RV: 1.112e-01 |

```
-----
| 1|1.136e-04|1.066e-02|5.937e-01|5.937e-01|
| 2|5.505e-05|7.419e-03|2.878e-01|8.815e-01|
-----
```


...

Table n°: 10 | Lig: 5 | Col: 50 | Rank: 4 |
Total co-inertia: 1.321e-04 | RV: 1.062e-01 |

| 1 | 6.765e-05 | 8.225e-03 | 5.123e-01 | 5.123e-01 |
| 2 | 3.910e-05 | 6.253e-03 | 2.961e-01 | 8.083e-01 |

...

Table n°: 11 | Lig: 7 | Col: 50 | Rank: 6 |
Total co-inertia: 6.909e-05 | RV: 1.450e-01 |

| 1 | 4.394e-05 | 6.629e-03 | 6.360e-01 | 6.360e-01 |
| 2 | 1.477e-05 | 3.843e-03 | 2.138e-01 | 8.498e-01 |

...

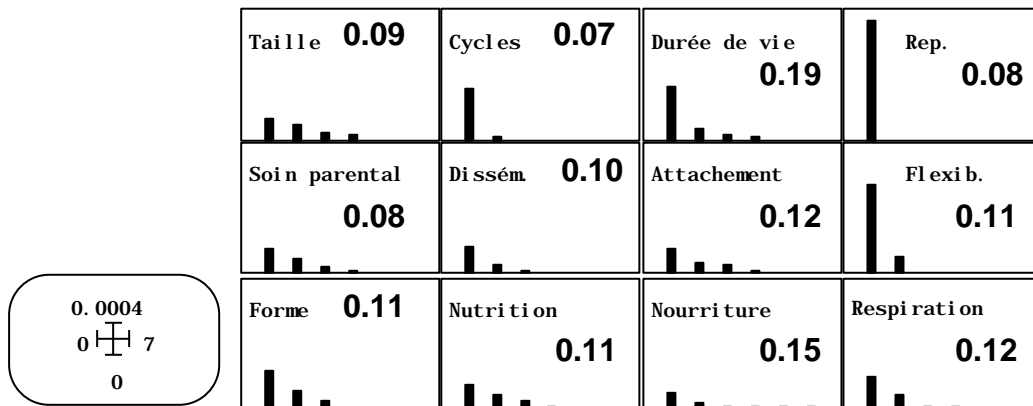
Table n°: 12 | Lig: 5 | Col: 50 | Rank: 4 |
Total co-inertia: 1.385e-04 | RV: 1.169e-01 |

| 1 | 9.945e-05 | 9.972e-03 | 7.181e-01 | 7.181e-01 |
| 2 | 3.400e-05 | 5.831e-03 | 2.455e-01 | 9.636e-01 |

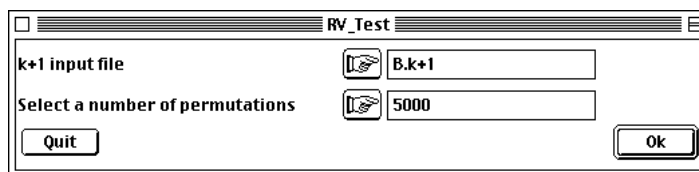
...

Eigenvalues (separate co-inertia) are stored by column into file B.k+1vp
Row : 7 - Col (tables): 12

File B.k+1lr contains rank, total co-inertia and RV
for each separate co-inertia analyse
It has 12 rows and 3 columns



Un test de signification s'impose. En effet ces graphes de valeurs propres sont peu comparables entre eux puisqu'il dépendent de l'inertie propre du tableau considéré. Les RV reportés sur la figure sont comparable entre eux. Un test doit dire lesquels sont significatifs.

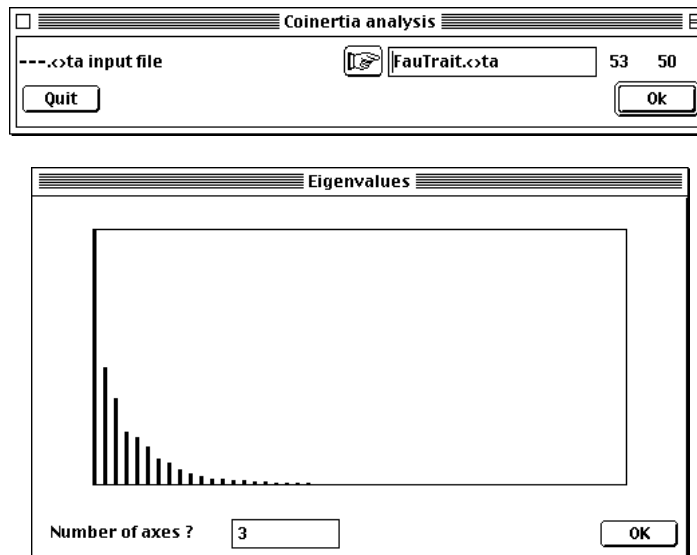


X = Sum of squared RV
number of random permutations: 5000 Observed: 1.579e-01

est plus nettement extraordinaire que le lien direct avec tout le tableau :

