

Fiche de Biostatistique

Méthodes K-tableaux

D. Chessel, A.B. Dufour & J. Thioulouse

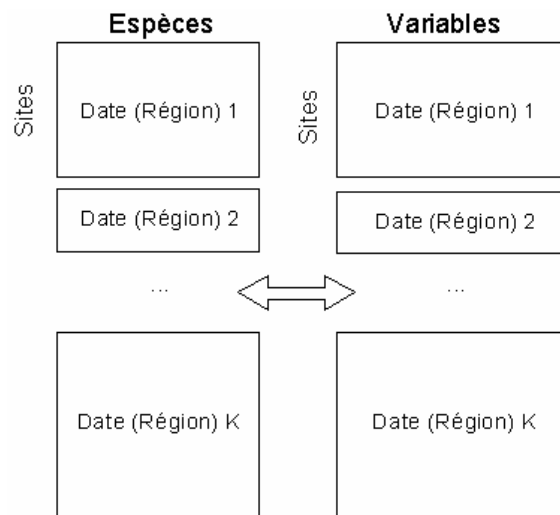
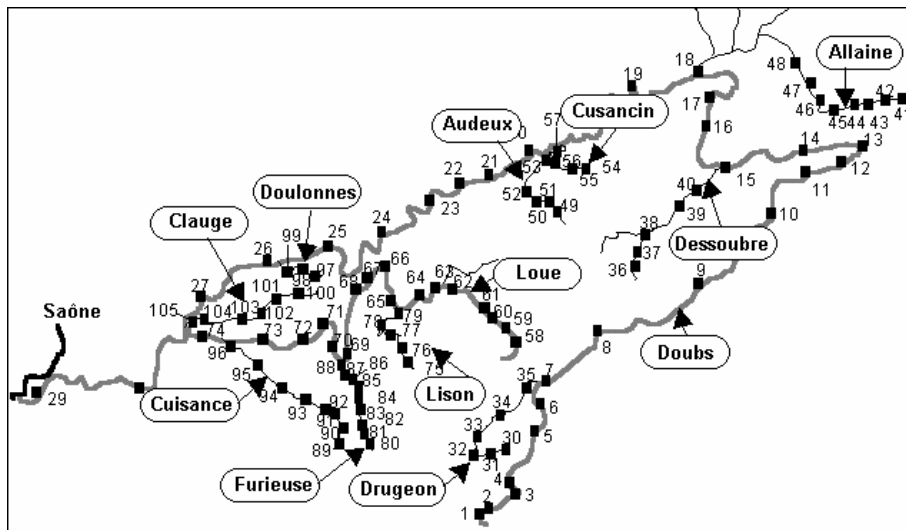
Résumé

La fiche donne des indications sur plusieurs méthodes d'analyse de données à K tableaux et des illustrations sur leur usage.

Plan

1.	INTRODUCTION.....	2
2.	TABLEAU COMME VARIABLE	7
	2.1. Analyse des correspondances multiples.....	7
	2.2. Analyse canonique généralisée	11
	2.3. Analyse factorielle multiple.....	14
	2.4. Variables floues	20
3.	TABLEAU COMME INDIVIDU	22
	3.1. Analyses inter-classe et analyses discriminantes	22
	3.2. Intra-classes.....	22
4.	TABLEAU COMME STRUCTURE.....	24
	4.1. Analyses triadiques partielles	24
	4.2. STATIS sur les tableaux	28
	4.3. STATIS sur les opérateurs.....	37
	4.4. Analyses de co-inertie multiple	48
	4.5. Extensions.....	54
5.	REFERENCES.....	56

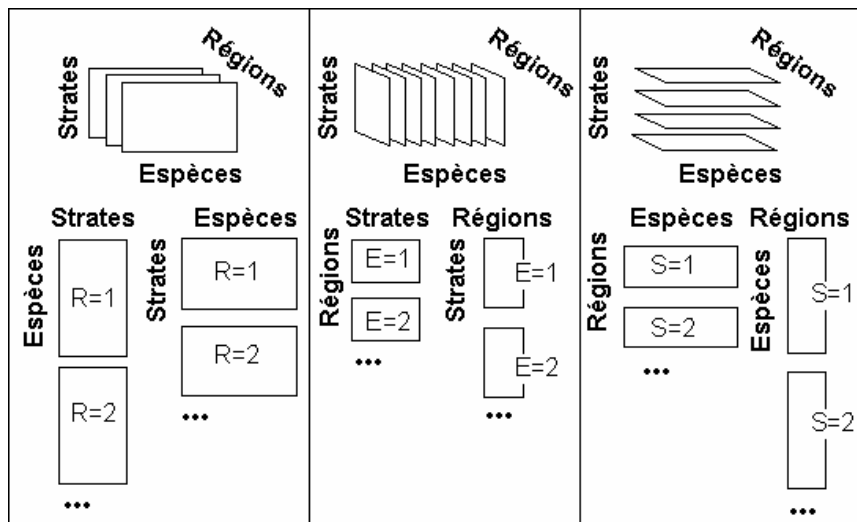
Il peut représenter le problème "mêmes individus et paquets distincts de variables". Le second représente la question de la reproduction d'une région à l'autre ou d'une période à l'autre de la structure ou de la co-structure écologique :



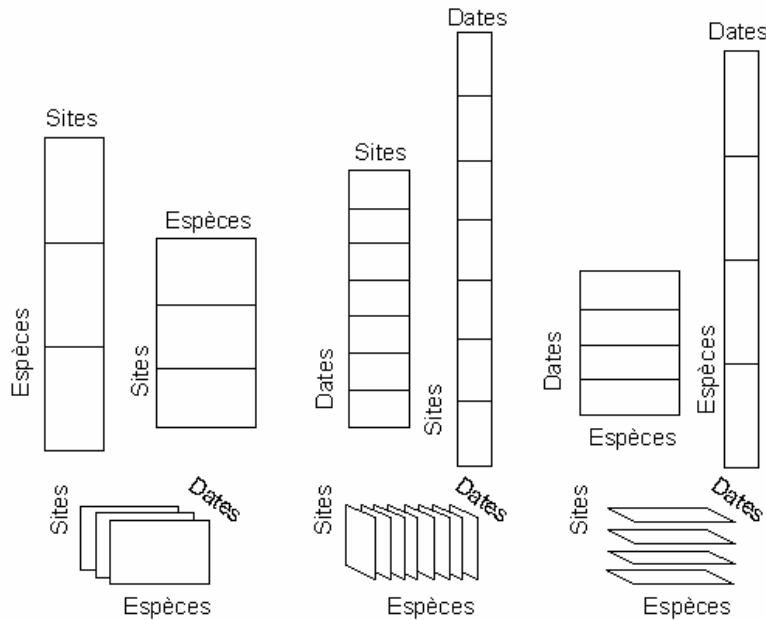
Il peut représenter le problème "mêmes variables et paquets distincts d'individus". Ces données sont exemplaires parce qu'elles représentent une multitude de situations écologiques, ce sont des exemples, et parce qu'elles sont parfaitement disponibles, ce sont des modèles. Les premières représentent la question « $K+1$ » et la seconde la question « $2K$ ». On s'intéressera pour commencer à la question « K ».

On voit immédiatement que la configuration "mêmes variables et mêmes individus" se présente naturellement. Dans ce champ, les cubes de données jouent un rôle central en exprimant la difficulté immédiate de la multiplicité des points de vue.

Les données de J. Blondel et H. Farré³ sont un troisième cas d'école, parce qu'elles ont été publiées en toute transparence. Il s'agit de mesurer la variabilité du cortège avifaunistique entre 4 régions (Pologne, Bourgogne, Provence et Corse), le long du gradient de fermeture de la végétation vu par six strates d'échantillonnage (1- végétation buissonnante basse (hauteur < 1 m) à 6- forêts de plus de 20 m de hauteur).



A défaut de méthodes appropriées, on peut toujours présenter les données dans une analyse à un tableau. La question qui vient immédiatement est lequel :



Il y a, en effet, 6 manières de faire d'un cube un tableau pour se ramener au cas précédent et on a peu de chances de tomber sur la bonne au hasard.

Au plan théorique, on voit donc que trois situations se présentent. K tableaux peuvent avoir les mêmes individus :

$$(X_1, Q_1, D) \quad (X_2, Q_2, D) \cdots (X_K, Q_K, D)$$

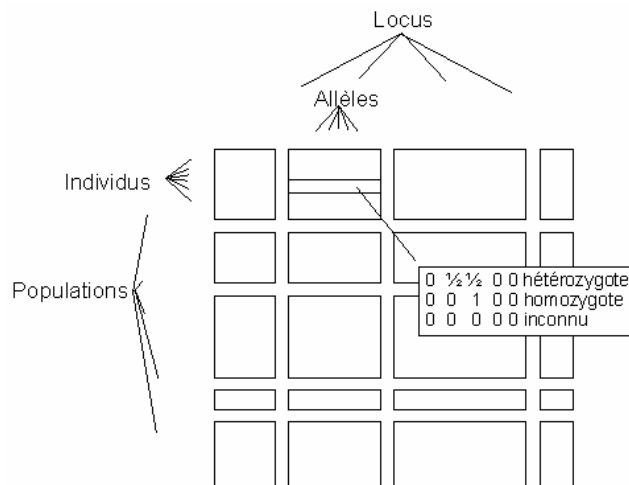
ou les mêmes variables :

$$(X_1, Q, D_1) \quad (X_2, Q, D_2) \cdots (X_K, Q, D_K)$$

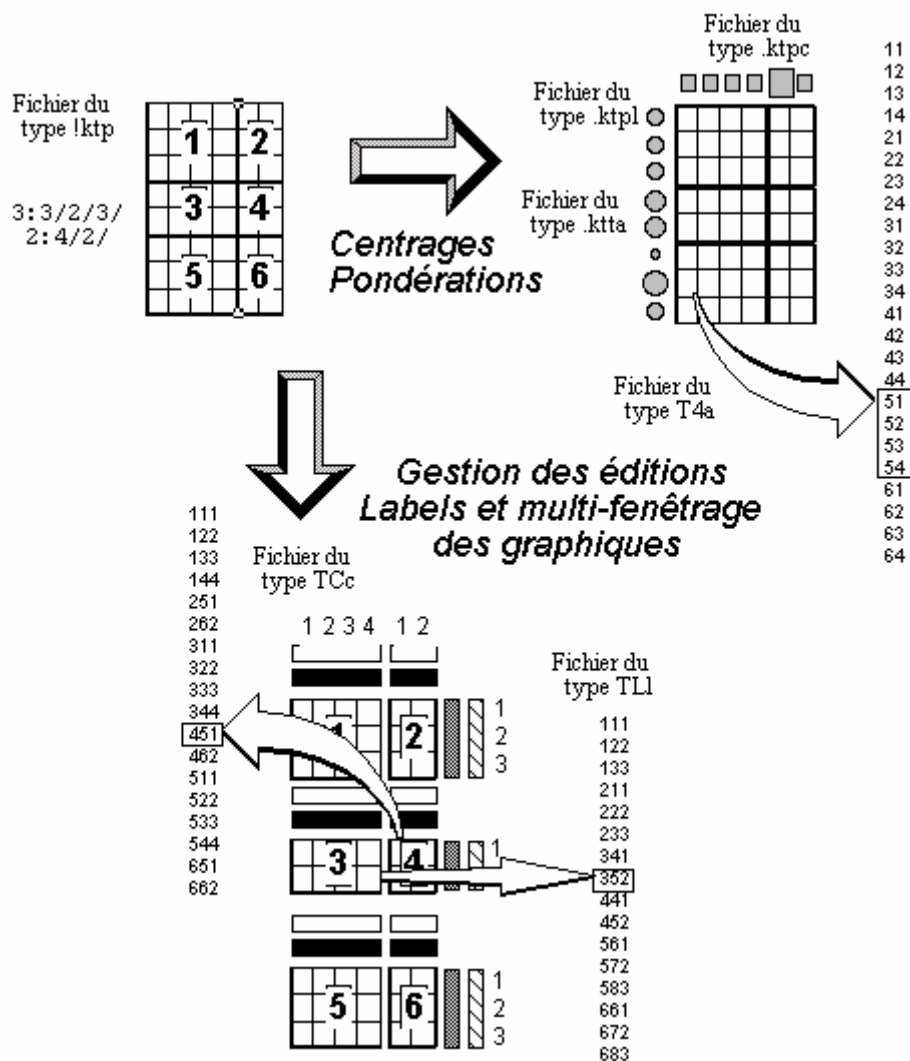
ou les mêmes marges :

$$(X_1, Q, D) \quad (X_2, Q, D) \cdots (X_K, Q, D)$$

La génétique propose même une structure doublement K -tableaux :



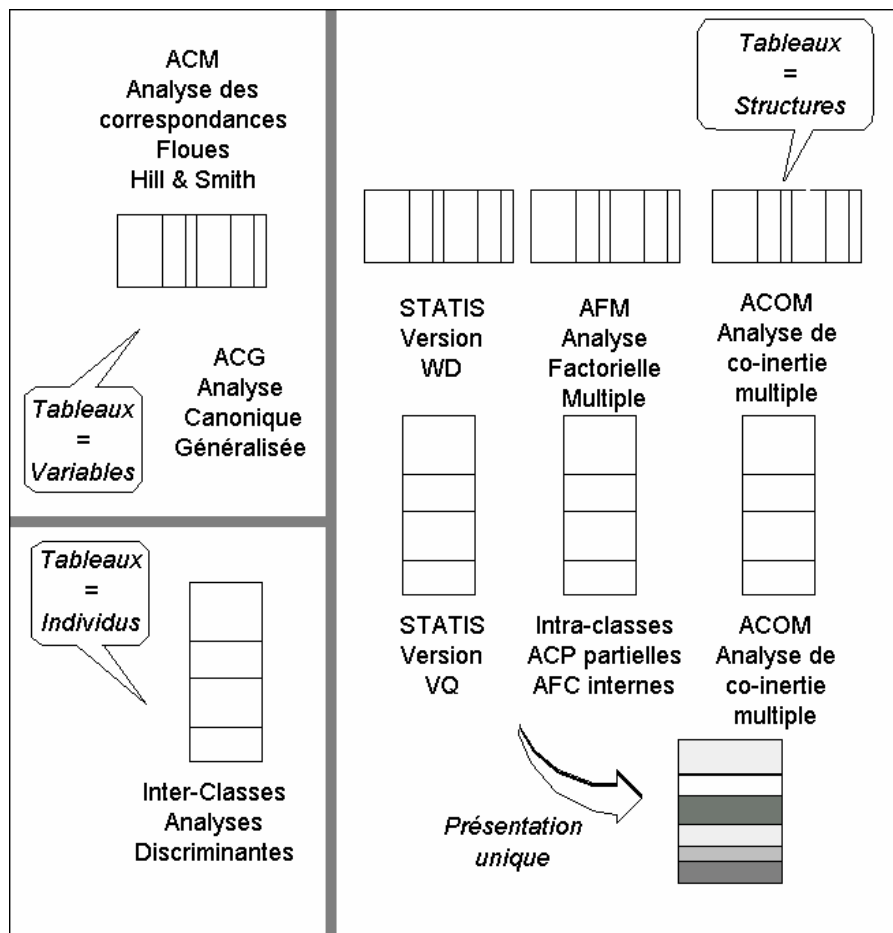
Ces questions soulèvent des problèmes théoriques et des questions pratiques. Dans ADE-4, un effort de mise en forme est exprimé dans la figure suivante :



Le code employé pour les noms de fichiers contient T pour numéro de tableau, L et C pour numéro de lignes et de colonnes comptés globalement, l et c pour numéro de lignes et de colonnes comptés par tableaux. Pour un K -tableaux simple, les tableaux sont les uns sous les autres, mais le paramétrage permet de se repérer dans les K -tableaux doubles.

On utilisera des indicateurs de blocs ici $3 : 3/2/3$ pour les lignes et $1 : 4/2$ pour indiquer que le tableau à 8 lignes et 6 colonnes est un K -tableau à 3 blocs de lignes et 2 blocs de colonnes. Les fichiers du type TCc contiendront l'information associée à chaque colonne de chaque tableau et les fichiers du type TLI contiendront l'information associée à chaque ligne de chaque tableau.

Ceci est purement formel et ne préjuge en rien de ce qu'on veut faire.



Points de repères

On peut introduire la question en disant que globalement

- un tableau peut être une « variable » en ce sens que plusieurs colonnes peuvent être nécessaires pour enregistrer l'information d'un type donné ;
- un tableau peut être un « individu » en ce sens que plusieurs lignes peuvent être nécessaires pour enregistrer l'information d'un type donné ;
- un tableau peut être une structure en ce sens qu'il définit à lui seul une typologie de variables et/ou d'individus.

Ainsi une distribution de granulométrie donnée par 20% Argile, 70% Limon et 10% Sable prendra 3 colonnes mais ne contiendra qu'une variable. Au contraire, une mesure de p variables sur n individus appartenant à la même population sera multivariée (chaque colonne est utile indépendamment des autres) mais ne représentera qu'un individu (une population, un groupe, un échantillon). Au contraire, enfin, un tableau faunistique représente un mode de dispersion des taxons entre les stations, mais rien n'empêche de reproduire la mesure à une autre date et le tableau représente alors une structure. Rien ne se ressemblent moins sur le fond que deux K -tableaux de nature différente, alors que formellement ce sont des objets très proches.

La principale difficulté vient d'un argument théorique. Il y a plusieurs méthodes associées à un seul tableau qui relèvent d'un modèle unique qui résout plusieurs problèmes à l'aide d'un seul algorithme. Cette unicité est perdue dans un couple de tableaux, plus encore avec K tableaux. La méthode n'existe pas et n'existera jamais.

2. Tableau comme variable

2.1. Analyse des correspondances multiples

L'analyse des correspondances multiples est la plus simple des méthodes K -tableaux mais elle pose parfaitement le problème des tableaux comme variables.

Un tableau de variables qualitatives (carte Meteq) indique pour chaque individu et chaque variable le numéro de la modalité utilisée :

```
2132333
2111233
1122233
1111233 ...
```

On peut le voir (CategVar: Categ->Disj) comme la juxtaposition de paquets d'indicatrices des classes :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	0	1	0	1	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1
2	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1
3	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	1
4	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1
5	1	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0
6	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0
7	0	0	1	0	1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	1	0
8	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0
9	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	0
10	0	1	0	1	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0
11	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	0
12	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	0	1	0	0	1	0
13	0	1	0	1	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1	0
14	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	0
15	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0
16	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0
17	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0
18	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0
19	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
20	0	1	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0
21	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0
22	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0
23	0	1	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0

Chaque tableau fait une typologie simple des individus (partition). L'objectif de l'ACM est de donner un point de vue coordonné sur toutes ces partitions. On note :

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_v]$$

le tableau disjonctif complet.

v est le nombre de variables qualitatives. n est le nombre d'individus (lignes) et m_1, m_2, \dots, m_v sont les nombres de modalités de chaque variable. Le nombre de modalités total est $m = m_1 + m_2 + \dots + m_v$. On peut considérer que l'ACM fait partie des analyses à un tableau (\mathbf{X}) mais son fonctionnement est une porte d'entrée dans la sphère K -tableaux et c'est pourquoi elle est présentée ici. On suppose que chaque ligne porte le poids $1/n$ (comme dans une ACP) ou le poids p_i ($\sum_{i=1}^n p_i = 1$). \mathbf{D} est comme d'habitude la diagonale des poids des lignes (matrice $n \times n$). On compte pour chaque modalité le nombre d'individus l'utilisant (le nombre de porteurs) en faisant :

$$\mathbf{X}' \mathbf{1}_n = [n_{11}, n_{12}, \dots, n_{1m_1}, n_{21}, n_{22}, \dots, n_{2m_2}, \dots, n_{v1}, n_{v2}, \dots, n_{vm_v}]$$

Il vient simplement que :

$$n_{11} + n_{12} + \dots + n_{1m_1} = n_{21} + n_{22} + \dots + n_{2m_2} = \dots = n_{v1} + n_{v2} + \dots + n_{vm_v} = n$$

Au lieu de compter les porteurs d'une modalité, on peut aussi sommer les poids des porteurs de cette modalité. Pour une pondération uniforme, le poids d'une modalité est la fréquence ($f_{ij} = n_{ij}/n$) du nombre de porteurs et on a :

$$f_{11} + f_{12} + \dots + f_{1m_1} = f_{21} + f_{22} + \dots + f_{2m_2} = \dots = f_{v1} + f_{v2} + \dots + f_{vm_v} = 1$$

mais en général, les poids des modalités sont dans :

$$[f_{11}, f_{12}, \dots, f_{1m_1}, f_{21}, f_{22}, \dots, f_{2m_2}, \dots, f_{v1}, f_{v2}, \dots, f_{vm_v}] = \mathbf{X}' \mathbf{D} \mathbf{1}_n$$

Les poids des modalités sont écrits sur la diagonale d'une matrice diagonale :

$$\mathbf{D}_m = \text{Diag}(f_{11}, f_{12}, \dots, f_{1m_1}, f_{21}, f_{22}, \dots, f_{2m_2}, \dots, f_{v1}, f_{v2}, \dots, f_{vm_v})$$

La somme de ces éléments vaut donc v . \mathbf{D}_m est une matrice à m lignes et m colonnes. Sa trace vaut v . On note $\mathbf{1}_{nm}$ la matrice à n lignes et m colonnes ne contenant que des 1. L'ACM (synthèse incontournable dans ⁴) est l'analyse du schéma :

$$\begin{array}{ccc} & \frac{1}{v} \mathbf{D}_m & \\ \boxed{m} & \rightarrow & \boxed{m} \\ \mathbf{Y}' \uparrow & & \downarrow \mathbf{Y} = \mathbf{X} \mathbf{D}_m^{-1} - \mathbf{1}_{nm} \\ \boxed{n} & \leftarrow & \boxed{n} \\ & \mathbf{D} & \end{array}$$

Le paramétrage de ce schéma a été choisi pour ses bonnes propriétés. En effet, on considère les vecteurs indicateurs de variables dans l'ensemble des modalités :

$$[\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_v] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

On a toujours $\mathbf{X}\mathbf{w}_k = \mathbf{1}_n$ (il y a une modalité présente et une seule pour chaque individu et chaque variable) donc $(\mathbf{X}\mathbf{D}_m^{-1} - \mathbf{1}_{nm})\frac{1}{v}\mathbf{D}_m\mathbf{w}_k = \mathbf{0}_n$ et les vecteurs \mathbf{w}_k sont propres pour 0. Il y a donc dans cette analyse au moins v valeurs propres nulles. Les axes principaux sont des vecteurs à m composantes formés de $m = m_1 + m_2 + \dots + m_v$ "morceaux". Chaque segment est un codage numérique des modalités de la variable correspondante. Par orthogonalité aux vecteurs \mathbf{w}_k , ces scores des modalités sont centrés (pour les poids des modalités) et $\frac{1}{v}\mathbf{D}_m$ -normés, donc de variance moyenne égale à 1. Ils maximisent successivement :

$$\left\| (\mathbf{X}\mathbf{D}_m^{-1} - \mathbf{1}_{nm})\frac{1}{v}\mathbf{D}_m\mathbf{a} \right\|_{\mathbf{D}}^2 = \left\| (\mathbf{X} - \mathbf{1}_{nm}\mathbf{D}_m)\frac{1}{v}\mathbf{a} \right\|_{\mathbf{D}}^2 = \left\| \frac{1}{v}\mathbf{X}\mathbf{a} \right\|_{\mathbf{D}}^2$$

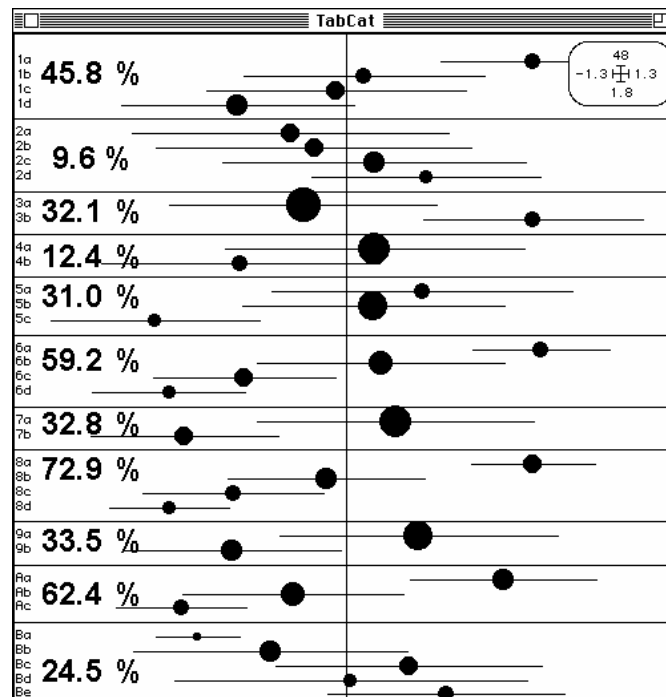
Or le vecteur $\mathbf{1}_n$ est composante principale pour la valeur propre 0 ($(\mathbf{D}_m^{-1}\mathbf{X}^t - \mathbf{1}_{mn})\mathbf{D}\mathbf{1}_n = 0$) donc les autres composantes principales sont centrées et $\left\| \frac{1}{v}\mathbf{X}\mathbf{a} \right\|_{\mathbf{D}}^2$ est la variance du score des individus obtenu en attribuant la moyenne des scores des modalités portées.

Inversement, une composante principale est un score des individus de moyenne 0 et variance 1 qui maximise :

$$\left\| (\mathbf{D}_m^{-1}\mathbf{X}^t - \mathbf{1}_{mn})\mathbf{D}\mathbf{b} \right\|_{\frac{1}{v}\mathbf{D}_m}^2 = \left\| \mathbf{D}_m^{-1}\mathbf{X}^t\mathbf{D}\mathbf{b} \right\|_{\frac{1}{v}\mathbf{D}_m}^2$$

c'est-à-dire la moyenne de la variance du score des modalités obtenu en attribuant à chaque modalité la moyenne des scores des individus porteurs.

L'objectif fondamental de l'ACM est de d'obtenir des scores numériques des individus qui maximisent les pourcentages de variance expliquée, en moyenne, pour toutes les variables qualitatives (rapport de corrélation) :



L'ACM fait la synthèse des liens entre les variables qualitatives par l'examen des liens entre les variables qualitatives q^j et un code numérique de synthèse appelée variable synthétique.

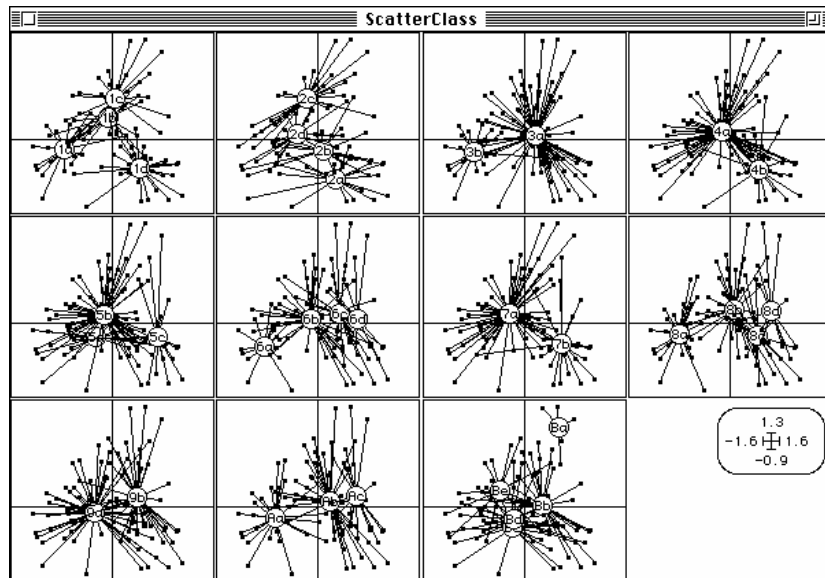
En ACP normée, on a trouvé une variable synthétique (la première composante principale, concrètement la première coordonnée normée des lignes) qui maximise :

$$\sum_{j=1}^p cor^2(x^j, y)$$

En ACM, on a trouvé une variable synthétique (la première composante principale, concrètement la première coordonnée normée des lignes) qui maximise :

$$\frac{1}{v} \sum_{j=1}^v \eta^2(q^j, y)$$

En utilisant deux coordonnées, on obtient le graphe canonique du type :



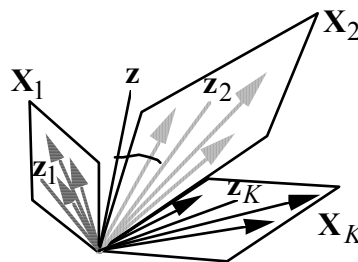
Chacune des variables qualitatives exprime sa partition sur le même plan des individus. Les deux coordonnées sont non corrélées et maximisent successivement la moyenne des rapports de corrélation (valeur propre).

L'ACM peut donc être vue comme l'ACP normée des variables qualitatives. On peut alors mélanger p variables qualitatives et v variables quantitatives (Analyse de Hill & Smith⁵) pour faire une analyse dont les coordonnées optimiseront :

$$\frac{1}{p} \sum_{j=1}^p \text{cor}^2(\mathbf{x}^j, \mathbf{y}) + \frac{1}{v} \sum_{j=1}^v \eta^2(\mathbf{q}^j, \mathbf{y})$$

2.2. Analyse canonique généralisée

L'objectif fondamental de l'ACG est de fournir des variables normalisées (scores canoniques) qui font en moyenne avec tous les sous-espaces engendrés par chaque tableau des angles minimaux ($\cos 2$) :

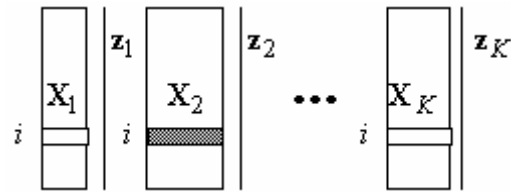


L'ACG est de Carroll⁶. Elle manipule K tableaux d'ACP centrés (ou normés, mais cela n'a aucune incidence) ayant les mêmes lignes-individus :

$$(\mathbf{X}_1, \mathbf{I}_1, \mathbf{D}) \quad (\mathbf{X}_2, \mathbf{I}_2, \mathbf{D}) \cdots (\mathbf{X}_K, \mathbf{I}_K, \mathbf{D})$$

Elle donne pour chaque tableau des combinaisons linéaires (variables canoniques) :

$$\mathbf{z}_1 = \mathbf{X}_1 \mathbf{a}_1, \mathbf{z}_2 = \mathbf{X}_2 \mathbf{a}_2, \dots, \mathbf{z}_K = \mathbf{X}_K \mathbf{a}_K$$

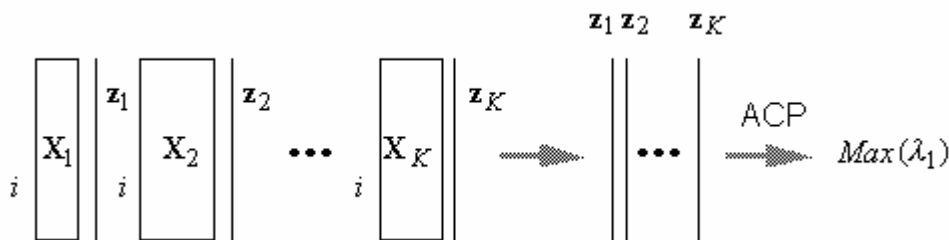


centrées et réduites et une variable synthétique \mathbf{z} (score canonique) centrée et réduite qui maximise :

$$\sum_{k=1}^K \text{cor}^2(\mathbf{z}, \mathbf{z}_k)$$

Si chaque tableau n'est formé que d'une variable quantitative, on retrouve l'ACP normée. Si chaque tableau n'est formé que d'une variable qualitative on retrouve l'ACM et pour un mélange des deux, on retrouve l'Analyse de Hill & Smith.