$$RV(\mathbf{Y}, [\mathbf{X}_1, \dots, \mathbf{X}_p])$$

Cette discussion est entièrement ouverte dans le domaine multi-tableau. Or la co-inertie est complexe avec un premier axe très important :



On a donc intérêt à décomposer le lien biologie -écologie par trait. On mesure ici la nature des innovations statistiques nécessaires à la manipulation des traits biologiques.

8 — Conclusion

On notera pour conclure cette première approche les points suivants.

1) L'AFC floue est une méthode de synthèse des tableaux de traits biologiques qui a l'avantage de la simplicité. C'est un premier pas dont la propriété principale d'AFC internes implicite dans la direction de la statistique multi-tableau.

2) La métrique du Khi2 qu'elle sous-tend est peu favorable à la mesure de la diversité et l'ACP centrée simple des tableaux de variables floues est plus simple de ce point de vue.

3) On peut avoir les deux en pratiquant l'analyse factorielle multiple qui ne soulève pas de difficultés particulières. Pour un examen exhaustif des possibilités, il faudra ajouter l'analyse canonique généralisée⁹.

4) La réduction d'un tableau de traits par variables floues peut ou non être une bonne chose. Elle est judicieuse en cas de redondance entre traits, elle ne l'est pas dans le cas contraire. Le tableau analysé ici a une structure caractérisée par la très grande indépendance des traits biologiques entre eux. Cela veut dire que sur plusieurs critères les taxons ont utilisé des combinaisons de modalités très diversifiées, chaque variable induisant une typologie d'espèce originale par rapport aux autres. Pour avoir un avis rapide sur ce point crucial, STATIS (opérateurs), dans sa partie inter-structure est efficace.

5) Indépendance ou redondance des traits influe fortement sur le lien avec un tableau écologique (sites-espèces). Si la réduction du tableau de traits est complexe, le lien avec un autre tableau le sera plus encore. Si globalement les méthodes de compromis entre tableaux-traits sont prises en défaut, les analyses de compromis entre liens comme l'analyse des concordances ¹⁰ disponible dans le module K+1 le seront plus encore. Il n'en a pas été fait état ici par absence de nécessité.

6) On peut alors prévoir qu'une bonne stratégie consiste à décomposer la partie biologique en morceaux indépendants contenant un seul trait ou plusieurs traits covariants, avec élimination des traits n'ayant pas d'impacts écologiques. La nécessité de méthodes de typologie de tableaux est ici très forte. C'est un problème ouvert.

Références

¹ Chevenet, F., Dolédec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309.

² Light, R.J. & Margolin, B.H. (1971) An analysis of variance for categorical data. *Journal of the American Statistical Association* : 66, 534-544.

³ Lebart, L. (1969) Analyse statistique de la contiguïté. *Publication de l'Institut de Statistiques de l'Université de Paris* : 28, 81-112.

⁴ Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* : 76, 5-13.

⁵ Ter Braak, C.J.F. (1983) Principal components biplots and alpha and beta diversity. *Ecology* : 64, 3, 454-462.

⁶ Cazes, P., Chessel, D. & Dolédec, S. (1988) L'analyse des correspondances internes d'un tableau partitionné : son usage en hydrobiologie. *Revue de Statistique Appliquée* : 36, 39-54.

⁷ Bournaud, M., Richoux, P. & Usseglio-Polatera, P. (1992a) An approach to the synthesis of qualitative ecological information from aquatic coleoptera communities. *Regulated rivers: Research and Management* : 7, 165-180.

⁸ Chessel, D. & Hanafi, M. (1996) Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* : 44, 35-60.

⁹ Carrol, J.D. (1968) A generalization of canonical correlation analysis to three or more sets of variables. *Proceeding of the 76th Convention of the American Psychological Association* : 3, 227-228.

¹⁰ Lafosse, R. & Hanafi, M. (1997) Concordance d'un tableau avec K tableaux : définition de $K+1$ uples synthétiques. *Revue de Statistique Appliquée* : (sous presse).