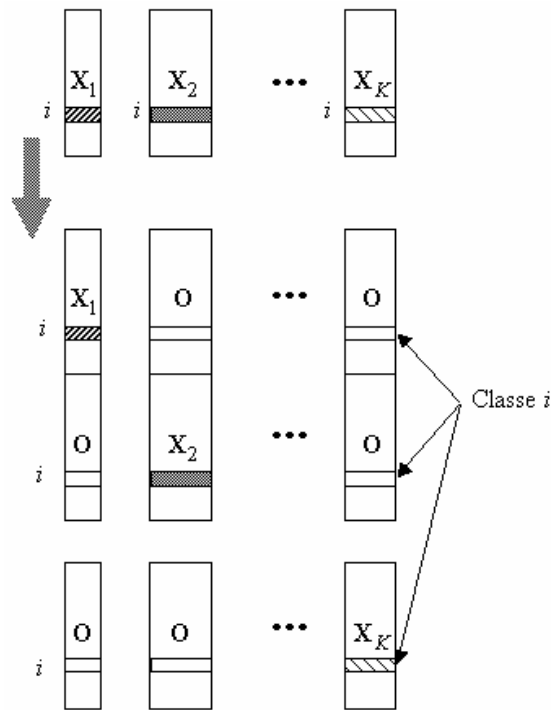
La première valeur propre de l'analyse $\sum_{k=1}^K \text{cor}^2(\mathbf{z}, \mathbf{z}_k)$ est en outre la première valeur propre de l'ACP normée de l'ensemble des variables canoniques de rang 1. On peut donc voir dans l'ACG la recherche d'une variable normée par sous-espace dont l'ensemble donne une ACP de première valeur propre optimale. Kettenring⁷ attribue ce principe à Vinograd⁸ et les premières approches numériques à Horst⁹.



Le dépouillement de l'ACG a été enrichi par Casin et Turlot¹⁰ qui y voient une analyse discriminante. Le schéma de l'ACG s'écrit en effet :

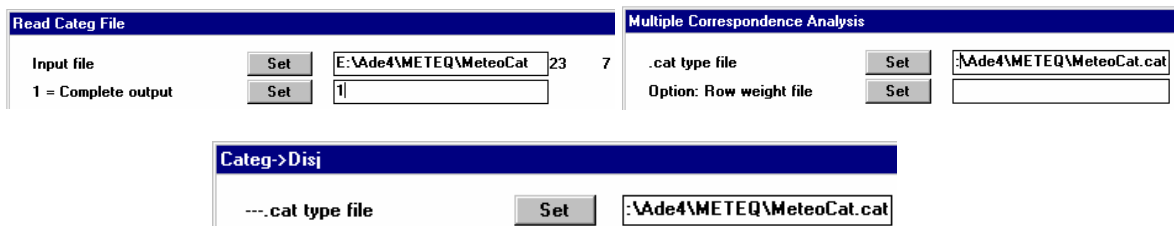
$$\begin{array}{ccc}
 \boxed{p_1 + p_2 + \dots + p_K} & \xrightarrow{\quad} & \boxed{p_1 + p_2 + \dots + p_K} \\
 \mathbf{X}' \uparrow & & \downarrow \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \\
 \boxed{n} & \xleftarrow{\quad} & \boxed{n} \\
 & \mathbf{D} &
 \end{array}$$

$$\begin{bmatrix}
 (\mathbf{X}'_1 \mathbf{D} \mathbf{X}_1)^{-1} & 0 & \dots & 0 \\
 0 & (\mathbf{X}'_2 \mathbf{D} \mathbf{X}_2)^{-1} & \dots & 0 \\
 \vdots & \vdots & \ddots & \vdots \\
 0 & 0 & \dots & (\mathbf{X}'_K \mathbf{D} \mathbf{X}_K)^{-1}
 \end{bmatrix}$$

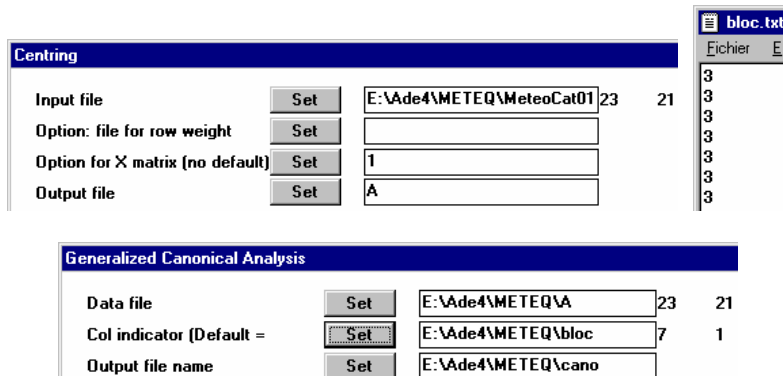


Le schéma renvoie au tableau déroulé dont le tableau juxtaposé contient les moyennes. Cette profonde unité de la théorie fait parfois sourire car l'analyse canonique généralisée est ici un cas particulier d'analyse discriminante qui est un cas particulier d'analyse canonique qui est un cas particulier d'analyse canonique généralisée. L'ACG est dans le module Canonical.

Saporta¹¹ a établi que l'ACM est un cas particulier de l'analyse canonique généralisée. Faire l'analyse des correspondances multiples (ACM) du tableau de variables qualitatives c'est faire l'analyse canonique généralisée (ACG) du K-tableau des indicatrices des classes centrées.



File E:\Ade4\METEQ\MeteoCat.cmrc contains the correlation ratios between each categorical variable and each factorial score
It has 7 rows and 2 columns



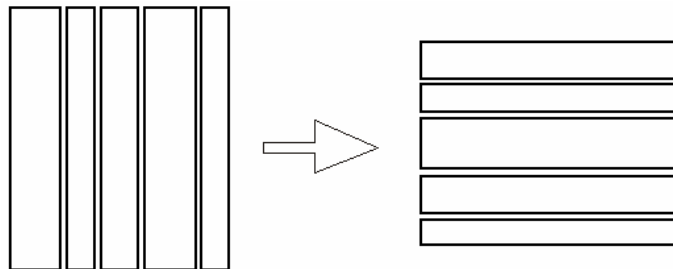
```

Cos 2 between scores and subspaces -----
|-----|-----|-----|
E:\Ade4\METEQ\cano_cos2 Binary input file:
|-----|-----|-----|
| Tab 1 | 5.391e-01 | 3.010e-01 | 1 | 0.5391 0.3010
| Tab 2 | 8.271e-01 | 1.659e-01 | 2 | 0.8271 0.1659
| Tab 3 | 5.514e-01 | 4.393e-01 | 3 | 0.5514 0.4393
| Tab 4 | 7.385e-01 | 2.207e-01 | 4 | 0.7385 0.2207
| Tab 5 | 5.697e-01 | 4.002e-01 | 5 | 0.5697 0.4002
| Tab 6 | 6.605e-01 | 4.618e-01 | 6 | 0.6605 0.4618
| Tab 7 | 4.405e-01 | 4.669e-01 | 7 | 0.4405 0.4669
|-----|-----|-----|
    
```

2.3. Analyse factorielle multiple

L'ACG est une méthode qui résume des paquets de variables. Chaque tableau donne une variable canonique de variance unité et le score canonique est la première composante principale du tableau des variables canoniques. Ces variables canoniques, combinaisons des variables d'un tableau sont normées. Elles ne donnent aucune information sur la structure de chaque tableau. Quand les tableaux ont un nombre de variables important, ces combinaisons sont instables numériquement et de peu de signification pratique.

C'est pourquoi on préfère alors l'analyse factorielle multiple (AFM) ¹². Pour l'exécution de l'analyse, on doit disposer d'un K-tableaux. Faisons l'AFM du tableau d'ACM. Utiliser DDUtil: Triplet Transpose qui conserve le triplet statistique de l'ACM en plaçant les blocs de colonnes les uns sous les autres :



puis KTabUtil: TripletToKtab qui définit une structure de K-tableaux :

Triplet Transpose

Input file Ade4\METEQ\MeteoCat.cmta 23 21

Generic output file name W

TripletToKTab

..-ta type file E:\Ade4\METEQ\W.trta 21 23

Row indicator (Default = 1) E:\Ade4\METEQ\bloc 7 1

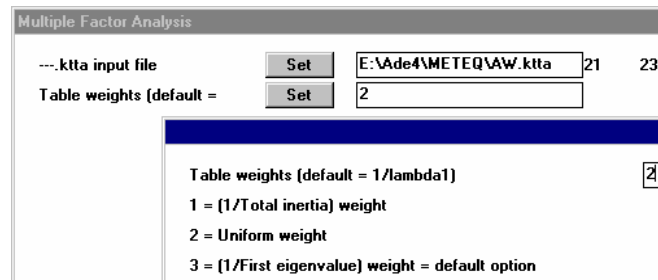
Col indicator (Default = 1)

Output file name AW

Output ASCII file: AW!.ktp

```

7:3/3/3/3/3/3/ Structure en blocs de lignes
1:23/          Structure en blocs de colonnes
    
```



```
DiagoRC: General program for two diagonal inner product analysis
Input file: E:\Ade4\METEQ\AW.++ta
--- Number of rows: 21, columns: 23
-----
Total inertia:          2
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.1812E-01  +0.3091  +0.3091  | 02  +3.5082E-01  +0.1754  +0.4845  |
03  +2.5032E-01  +0.1252  +0.6096  | 04  +2.1054E-01  +0.1053  +0.7149  |
05  +1.5776E-01  +0.0789  +0.7938  | 06  +1.1827E-01  +0.0591  +0.8529  |
...

```

Comparer avec l'ACM :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: E:\Ade4\METEQ\MeteoCat.cmta
--- Number of rows: 23, columns: 21
-----
Total inertia:          2
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.1812E-01  +0.3091  +0.3091  | 02  +3.5082E-01  +0.1754  +0.4845  |
03  +2.5032E-01  +0.1252  +0.6096  | 04  +2.1054E-01  +0.1053  +0.7149  |
05  +1.5776E-01  +0.0789  +0.7938  | 06  +1.1827E-01  +0.0591  +0.8529  |

```

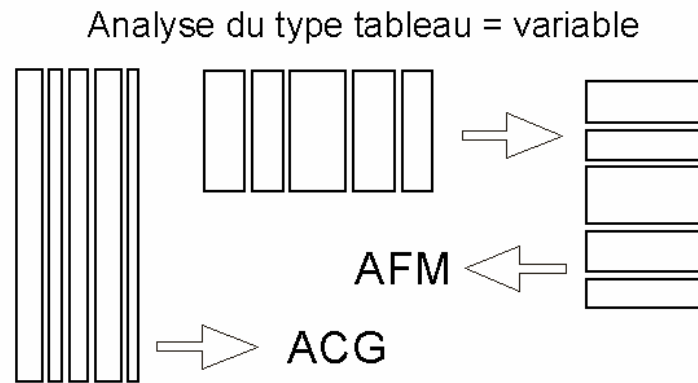
Dans l'AFM on retrouve :

```
R2 coefficients = cos2 between canonical scores and subspaces
|-----|-----|-----|
|          | fac 1 | fac 2 |
|-----|-----|-----|
| Tab 1 | 5.391e-01 | 3.010e-01 |
| Tab 2 | 8.271e-01 | 1.659e-01 |
| Tab 3 | 5.514e-01 | 4.393e-01 |
| Tab 4 | 7.385e-01 | 2.207e-01 |
| Tab 5 | 5.697e-01 | 4.002e-01 |
| Tab 6 | 6.605e-01 | 4.618e-01 |
| Tab 7 | 4.405e-01 | 4.669e-01 |
|-----|-----|-----|

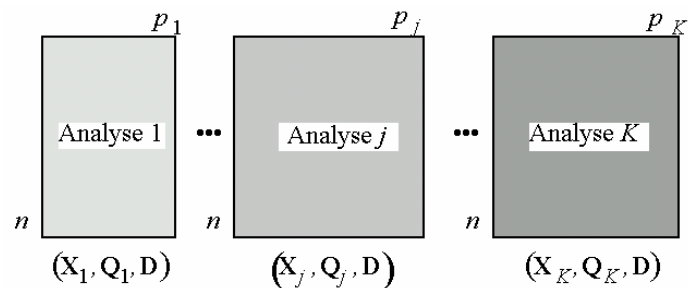
```

L'ACM, l'ACG et l'AFM donnent les mêmes résultats *sans être les mêmes méthodes*. Cela vient de la propriété très particulière des indicatrices des classes qui sont orthogonales par bloc de variables. La décorrélation propre à l'ACG est sans effet.

De même l'ACP normée, l'ACG avec une variable par tableaux et l'AFM avec une variable par tableau sont identiques. Noter le point d'orientation :



Les concepts sous-jacents à l'AFM sont simples et méritent d'être explicités. On dispose de K tableaux ayant en commun les lignes-individus, chacun d'entre eux correspondant à un groupe de variables-colonnes. On dispose donc de K triplets statistiques $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$ ($1 \leq k \leq K$) :



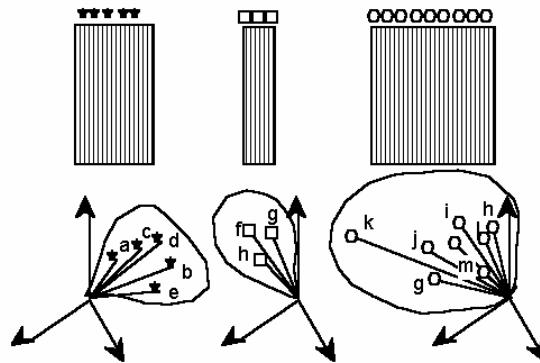
On considère le tableau global :

$$\begin{array}{ccc}
 \boxed{p_1 + p_2 + \dots + p_K} & \begin{bmatrix} \mathbf{Q}_1 & 0 & \dots & 0 \\ 0 & \mathbf{Q}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{Q}_K \end{bmatrix} & \boxed{p_1 + p_2 + \dots + p_K} \\
 \mathbf{X}' \uparrow & \rightarrow & \downarrow \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \\
 \boxed{n} & \leftarrow & \boxed{n} \\
 & \mathbf{D} &
 \end{array}$$

Si les \mathbf{Q}_k sont des matrices diagonales, leur réunion aussi et le schéma ci-dessus est l'analyse simple du tableau global qu'on appelle analyse simultanée. Elle a le défaut de dépendre de l'importance respective des analyses séparées $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$, les tableaux ayant plus de colonnes (p_k), plus d'inertie ($Trace(\mathbf{X}_k' \mathbf{D} \mathbf{X}_k \mathbf{Q}_k)$) ou des valeurs propres plus marquées imposant systématiquement leur point de vue. C'est pourquoi on pondère le rôle de chaque tableau en utilisant en AFM un schéma du type :

$$\begin{array}{ccc}
 \boxed{p_1 + p_2 + \dots + p_K} & \begin{bmatrix} \pi_1 \mathbf{Q}_1 & 0 & \dots & 0 \\ 0 & \pi_2 \mathbf{Q}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \pi_K \mathbf{Q}_K \end{bmatrix} & \boxed{p_1 + p_2 + \dots + p_K} \\
 \mathbf{X}' \uparrow & \rightarrow & \downarrow \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \\
 \boxed{n} & \leftarrow & \boxed{n} \\
 & \mathbf{D} &
 \end{array}$$

L'analyse fonctionne alors comme une analyse d'un nuage de variables.



Notons p le nombre total de variables, P l'ensemble des p premiers entiers, à savoir $\{1, 2, \dots, p\}$. Notons p_1, p_2, \dots, p_K les effectifs des variables par tableau et P_1, P_2, \dots, P_K les parties de P correspondant à chaque groupe. Cela permet d'écrire $j \in P_k$ ($1 \leq j \leq p$) ($1 \leq k \leq K$) pour dire que la variable j dans la numérotation complète appartient au tableau k . x^j désigne la variable j , vecteur de \mathbb{R}^n (n est le nombre d'individus), de poids initial m_j dans l'analyse globale (souvent l'unité). \mathbb{R}^n est muni du produit scalaire \mathbf{D} associé à la pondération des individus.

On note encore \mathbf{Q}_k la matrice diagonale à p_k lignes et p_k colonnes qui contient les poids m_j des variables du groupe k . On note enfin α_k la première valeur propre de l'analyse séparée du triplet $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$, où \mathbf{X}_k est le tableau du groupe k .

Si \mathbf{z} est un vecteur de \mathbb{R}^n \mathbf{D} -normé, l'inertie projetée associée aux p_k colonnes de \mathbf{X}_k est :

$$I_k(\mathbf{z}) = \sum_{j \in P_k} p_j \langle \mathbf{x}^j | \mathbf{z} \rangle_{\mathbf{D}}^2 = \|\mathbf{X}_k^t \mathbf{D} \mathbf{z}\|_{\mathbf{Q}_k}^2 = \langle \mathbf{X}_k^t \mathbf{D} \mathbf{z} | \mathbf{X}_k^t \mathbf{D} \mathbf{z} \rangle_{\mathbf{Q}_k} = \langle \mathbf{X}_k \mathbf{D} \mathbf{X}_k^t \mathbf{Q}_k \mathbf{z} | \mathbf{z} \rangle_{\mathbf{D}} = \langle \mathbf{W}_k \mathbf{D} \mathbf{z} | \mathbf{z} \rangle_{\mathbf{D}}$$

On appellera valeur du lien entre le vecteur \mathbf{z} et le tableau \mathbf{X}_k cette inertie projetée ramenée au maximum potentiel, c'est-à-dire α_k . D'où la notation :

$$L(\mathbf{z}, k) = \sum_{j \in P_k} \frac{p_j}{\alpha_k} \langle \mathbf{x}^j | \mathbf{z} \rangle_{\mathbf{D}}^2 = \frac{1}{\alpha_k} \|\mathbf{X}_k^t \mathbf{D} \mathbf{z}\|_{\mathbf{Q}_k}^2 = \frac{1}{\alpha_k} \langle \mathbf{W}_k \mathbf{D} \mathbf{z} | \mathbf{z} \rangle_{\mathbf{D}}$$

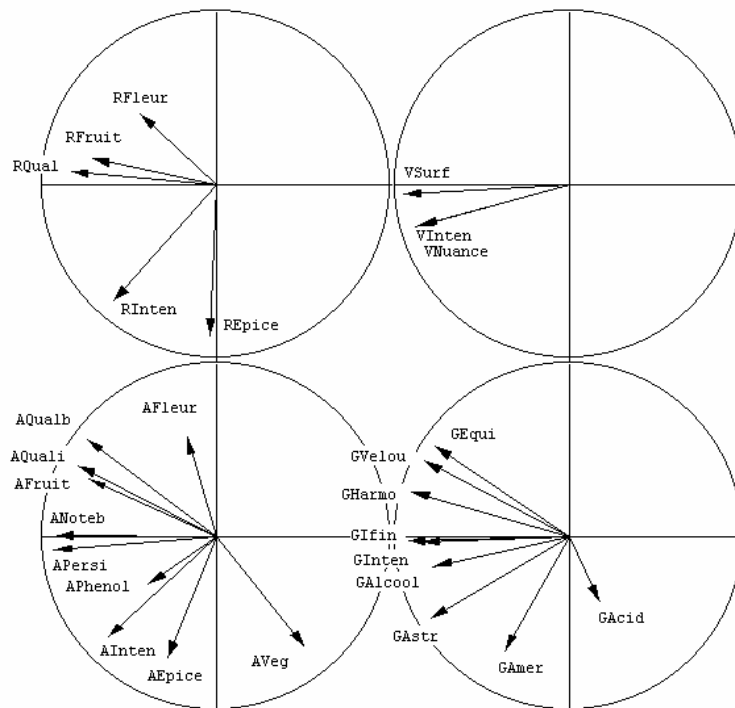
Ce lien ne peut donc dépasser l'unité (pourcentage d'un optimum). La valeur globale du vecteur \mathbf{z} , sur l'ensemble des nuages de variables (des tableaux) est la somme des liens, soit :

$$L(\mathbf{z}) = \sum_{k=1}^K L(\mathbf{z}, k) = \sum_{k=1}^K \left\langle \frac{1}{\alpha_k} \mathbf{W}_k \mathbf{Dz} \middle| \mathbf{z} \right\rangle_{\mathbf{D}} = \left\langle \left(\sum_{k=1}^K \frac{1}{\alpha_k} \mathbf{W}_k \right) \mathbf{Dz} \middle| \mathbf{z} \right\rangle_{\mathbf{D}}$$

Il est donc logique de choisir :

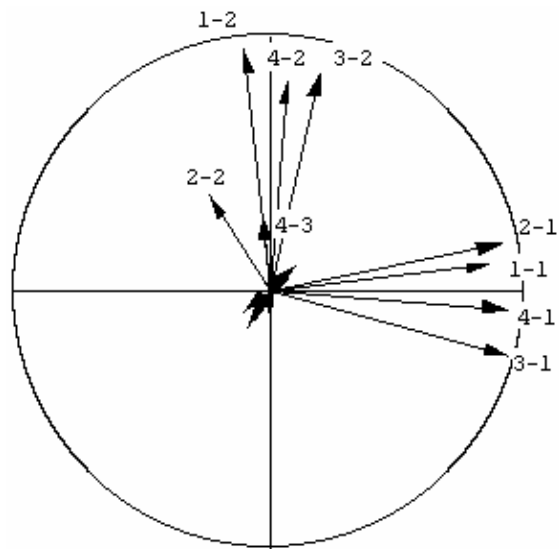
$$\pi_k = \frac{1}{\alpha_k}$$

L'optimum est atteint pour la première composante principale normée du schéma surpondéré de l'AFM et cet optimum ne peut dépasser le nombre de tableaux (somme de k valeurs inférieures ou égales à 1). On cherche ensuite un second vecteur \mathbf{z} , orthogonal au précédent qui optimise la même quantité et on trouve la seconde composante de l'AFM. En cherchant un plan qui optimise l'inertie projetée, on trouve le sous-espace engendré par les deux premières composantes.

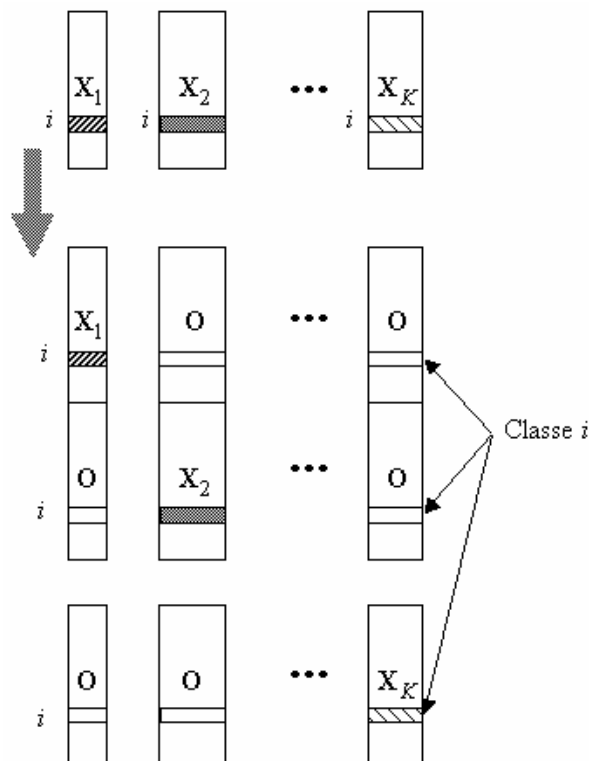


Projection sur les composantes principales d'une AFM des variables (représentation par tableau)

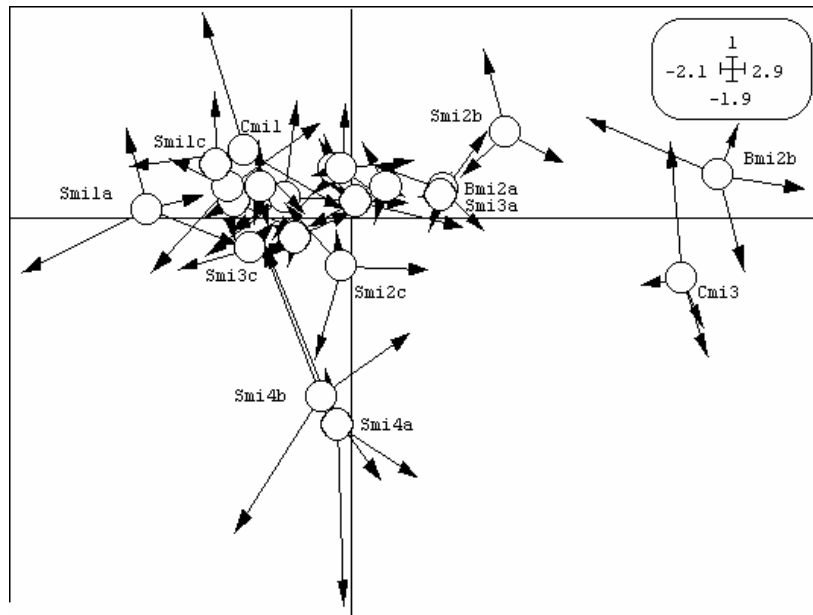
Tout se passant dans le même espace \mathbb{R}^n , il est aussi logique de projeter sur le plan des composantes de l'AFMULT les composantes principales des sous-nuages de variables par tableau :



Enfin l'argument utilisé par Casin et Turlot avait été mis en œuvre par les auteurs de l'AFM :



Le tableau global est formé des moyennes par individu du tableau en dessous. En ACG on inverse les normes et on obtient une analyse discriminante. En AFM, on conserve les métriques de départ (à la surpondération près) et on obtient une ACP inter-classes qui permet de juxtaposer sur la carte des lignes chaque point vu par chaque tableau.



Représentation simultanée en AFM des points-lignes et de leur représentation par chaque tableau.

Pour des K -tableaux de paquets de variables quantitatives, l'AFM est simple de mise en œuvre et les principes d'interprétation sont accessibles. Tous les détails (mais sans bibliographie et sans données) sont dans ¹².

2.4. Variables floues

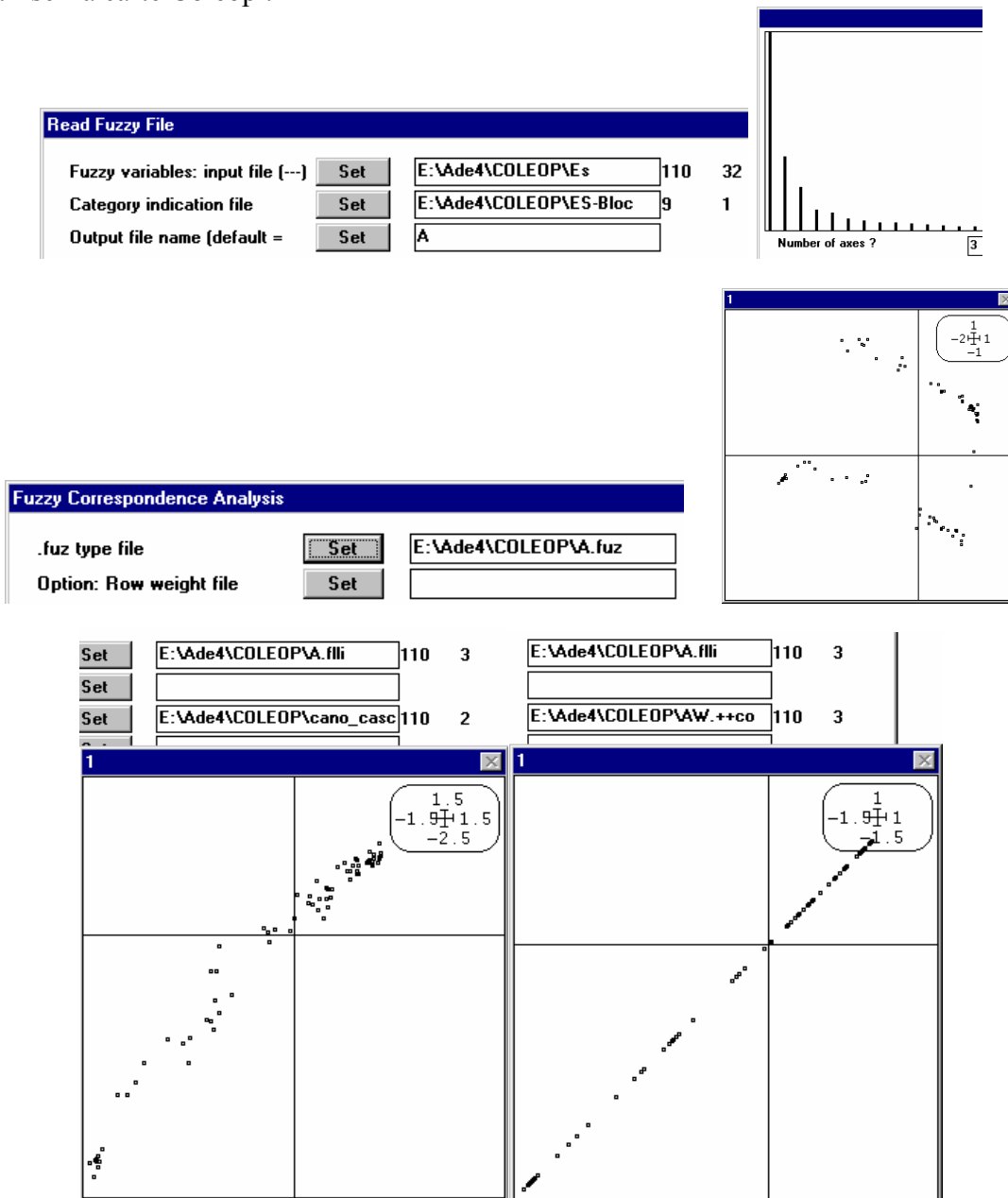
Un archétype des tableaux ayant valeur expérimentale de variables est celle des variables floues. Quand on veut noter le lien entre une espèce et l'altitude, on peut dire 670 m pour signifier que l'espèce se trouve en moyenne à 670 m. On peut dire 1|3|2|0|0 sur le code [0,400], [400,800], [800,1200], [1200,1600] et [1600,200] pour dire qu'une fois sur 6 on trouve l'espèce dans la première classe, 1 fois sur 2 dans la seconde et 1 fois sur 3 dans la troisième. Le profil écologique est une variable floue. Sur plusieurs variables et plusieurs espèces, on obtient des tableaux de variables floues :

	1	2	3		1	2	3		1	2	3
1	1	3	0	1	0.250	0.750	0.000	1	0.250	0.750	0.000
2	1	3	0	2	0.250	0.750	0.000	2	0.250	0.750	0.000
3	3	0	0	3	1.000	0.000	0.000	3	1.000	0.000	0.000
4	1	3	0	4	0.250	0.750	0.000	4	0.250	0.750	0.000
5	3	0	0	5	1.000	0.000	0.000	5	1.000	0.000	0.000
6	3	0	0	6	1.000	0.000	0.000	6	1.000	0.000	0.000
7	0	1	3	7	0.000	0.250	0.750	7	0.000	0.250	0.750
8	2	2	0	8	0.500	0.500	0.000	8	0.500	0.500	0.000
9	0	1	3	9	0.000	0.250	0.750	9	0.000	0.250	0.750
10	1	3	0	10	0.250	0.750	0.000	10	0.250	0.750	0.000
11	0	0	0	11	0.000	0.000	0.000	11	0.339	0.407	0.254
12	1	3	0	12	0.250	0.750	0.000	12	0.250	0.750	0.000
13	1	3	0	13	0.250	0.750	0.000	13	0.250	0.750	0.000
14	1	3	0	14	0.250	0.750	0.000	14	0.250	0.750	0.000
15	3	0	0	15	1.000	0.000	0.000	15	1.000	0.000	0.000
16	1	3	0	16	0.250	0.750	0.000	16	0.250	0.750	0.000
17	1	3	0	17	0.250	0.750	0.000	17	0.250	0.750	0.000

ADE-4 accepte les données manquantes sur ce type de variables en proposant de remplacer le profil 0|0|...|0 par le profil moyen des autres.

Les variables floues transfèrent la notion expérimentale de variable de la notion technique de colonne à la notion technique de tableau. On étend ainsi l'usage des variables qualitatives. *Mais* l'AFC d'un tel tableau (MCA: Fuzzy Correspondence Analysis) n'est plus identique à l'ACG de ce tableau (Canonical: Canonical correlations Analysis) elle-même différente de l'AFM (KTA-MFA: Multiple Factorial Analysis) du même K -tableaux. On est là au centre des difficultés associées à la statistique K -tableaux. Une analyse simple résout avec un seul algorithme plusieurs problèmes différents. Ceci devient impossible avec plusieurs tableaux et en décidant de conserver la solution d'un problème on perd celle des autres. Il y a alors plusieurs méthodes concurrentes, donc plus de choix à faire, donc plus de problèmes.

Utiliser la carte Coleop :



Observer que l'AFM et l'ACF sont quasiment identiques et ne diffèrent que par certaines aides à l'interprétation tandis que l'ACG se sépare nettement des deux précédentes. L'ACF a été dupliquée sous le terme RCT-MCA en génétique (tableaux de fréquences alléliques) dans ¹³. Après examen, ce transfert est peu légitime et déconseillé. L'ACP (PCA: Fuzzy PCA) est beaucoup plus proche des calculs de Fst.

3. Tableau comme individu

3.1. Analyses inter-classes et analyses discriminantes

Si on déconseille de travailler sur des tableaux de fréquences alléliques, c'est simplement parce que ce sont des tableaux de moyennes. Représenter une population par sa moyenne directement, c'est se priver de pouvoir représenter sa diversité, sa variabilité, ce qui est bizarre en génétique. Il se peut qu'un tableau soit décomposé en sous-tableaux pour que chacun d'entre eux soit le descriptif d'un groupe d'individus, la typologie de groupes étant l'objectif essentiel.

Le module Discrimin contient tout ce qui touche à ce point de vue. Seule l'option Discrimin: Within Analysis/Run poursuit un objectif radicalement inverse. C'est le point d'entrée de la troisième partie.

3.2. Intra-classes

Nous avons rencontré les analyses intra-classes dans le champ des analyses à un tableau comme une transformation préalable des données d'un type particulier dans :

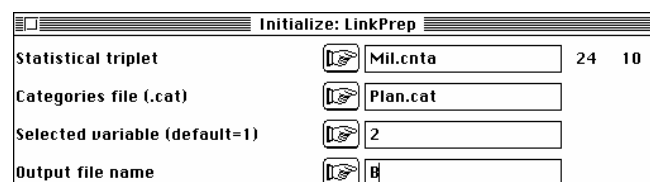
PCA: Partial normed PCA (centrage et réduction par blocs)

PCA: Within group normalized PCA (centrage par blocs et normalisation)

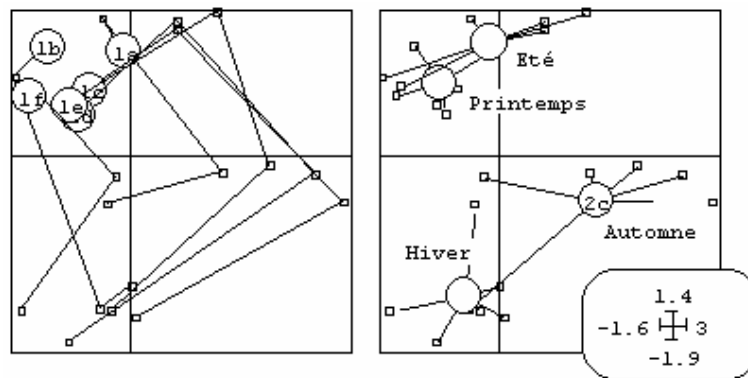
COA: Internal COA (centrage par blocs sur une ou deux dimensions avant AFC)

Ces options considèrent simplement qu'on ne veut pas voir les effets d'une partition externe. Une partie de l'information est un parasite éliminé. On peut cependant voir dans les intra-classes une introduction simple à ce qui fait l'essence des méthodes à compromis. Un bloc de lignes dans un tableau de données est alors vu comme un tableau et on superpose des typologies.

Utiliser la carte Méaudret pour obtenir le fichier Mil (24-10) et Plan (24-2). Lire ce dernier avec CategVar : Read Categ File. Faire l'ACP normée du tableau Mil (PCA : Correlation matrix PCA). Initialiser le lien entre cette analyse et la partition des relevés par dates (Discrimin : Initialize/LinkPrep) :

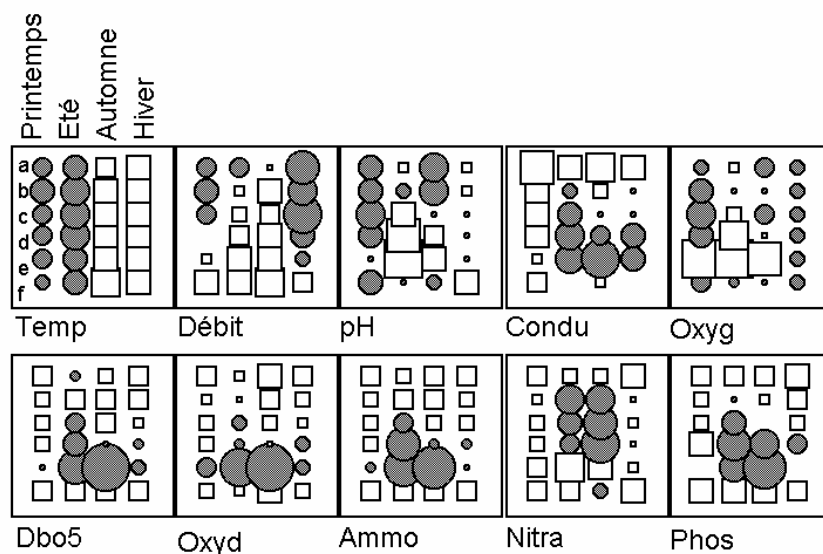


Observer (ScatterClass : Trajectories et ScatterClass : Stars) que le plan 2-3 de l'ACP normée décrit des variations temporelles :



Cette figure est fondamentale. Il s'agit du même plan factoriel vu à gauche comme intra-stations et à droite comme inter-dates. Les données étant très simples, on trouve ceci directement. Mais il est possible dans les jeux de données complexes qu'une partie des variables définissent une structure et qu'une partie des variables définissent l'autre. Dans ce cas séparer les saisons est un objectif, superposer les évolutions en est un autre. A gauche on poursuit un objectif de *compromis* entre tableaux comme structures (identifier par blocs une typologie qui est l'expression d'une typologie commune), à droite on poursuit un objectif de typologies de tableaux considérés comme individus eux-mêmes (identifier ce qui en moyenne distingue les blocs).

L'analyse sans *a priori* décrit simultanément rythme saisonnier et typologie spatiale intervenant dans le processus spatio-temporel généré par le déplacement de l'eau et l'évolution de la température externe. On peut examiner ce processus sous plusieurs angles au sens géométrique du terme c'est-à-dire choisir de privilégier ou d'éliminer une des composantes du plan d'expérience. Représenter les données par Scatters : Values :



Comparer alors les analyses intra-stations et inter-dates. Les analyses intra-classes sont les plus simples des approches d'un *K*-tableaux comme ensemble de typologies à coordonner. Elles sont simples mais rustiques et souvent insuffisantes.

Ce paragraphe très court n'a qu'un seul objectif : souligner qu'un paquet de tableaux superposés est une structure de données qui conduit à l'emploi de méthodes profondément différentes dans leurs objectifs. Si les colonnes communes sont des variables, on peut séparer les populations par leur moyenne (inter-classes, discriminante). On peut rechercher au contraire un compromis entre leur structure interne (intra-classes, STATIS). Inter-classes et intra-classes sont des jumelles au plan théorique (deux ACPVI projetant soit sur un espace soit sur son orthogonal). Elles sont opposées au plan expérimental (ce qui sépare contre ce qui rapproche). Intra-classes et STATIS sont fort éloignées au plan théorique, elle sont voisines au plan expérimental. Plus qu'ailleurs il est indispensable de savoir ce que l'on veut.

4. Tableau comme structure

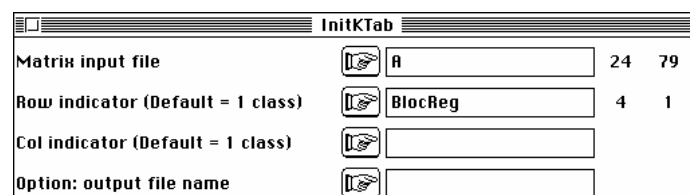
4.1. Analyses triadiques partielles

On s'intéresse d'abord aux méthodes d'analyse des cubes de données. La principale difficulté de ces approches est le choix de la face du cube qui sert de référence. Il y a toujours 6 dispositions fondamentales.

En AFC elles sont équivalentes deux à deux et il y a trois stratégies d'analyse qui donnent des résultats radicalement différents. Quand on veut faire simultanément les AFC des « feuilles » d'un cube, on utilise STATIS : Foucart's COA. L'objectif est de faire simultanément plusieurs analyses des correspondances sur des tableaux portant sur les mêmes lignes et les mêmes colonnes. Cette pratique est définie par T. Foucart qui donne son nom à l'option ¹⁴. Utiliser la carte Succes ³. Transposer le fichier (FilesUtil : Transpose) :



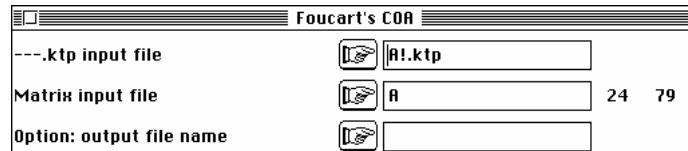
Définir avec le résultat un K-tableaux (KTabUtil : InitKTab) :



4 tableaux sont superposés. Ils ont en commun 79 colonnes (espèces) et 6 lignes (strates, types de végétation). L'option STATIS : Foucart's COA suppose que les tableaux sont appariés par les lignes et par les colonnes. Elle contrôle que les tableaux ont tous mêmes dimensions mais il appartient à l'utilisateur de vérifier que les objets de mesure sont les mêmes dans chaque tableau.

```
Option: InitKTab
Input file: A
Row indicator file: BlocReg
Col indicator file:
Output ASCII file: A!.ktp
4 : 6/6/6/6/
1 : 79/
```

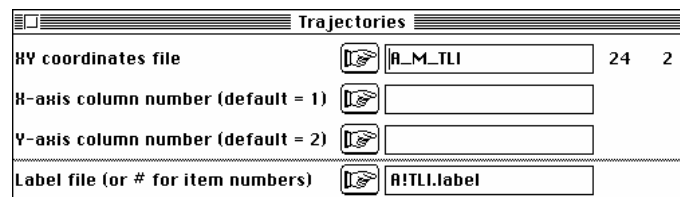

Exécuter l'analyse :



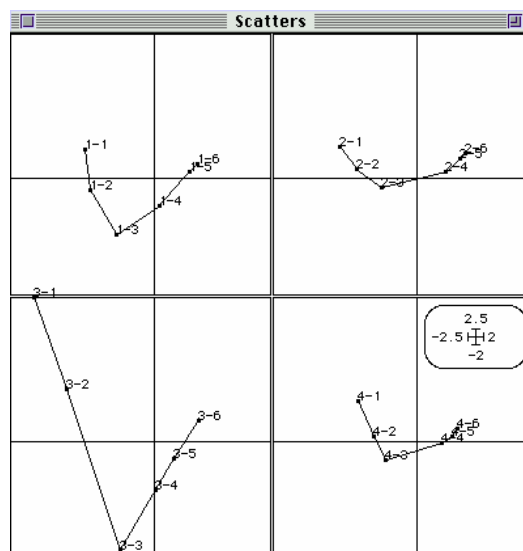
Chaque tableau est passé en pourcentage de son abondance totale. Un tableau compromis est calculé en faisant simplement la moyenne des tableaux du type $[p_{ij}]$. Cette moyenne est encore une distribution de fréquence bivariée du type $[p_{ij}]$. Son analyse des correspondances est exécutée. Le titre de cette analyse est A_M.fc. Les valeurs propres sont éditées :

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+5.7435E-01	+0.5794	+0.5794	02	+2.2225E-01	+0.2242	+0.8036
03	+9.1626E-02	+0.0924	+0.8960	04	+7.2898E-02	+0.0735	+0.9695
05	+3.0223E-02	+0.0305	+1.0000	06	+0.0000E+00	+0.0000	+1.0000

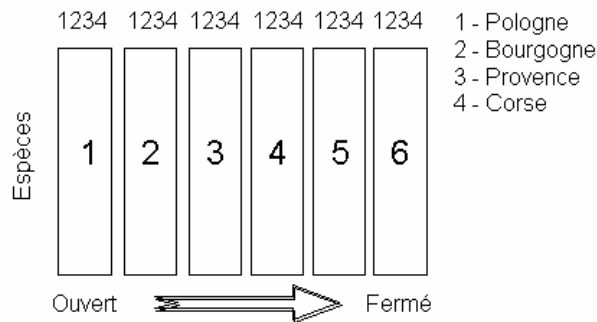
Cette AFC classique est simplement complétée par la projection en lignes supplémentaires de toutes les lignes de tableaux de départ et en colonnes supplémentaires de toutes les colonnes du tableau de départ. Le dépouillement ne pose pas de problèmes particuliers. Utiliser de préférence le multi-fenêtrage par tableaux (Scatters : Trajectories) :



Ce qui concerne les lignes est dans la structure TLI (T numéro du tableau, L numéro de la ligne globale, l numéro de la ligne par tableau). L'analyse a caractérisé le gradient commun de végétation sur lequel se distribue la niche des espèces d'oiseaux dans chaque région, ce n'est pas l'objectif des auteurs :



Reprendre le même exemple pour changer radicalement la question.

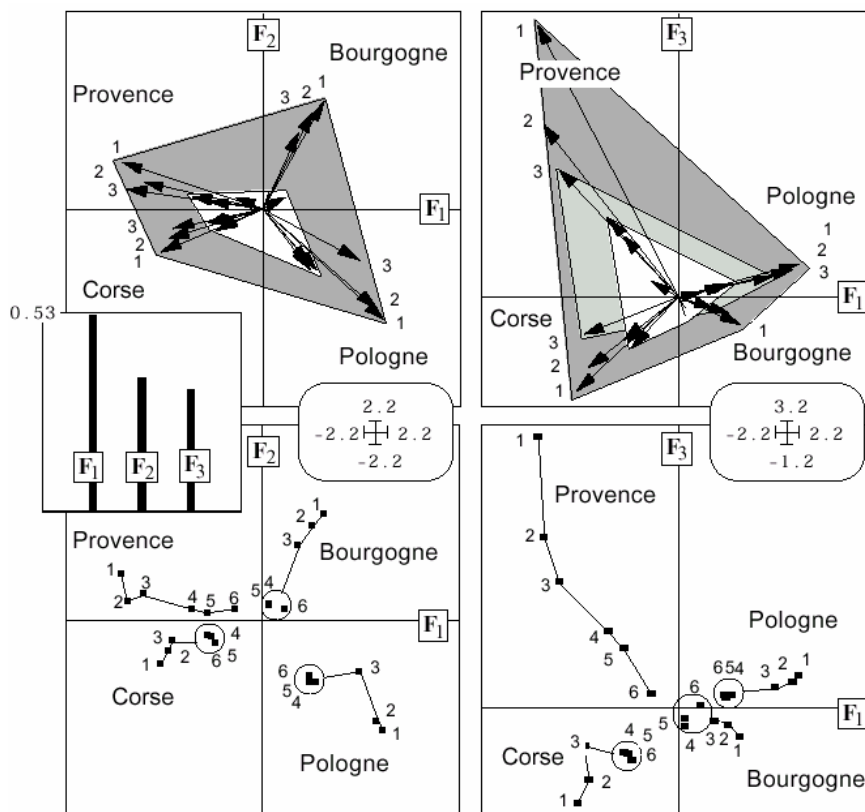


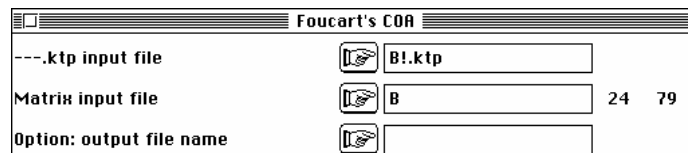
Trier les lignes (regions•strates) du tableau A par strate (tri par numéro de la ligne par tableaux, donc sur la colonne 3 de A!TLL). L'option KTabUtil : SortKTab permet de transférer la structure de K-tableaux à cette nouvelle disposition :



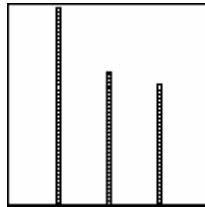
Output file: B
 Output ASCII file: B!.ktp
 -> Rows: 24, columns: 79
 -> Content: 6 tables with 4 rows and 79 columns
 Output ASCII file: B!.ktp
6 : 4/4/4/4/4/4/
1 : 79/

Refaire l'analyse pour obtenir :





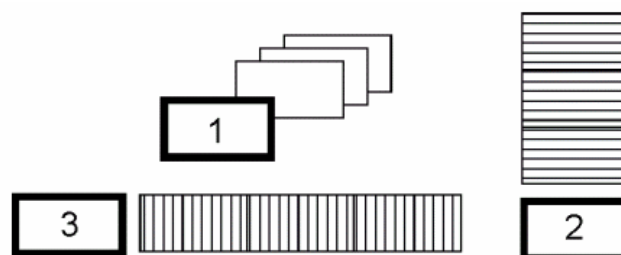
Num.	Eigenval.	R.Iner.	R.Sum		Num.	Eigenval.	R.Iner.	R.Sum	
01	+5.2777E-01	+0.4361	+0.4361		02	+3.5909E-01	+0.2967	+0.7327	
03	+3.2346E-01	+0.2673	+1.0000		04	+0.0000E+00	+0.0000	+1.0000	



Il n'y a pas réduction de dimension et ceci est rarissime. Il est possible de multifenêtrer par tableau (6) mais il faut 2 plans pour 3 facteurs. On préfère directement superposer (ScatterClass : Labels) en gardant la même échelle dans les trois directions. L'interprétation de la figure (ci-dessus) se fait sans difficulté et l'analyse est particulièrement adaptée aux objectifs biologiques.

Les trois strates forestières sont quasiment équivalentes. La disparité inter-régionale y est faible, sans être nulle. La typologie se renforce dans le gradient d'ouverture et on peut parler soit de convergence des communautés dans les milieux fermés, soit de divergence des contenus faunistiques dans les milieux ouverts. La deuxième formulation est plus en accord avec la discussion des auteurs. Cet exemple contient une grande quantité d'information. L'analyse du *K*-tableaux n'est plus une analyse simple mais une analyse coordonnée de plusieurs tableaux. Il y en a plusieurs possibles et *il faut définir clairement son objectif*. On y cherche une structure commune et on voit comment elle se reproduit. Ici les différences entre régions (tableaux espèces-régions) varient en intensité d'un tableau à l'autre. On parle de compromis et de reproduction du compromis.

Noter le principe :



- 1 - Combinaison des tableaux de départ pour définir un compromis **X**
- 2 - Projections en lignes supplémentaires dans l'analyse de **X**
- 3 - Projections en colonnes supplémentaires dans l'analyse de **X**.

Cette stratégie est étendue à tous les types d'analyses dans STATIS: Table averaging. L'exemple de la documentation de l'option précise la nature des sorties nombreuses du programme. On peut en particulier projeter sur les plans du compromis les axes principaux et les composantes principales de chacun des tableaux, ce qui introduit aux pratiques plus élaborées de STATIS et de l'AFM complète.

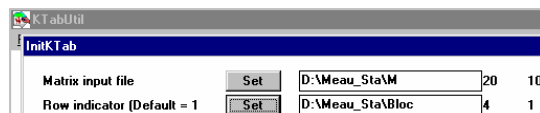
4.2. STATIS sur les tableaux

La situation est celle de K tableaux ayant mêmes lignes et mêmes colonnes, donc de K schémas ayant même pondération des lignes et même pondération des colonnes du type :

$$(X_1, Q, D) \quad (X_2, Q, D) \cdots (X_K, Q, D)$$

On peut suivre les raisonnements sur un exemple : 4 dates d'échantillonnage, 5 stations sur une rivière, 10 variables environnementales (carte Meau_Sta).

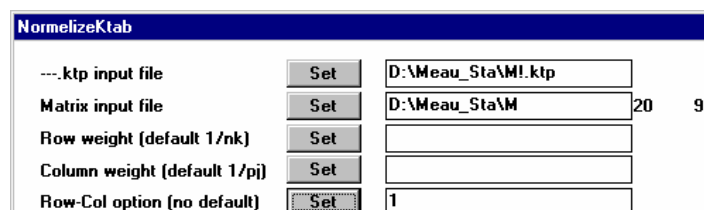
1	10.00	41.00	8.50	295.00	2.30	1.40	0.12	3.40	0.11
2	11.00	158.00	8.30	315.00	7.60	3.30	2.85	2.70	1.50
3	11.00	198.00	8.50	290.00	3.30	1.50	0.40	4.00	0.10
4	12.00	280.00	8.60	290.00	3.50	1.50	0.45	4.00	0.73
5	13.00	322.00	8.50	285.00	3.60	1.60	0.48	4.60	0.84
6	13.00	62.00	8.30	325.00	2.30	1.80	0.11	3.00	0.13
7	13.00	80.00	7.60	380.00	21.00	5.70	9.80	0.80	3.65
8	15.00	100.00	7.80	385.00	15.00	2.50	7.90	7.70	4.50
9	16.00	140.00	8.00	360.00	12.00	2.60	4.90	8.40	3.45
10	15.00	160.00	8.40	345.00	1.70	1.90	0.22	10.00	1.74
11	1.00	25.00	8.40	315.00	1.60	0.50	0.07	6.40	0.03
12	3.00	63.00	8.00	425.00	36.00	8.00	12.50	2.20	6.50
13	2.00	79.00	8.10	350.00	7.10	1.90	2.70	13.20	3.70
14	3.00	85.00	8.30	330.00	2.00	1.40	0.42	12.00	1.60
15	2.00	72.00	8.60	305.00	1.60	0.90	0.10	9.50	1.25
16	3.00	118.00	8.00	325.00	1.60	1.20	0.17	1.80	0.19
17	3.00	252.00	8.30	360.00	9.50	2.90	2.52	4.60	1.60
18	3.00	315.00	8.30	370.00	8.70	2.80	2.80	4.80	2.85
19	3.00	498.00	8.30	330.00	4.80	1.60	1.04	4.40	0.82
20	2.00	390.00	8.20	330.00	1.70	1.20	0.56	5.00	0.60



```
Option: InitKTab
Input file: D:\Meau_Sta\M
Row indicator file: D:\Meau_Sta\Bloc
Col indicator file:
Output ASCII file: D:\Meau_Sta\M!.ktp
```

```
4:5/5/5/5/
1:9/
```

Les tableaux superposés ont en commun les variables colonnes. Normaliser par tableau :



On a ainsi quatre schémas d'ACP normée (à la pondération près des colonnes) portant sur des tableaux à 5 lignes et 10 colonnes avec une pondération uniforme (1/5) des lignes et une pondération uniforme (1/10) des colonnes. On aurait pu utiliser une normalisation de type particulier (variables centrées par tableau puis normées sur

l'ensemble ¹⁵ pour éliminer l'effet des unités de mesure mais on laisse varier l'hétérogénéité entre blocs). Les schémas individuels sont donc des ACP normées comparables.

On peut faire les quatre ACP séparées d'un seul coup :

```

KTA-MFA
Separate analyses
---.ktt input file  Set  D:\Meau_Sta\M.ktt 20 9

Row bloc: 1 - Col bloc: 1 - Total inertia: 1
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum |
01  +6.7775E-01 +0.6778  +0.6778  |02  +2.8924E-01 +0.2892  +0.9670 |
03  +2.3521E-02 +0.0235  +0.9905  |04  +9.4891E-03 +0.0095  +1.0000 |
05  +0.0000E+00 +0.0000  +1.0000

-----

Row bloc: 2 - Col bloc: 1 - Total inertia: 1
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum |
01  +6.1165E-01 +0.6116  +0.6116  |02  +3.3144E-01 +0.3314  +0.9431 |
03  +4.4647E-02 +0.0446  +0.9877  |04  +1.2268E-02 +0.0123  +1.0000 |
05  +0.0000E+00 +0.0000  +1.0000

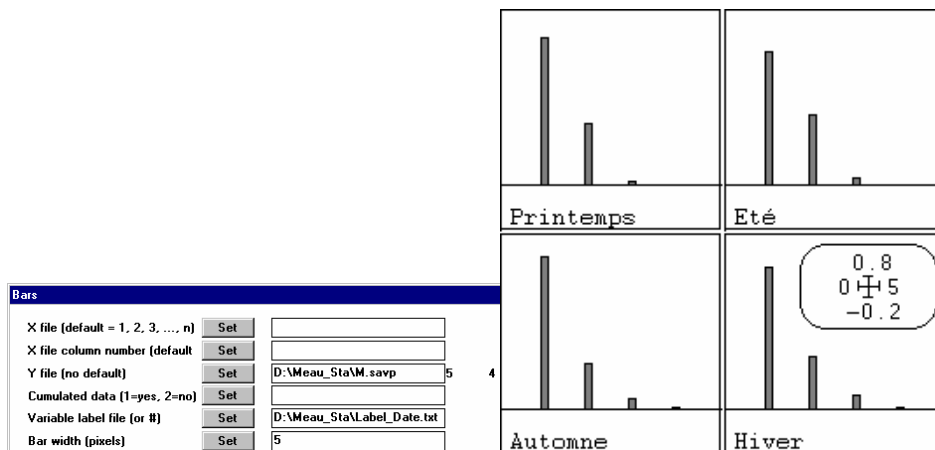
-----

Row bloc: 3 - Col bloc: 1 - Total inertia: 1
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum |
01  +7.0308E-01 +0.7031  +0.7031  |02  +2.1877E-01 +0.2188  +0.9219 |
03  +6.1303E-02 +0.0613  +0.9832  |04  +1.6845E-02 +0.0168  +1.0000 |
05  +0.0000E+00 +0.0000  +1.0000

-----

Row bloc: 4 - Col bloc: 1 - Total inertia: 1
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum |
01  +6.5408E-01 +0.6541  +0.6541  |02  +2.4861E-01 +0.2486  +0.9027 |
03  +7.7171E-02 +0.0772  +0.9799  |04  +2.0138E-02 +0.0201  +1.0000 |
05  +0.0000E+00 +0.0000  +1.0000
    
```

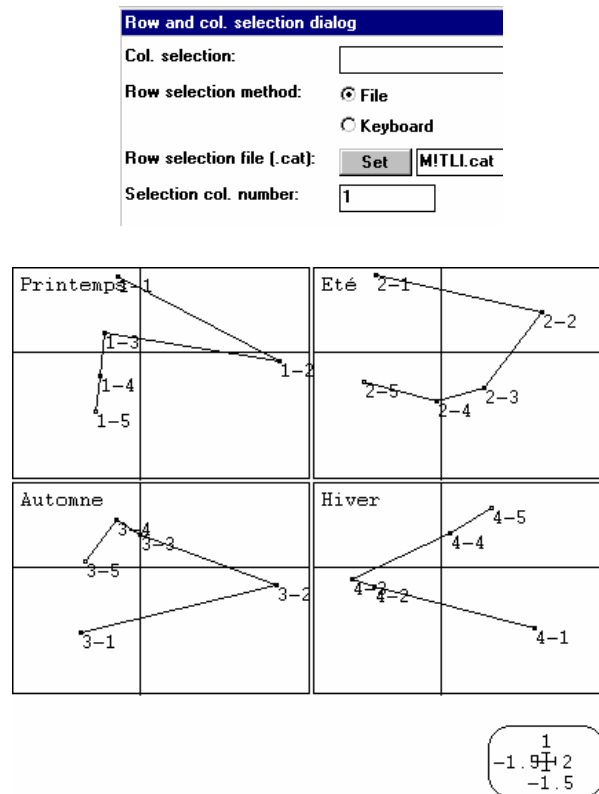
Les tableaux sont étiquetés Printemps, Été, Automne et Hiver. Représenter les quatre paquets de valeurs propres :



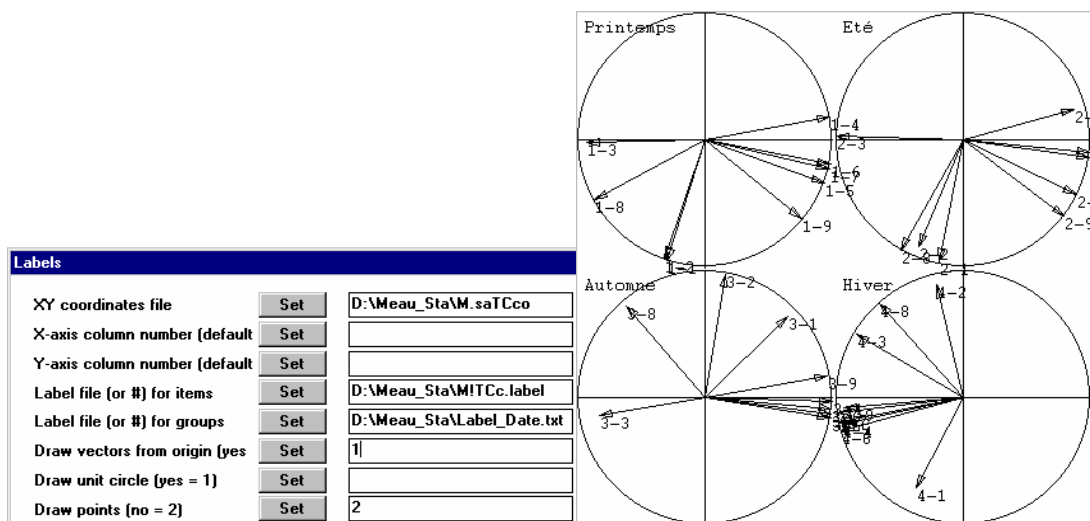
Représenter les quatre cartes individus :

```

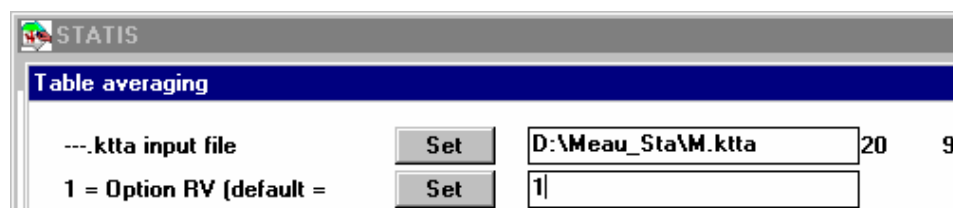
Trajectories
XY coordinates file  Set  D:\Meau_Sta\M.saTLLi 20 4
X-axis column number (default)  Set
Y-axis column number (default)  Set
Label file (or #) for items  Set  D:\Meau_Sta\MITLLi.label
Label file (or #) for groups  Set  D:\Meau_Sta\Label_Date.txt
Draw points (no = 2)  Set
Constrain H/V ratio (yes = 1)  Set  1
    
```



Noter l'utilisation des fichiers utilitaires des structures K-tableaux. Représenter les 4 cartes de variables :



Le but est de coordonner ces quatre analyses et représenter simultanément les individus et les variables des quatre tableaux. On utilise pour ce faire une analyse de référence dite analyse du compromis.



Paramètres d'entrée :

Option : Table averaging
 Input file D:\Meau_Sta\M
 Number of rows: 20, columns: 9

Notons n le nombre de lignes et p le nombre de colonnes de chacune des analyses séparées, \mathbf{D}_n et \mathbf{D}_p les normes associées. Les schémas séparés sont :

$$(\mathbf{X}_1, \mathbf{D}_p, \mathbf{D}_n) (\mathbf{X}_2, \mathbf{D}_p, \mathbf{D}_n) \cdots (\mathbf{X}_K, \mathbf{D}_p, \mathbf{D}_n)$$

On peut calculer un produit scalaire **entre tableaux** :

$$Covv(\mathbf{X}_k, \mathbf{X}_j) = Trace(\mathbf{X}_k^t \mathbf{D}_n \mathbf{X}_j \mathbf{D}_p) = Trace(\mathbf{X}_j^t \mathbf{D}_n \mathbf{X}_k \mathbf{D}_p)$$

Cette fonction est bilinéaire, symétrique, positive ($Vav(\mathbf{X}) = Trace(\mathbf{X}^t \mathbf{D}_n \mathbf{X} \mathbf{D}_p) \geq 0$) et non dégénérée.

Cette quantité a une signification expérimentale simple. C'est la somme pondérée par \mathbf{D}_p (donc ici la moyenne ordinaire) des corrélations entre les variables du tableau k et les variables correspondantes du tableau j . La variance vectorielle associée est simplement l'inertie totale de chaque analyse.

D'où le coefficient de corrélation entre deux tableaux :

$$RV(\mathbf{X}_k, \mathbf{X}_j) = \frac{Covv(\mathbf{X}_k, \mathbf{X}_j)}{\sqrt{Vav(\mathbf{X}_k)} \sqrt{Vav(\mathbf{X}_j)}}$$

```
----- Correlation matrix -----
[ 1] 1000
[ 2] 693 1000
[ 3] 789 767 1000
[ 4] 283 534 479 1000
-----
```

Ceci signifie qu'en moyenne les variables du tableau 1 et celles du tableau 2 présente une corrélation de 0.693. Les RV sont élevés, mais la structure du tableau 4 est manifestement la plus éloignée du groupe des 3 autres.

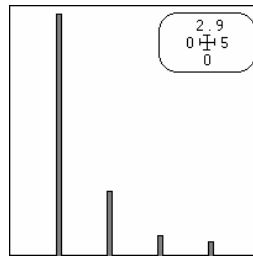
```
File D:\Meau_Sta\M.xa+RV contains the cosines between tables
It has 4 rows and 4 columns
File D:\Meau_Sta\M.xa+CV contains the inner products between tables
It has 4 rows and 4 columns
```

La matrice diagonalisée est la matrice des RV . Dans l'option *Covv*, on suppose que des tableaux de mêmes individus et mêmes variables présentent des inerties de même ordre de grandeur et que c'est un fait qui rentre en ligne de compte. Ici, les deux options sont identiques, car les variances vectorielles sont les inerties des nuages, donc les moyennes des variances des variables (chaque colonne a un poids uniforme) et valent l'unité, car les variables sont normalisées. La matrice des RV est diagonalisée :

```
Compromise with the RV coefficient matrix
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum  |
01  +2.8121E+00 +0.7030 +0.7030  |02  +7.5415E-01 +0.1885 +0.8916  |
03  +2.5368E-01 +0.0634 +0.9550  |04  +1.8003E-01 +0.0450 +1.0000  |
```

```
File D:\Meau_Sta\M.xa+EV contains the eigenvalues of the interstructure analysis
```

It has 4 rows and 1 columns

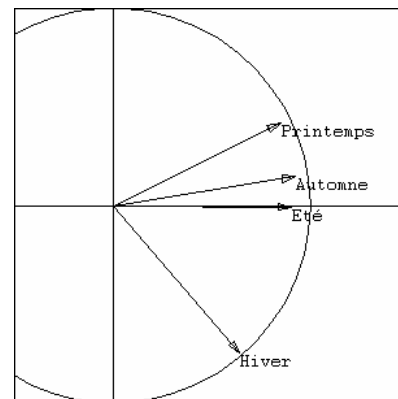


File D:\Meau_Sta\M.xa+IS contains the table scores of the interstructure analysis

It has 4 rows and 4 columns

On obtient ainsi une représentation euclidienne des tableaux :

Labels		
XY coordinates file	Set	D:\Meau_Sta\M.xa+IS
X-axis column number (default)	Set	
Y-axis column number (default)	Set	
Label file (or #) for items	Set	D:\Meau_Sta\Label_Date.txt
Label file (or #) for groups	Set	
Draw vectors from origin (yes)	Set	1
Draw unit circle (yes = 1)	Set	1
Draw points (no = 2)	Set	2



Le premier axe représente la ressemblance entre tableaux. Les suivants expriment des différences de structure (l'hiver est seul concerné). Le premier vecteur propre a, de plus, pour fonction d'attribuer à chaque tableau un poids (Cf. weights ci-dessous). Le poids attribué au tableau 4 est moindre que celui des trois autres. La combinaison des tableaux utilisant ces poids est un nouveau tableau de synthèse combinant les tableaux initiaux à proportion de leurs apports à la description de la structure commune dite compromis.

File D:\Meau_Sta\M.xata contains the compromise

It has 5 rows and 9 columns

```
File :D:\Meau_Sta\M.xata
|Col.| Mini | Maxi |
|----|-----|-----|
| 1|-2.017e+00| 1.800e+00|
| 2|-3.128e+00| 1.956e+00|
...
| 8|-2.359e+00| 1.704e+00|
| 9|-2.462e+00| 2.251e+00|
|----|-----|-----|
```

Ce nouveau tableau, dont le contenu importe peu (ce sont des combinaisons des valeurs des tableaux initiaux avec des coefficients tous positifs), a pour fonction de définir des axes et des composantes, donc des vecteurs de \mathbb{R}^p et de \mathbb{R}^n , qui expriment la structure compromis. Le programme est donc consacré essentiellement à une recherche d'un compromis inter-tableaux et à l'étude de la structure de ce compromis. Le compromis est muni des poids des lignes et des poids des colonnes communs à tous les tableaux. On démontre facilement que, sous la contrainte $\sum_{k=1}^K \alpha_k^2 = 1$, le compromis est la combinaison $\sum_{k=1}^K \alpha_k \mathbf{X}_k$ qui maximise l'inertie totale du schéma $(\sum_{k=1}^K \alpha_k \mathbf{X}_k, \mathbf{D}_p, \mathbf{D}_n)$.

File D:\Meau_Sta\M.xapl contains the row weights

It has 5 rows and 1 column
 File D:\Meau_Sta\M.xapc contains the column weights
 It has 9 rows and 1 column

Ce schéma subit l'analyse standard et peut être traité comme une analyse ordinaire. Les valeurs propres de ce compromis sont :

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.9112E+00	+0.6796	+0.6796	02	+8.1084E-01	+0.2883	+0.9680
03	+6.7772E-02	+0.0241	+0.9921	04	+2.2308E-02	+0.0079	+1.0000
05	+0.0000E+00	+0.0000	+1.0000				

File D:\Meau_Sta\M.xavp contains the eigenvalues and relative inertia for each axis
 --- It has 5 rows and 2 columns

On a conservé les coordonnées de variance 1 et de variance lambda pour les lignes et les colonnes :

File D:\Meau_Sta\M.xac1 contains the column scores (Norm = 1 for column weights)

in the analysis of the compromise
 It has 9 rows and 2 columns

File :D:\Meau_Sta\M.xac1

Col.	Mini	Maxi
1	-1.341e+00	8.629e-01
2	-2.037e+00	-1.506e-02

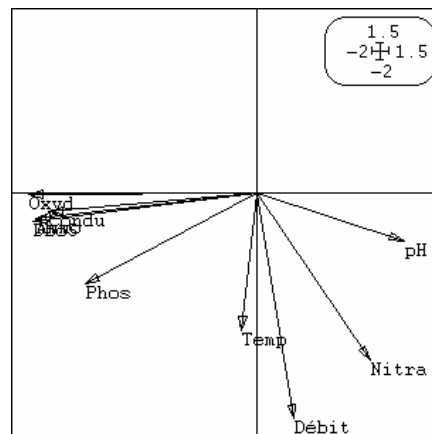
File D:\Meau_Sta\M.xaco contains the column scores (Norm = eigenvalue)

in the analysis of the compromise
 It has 9 rows and 2 columns

File :D:\Meau_Sta\M.xaco

Col.	Mini	Maxi
1	-1.854e+00	1.193e+00
2	-1.834e+00	-1.356e-02

Labels	
XY coordinates file	Set D:\Meau_Sta\M.xaco 9 2
X-axis column number (default	Set
Y-axis column number (default	Set
Label file (or #) for items	Set D:\Meau_Sta\Label_Var.txt



File D:\Meau_Sta\M.xall contains the row scores (Norm = 1 for row weights)

in the analysis of the compromise
 It has 5 rows and 2 columns

File :D:\Meau_Sta\M.xall

Col.	Mini	Maxi
1	-1.834e+00	8.943e-01
2	-9.731e-01	1.817e+00

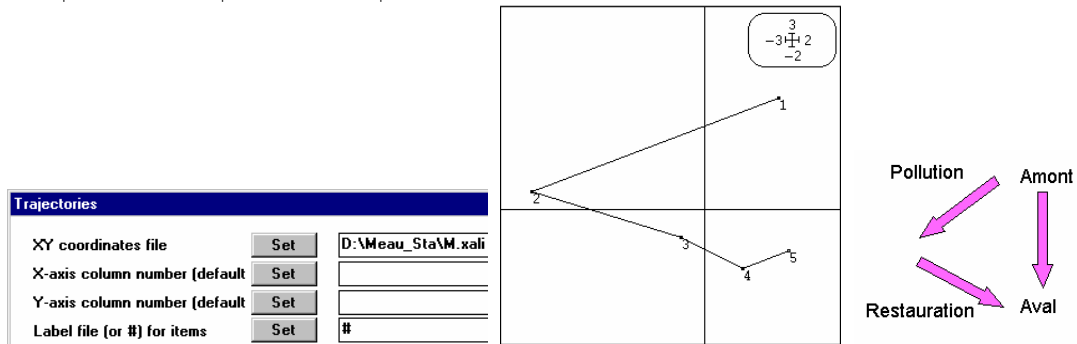
File D:\Meau_Sta\M.xali contains the row scores (Norm = eigenvalue)

in the analysis of the compromise
 It has 5 rows and 2 columns

File :D:\Meau_Sta\M.xali

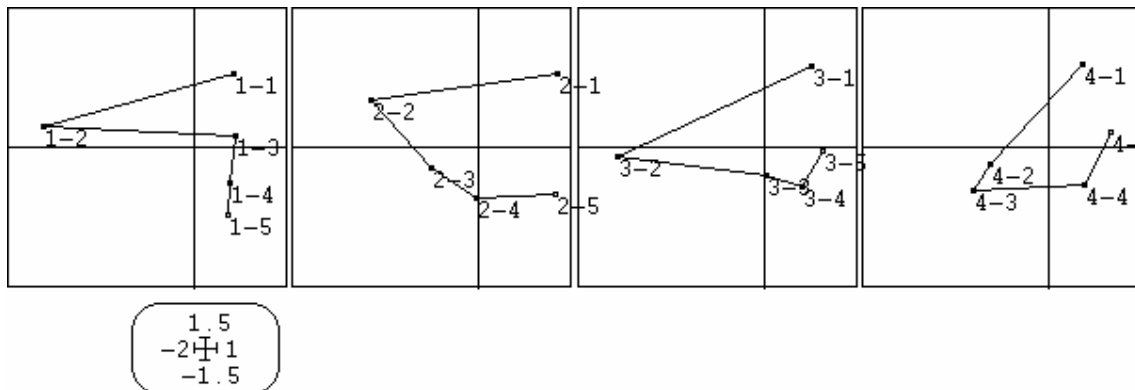
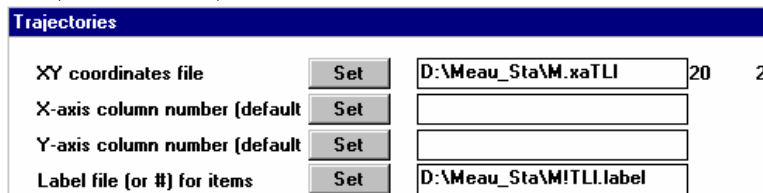
Col.	Mini	Maxi
1	-2.536e+00	1.236e+00

```
| 2|-8.762e-01| 1.637e+00|
|-----|-----|-----|
```



Ces coordonnées définissent la structure compromis. Ils sont liés aux axes et composantes de la manière habituelle. Pour voir comment chaque tableau reproduit la structure commune, on projette lignes et colonnes de chacun des tableaux : il s'agit alors de véritables projections.

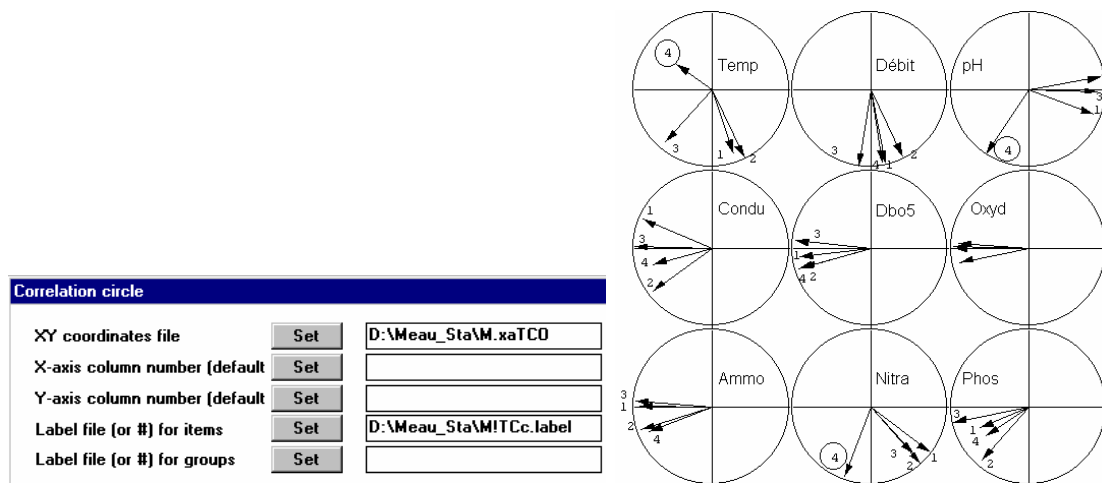
```
File D:\Meau_Sta\M.xaTl1 contains the standard row scores
It has 20 rows and 2 columns
It is to be used with --Tl1.label and --Tl1.cat files
File :D:\Meau_Sta\M.xaTl1
|Col.| Mini | Maxi |
|-----|-----|-----|
| 1|-1.610e+00| 8.389e-01|
| 2|-7.287e-01| 8.837e-01|
|-----|-----|-----|
```



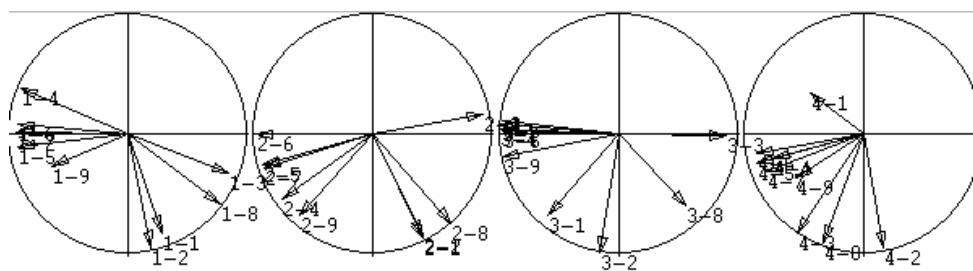
On voit la déformation du nuage de points au cours du temps.

```
File D:\Meau_Sta\M.xaTc0 contains the standard column scores
It has 36 rows and 2 columns
It is to be used with --Tc0.label and --Tc0.cat files
File :D:\Meau_Sta\M.xaTc0
|Col.| Mini | Maxi |
|-----|-----|-----|
| 1|-9.970e-01| 9.188e-01|
| 2|-9.794e-01| 3.755e-01|
|-----|-----|-----|
```

On verra de même la déformation de la structure de corrélations des variables :



par variable



par tableau

Le programme édite ensuite trois paramètres résumant la position d'un tableau par rapport au compromis. Les normes (inertie totale) sont égales à l'unité, les poids et les Cosinus carrés soulignent la ressemblance de trois des quatre tableaux et l'originalité relative des mesures d'hiver (tableau 4).

Typological value indices
 Rows = row number of each table
 Weights = Weights of tables in the compromise
 NS norm2 = Square norm (Hilbert-Schmidt) of tables
 Cos2 = Square cosine between table and approximated compromise

Number	Rows	Weights	NS norm2	Cos2
1	5	5.067e-01	1.000e+00	0.743
2	5	5.404e-01	1.000e+00	0.779
3	5	5.510e-01	1.000e+00	0.815
4	5	3.843e-01	1.000e+00	0.386

On peut encore projeter sur le plan d'inertie de synthèse les axes principaux de chaque analyse (\mathbb{R}^p) ou les composantes principales de chaque analyse (\mathbb{R}^n)

File D:\Meau_Sta\M.xaTComp contains the component scores issued from the separate analysis of each table
 It has 16 rows and 2 columns
 It is to be used with --T4A.label and --T4A.cat files

File :D:\Meau_Sta\M.xaTComp

Col.	Mini	Maxi
1	-5.702e-01	9.906e-01
2	-4.765e-01	9.595e-01

Correlation circle			Correlation circle		
XY coordinates file	Set	D:\Meau_Sta\M.xaTAxe	XY coordinates file	Set	D:\Meau_Sta\M.xaTComp
X-axis column number (default	Set		X-axis column number (default	Set	
Y-axis column number (default	Set		Y-axis column number (default	Set	
Label file (or #) for items	Set	D:\Meau_Sta\MIT4a.label	Label file (or #) for items	Set	D:\Meau_Sta\MIT4a.label
			Label file (or #) for groups	Set	

File D:\Meau_Sta\M.xaTAxe contains the axis scores issued from the separate analysis of each table

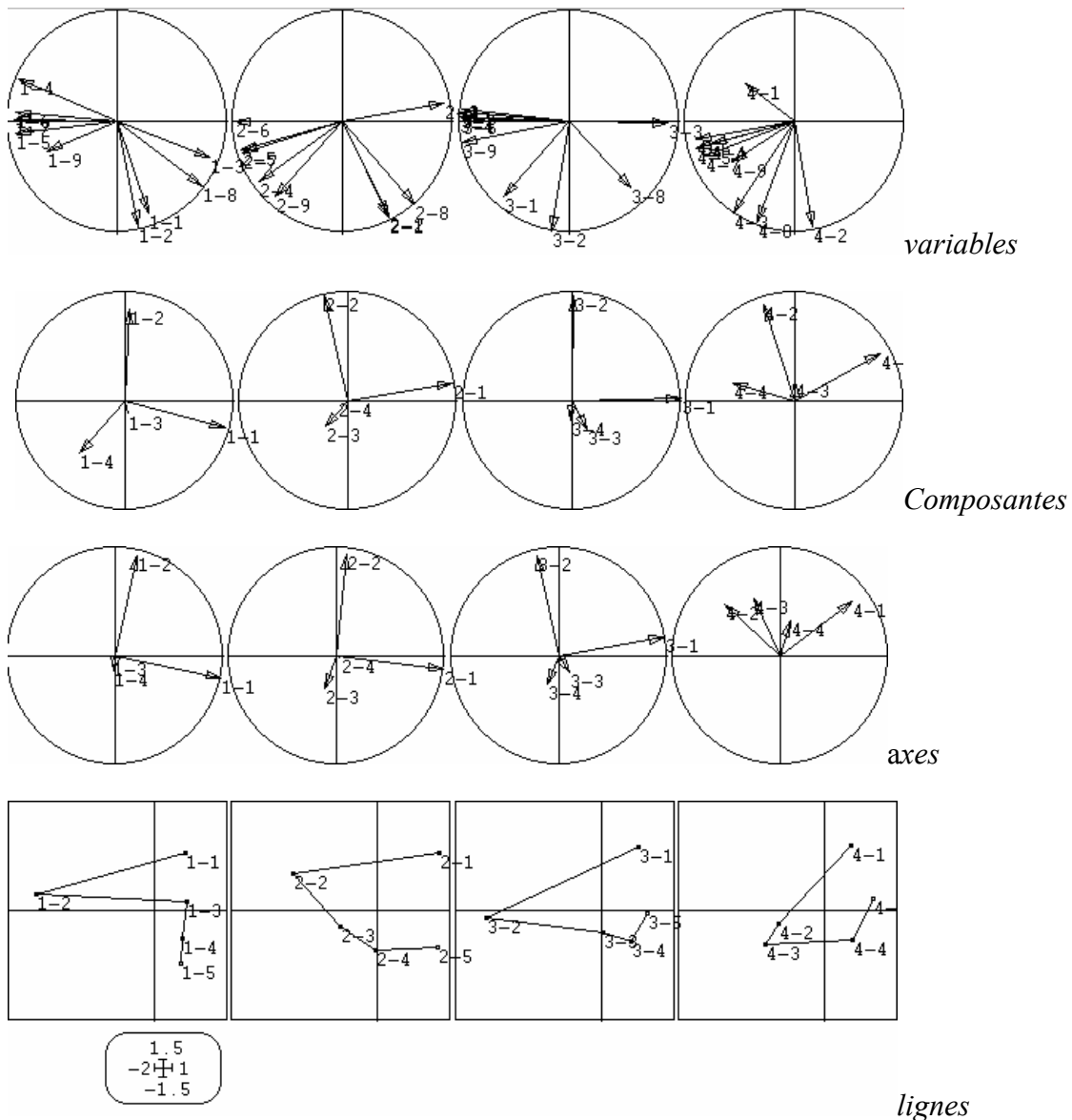
It has 16 rows and 2 columns

It is to be used with --T4A.label and --T4A.cat files

File :D:\Meau_Sta\M.xaTAxe

Col.	Mini	Maxi
1	-5.056e-01	9.775e-01
2	-3.010e-01	9.270e-01

On obtient une vue complète de quatre ACP simultanées :



Analyse triadique partielle

Sources dans ¹⁶, réponse en direction de l'ACP 3-modes dans ¹⁷, exemple dans ¹⁸ précisions dans ¹⁹. L'ACP 3-modes fait partie des méthodes qui étendent les propriétés de reconstitution de données. On sait approximer un tableau par :

$$x_{ij} = a_i b_j + r_{ij}$$

On peut donc chercher des modèles des cubes de données dont le plus simple est :

$$x_{ijk} = a_i b_j c_k + r_{ijk}$$

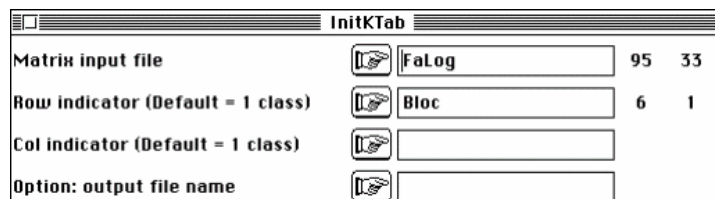
Une bibliographie considérable est associée à cette problématique ²⁰, ²¹, ²².

4.3. STATIS sur les opérateurs

La méthode d'analyse multi-tableaux STATIS tient son nom de Structuration des Tableaux À Trois Indices de la Statistique : bases théoriques dans les travaux de Y. Escoufier ²³, diffusion assurée par Ch. Lavit ²⁴. La méthode est reconnue comme un standard conceptuel dans plusieurs revues comparatives récentes.

L'objectif est de comparer plusieurs analyses de tableaux ayant en commun les lignes-individus (on dit STATIS sur les **WD**) ou les colonnes-variables (on dit STATIS sur les **VQ**). Le paragraphe précédent considérait les deux marges communes. Il n'y en a maintenant qu'une seule.

On peut accéder aux principes de cette méthode par le biais de l'exemple de la carte Fatala ²⁵. Passer le tableau Fatala en $\text{Log}(x + 1)$ pour limiter l'effet des dénombrements (Bin->Bin : $c * \text{Log}[a * x + b]$). Le fichier FaLog a 95 lignes (coups de filet) et 33 colonnes (espèces). Définir la structure du *K*-tableaux (**KTabUtil** : **InitKTab**) :



```
Option: InitKTab
Input file: FaLog
Row indicator file: Bloc
Col indicator file:
Output ASCII file: FaLog!.ktp
```

6 tableaux sont superposés. Ils ont en commun 33 colonnes (espèces). Le nombre de pêches par date est indiqué dans la partition en 6 blocs de lignes :

6 : 16/15/16/16/16/16/
1 : 33/

L'option STATIS : Operator averaging suppose que les tableaux sont appariés par les colonnes et prend un *K*-tableaux superposés. Mais cette structure ne préjuge pas si les

objets en commun sont des variables ou des individus (au sens statistique). Cela est décidé par l'utilisateur au moment du centrage.

```
Qualitative variable file: FaLog!TL1
Number of rows: 95, variables: 3, categories: 117
Auxiliary ASCII output file FaLog!TL1.label: label file
```

Numérotation des lignes par tableau (6 modalités-dates), au total (95 modalités-coups de pêche), dans un tableau (16 modalités-coups de pêche, dans le plus grand tableau) soit un total de 117 modalités :

```
Qualitative variable file: FaLog!TCc
Number of rows: 198, variables: 3, categories: 72
Auxiliary ASCII output file FaLog!TCc.label: label file
```

Numérotation des colonnes par tableau (6 modalités-dates), au total (33 modalités-espèces), dans un tableau (33 modalités-espèces) soit un total de 72 modalités :

```
Qualitative variable file: FaLog!T4a
Number of rows: 24, variables: 2, categories: 10
Auxiliary ASCII output file FaLog!T4a.label: label file
```

Numérotation des axes par tableau (6 modalités-dates) et dans un tableau (4 modalités-axes) soit un total de 10 modalités. Centrer le K -tableaux. L'opération (KTabUtil : CentringKtab, KTabUtil : NormelizeKtab, KTabUtil : COAKtab) est fondamentale. Les tableaux sont des tableaux faunistiques. On peut opter pour l'ACP (centrée par espèce) ou l'AFC (doublement centrée). On utilise l'AFC avec un double centrage par espèce et par coup de pêche :



```
Option coaktab
Input file FaLog
Number of rows: 95, columns: 33
-----
File FaLog_CA.ktpc contains the column weights
It has 33 rows and 1 column
-----
File FaLog_CA.ktpl contains the row weights (Sum inside block = 1)
It has 95 rows and 1 column
-----
File FaLog_CA.ktta contains the transformed table
It has 95 rows and 33 columns
-----
```

On récupère un K -tableaux formé de 6 triplets statistiques d'ACP doublement centrée avec les mêmes variables (colonnes) et des individus distincts (coups de pêche). On veut comparer ces 6 ACP, donc comparer 6 structures espèces-relevés dans 6 occasions différentes. Les relevés sont distribués sur 4 stations dans un gradient d'éloignement à l'océan (structure estuarienne).

L'enjeu est de savoir si l'hydrologie (débit variant de 10 à 400 m³/s) influe sur **la structure**, c'est-à-dire l'intensité de la séparation des espèces entre les biotopes utilisés. Ce passage du facteur de milieu intervenant sur la composition du peuplement (problème classique) au facteur de milieu intervenant sur la structuration est caractéristique du passage des méthodes de couplage aux méthodes de K -tableaux. Le

facteur (ici le temps qui intègre le débit, donc la salinité, et la température) définit les tableaux. Chaque tableau mérite une analyse séparée, qu'il convient de faire au moins pour bien comprendre l'objectif poursuivi.

Exécuter ces analyses séparées (KTA : Separate analyses) :

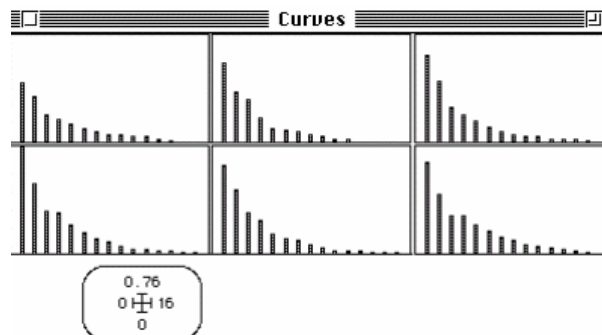


```
Row bloc: 1 - Col bloc: 1 - Total inertia: 1.7031
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum |
01  +4.2532E-01 +0.2497 +0.2497 |02  +3.2466E-01 +0.1906 +0.4404 |
03  +2.0163E-01 +0.1184 +0.5588 |04  +1.6415E-01 +0.0964 +0.6551 |
05  +1.3060E-01 +0.0767 +0.7318 |06  +1.0678E-01 +0.0627 +0.7945 |
...
```

```
Row bloc: 2 - Col bloc: 1 - Total inertia: 1.87619
Num. Eigenval.  R.Iner.  R.Sum  |Num. Eigenval.  R.Iner.  R.Sum |
01  +5.6552E-01 +0.3014 +0.3014 |02  +3.5924E-01 +0.1915 +0.4929 |
03  +3.0876E-01 +0.1646 +0.6575 |04  +1.8002E-01 +0.0960 +0.7534 |
05  +1.0634E-01 +0.0567 +0.8101 |06  +9.3390E-02 +0.0498 +0.8599 |
07  +8.4263E-02 +0.0449 +0.9048 |08  +5.9774E-02 +0.0319 +0.9366 |
...
```

Eigenvalues are stored by column into file FaLog_CA.savp
Row : 16 - Col (tables): 6

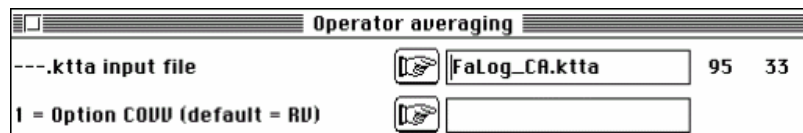
Représenter les valeurs propres des 6 analyses séparées (Curves : Bars) :



5 des 6 tableaux méritent une réduction en dimension 2. Faire les cartes espèces et relevés des 6 analyses séparées et noter combien une analyse de synthèse, coordonnée pour orienter les axes de manière unique est souhaitable. Tout est disponible dans les fichiers :

```
File FaLog_CA.saTlli contains row scores
in the separate analyses
It has 95 rows and 4 columns
It is to be used with --Tll.label and --Tll.cat files
File FaLog_CA.saTCco contains column scores
in the separate analyses
It has 198 rows and 4 columns
It is to be used with --TCc.label and --TCc.cat files
File FaLog_CA.sapa contains total inertia and rank
for each separate analyses
It has 6 rows and 2 columns
```

Exécuter l'analyse STATIS :



On ne peut plus comparer les tableaux qui ont des dimensions variables. Supposons, en toute généralité que les triplets séparés soient du type :

$$\left(\underset{n_1 \times p}{\mathbf{X}_1, \mathbf{D}, \mathbf{D}_1} \right), \left(\underset{n_2 \times p}{\mathbf{X}_2, \mathbf{D}, \mathbf{D}_2} \right), \dots, \left(\underset{n_K \times p}{\mathbf{X}_K, \mathbf{D}, \mathbf{D}_K} \right) \text{ (colonnes en commun)}$$

ce qui équivaut à :

$$\left(\underset{p \times n_1}{\mathbf{X}_1^t, \mathbf{D}, \mathbf{D}_1} \right), \left(\underset{p \times n_2}{\mathbf{X}_2^t, \mathbf{D}, \mathbf{D}_2} \right), \dots, \left(\underset{p \times n_K}{\mathbf{X}_K^t, \mathbf{D}, \mathbf{D}_K} \right) \text{ (lignes en commun)}$$

Le raisonnement est donc totalement symétrique entre appariement par les colonnes ou par les lignes. On retiendra le premier cas qui correspond à la disposition pratique des données. Ce qui peut être comparé d'une analyse à l'autre est dans la matrice :

$$\mathbf{E}_k = \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D}$$

Elles sont diagonalisées dans les analyses séparées. L'inertie du $k^{\text{ième}}$ triplet est :

$$I_k = \text{Trace}(\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D}) = \sum_{i=1}^p \lambda_{ik}$$

où λ_{ik} est la $i^{\text{ième}}$ valeur propre du $k^{\text{ième}}$ triplet.

celle de variance vectorielle ou norme d'opérateur (Escoufier 1973, op. cit.) :

$$Vav(\mathbf{X}_k) = \text{Trace}(\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D} \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D}) = \sum_i \lambda_i^2$$

A la notion de mesure de structure, on ajoute le mesure de co-structure par le produit scalaire utilisé dans la norme précédente :

$$\begin{aligned} \text{Covv}(\mathbf{X}_k, \mathbf{X}_j) &= \text{Trace}(\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D} \mathbf{X}_j^t \mathbf{D}_j \mathbf{X}_j \mathbf{D}) \\ &= \text{Trace}(\mathbf{X}_j \mathbf{D}_j \mathbf{X}_j^t \mathbf{D}_j \mathbf{X}_k \mathbf{D} \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_j) = \sum_i \mu_i \end{aligned}$$

Il s'agit exactement de la co-inertie totale associée au couple de tableaux qui se décompose dans la somme des valeurs propres de l'analyse de co-inertie.

Il s'en suit qu'on peut mesurer la corrélation entre deux triplets par le coefficient de corrélation vectoriel :

$$Rv(\mathbf{X}_k, \mathbf{X}_j) = \frac{\text{Covv}(\mathbf{X}_k, \mathbf{X}_j)}{\sqrt{Vav(\mathbf{X}_k)} \sqrt{Vav(\mathbf{X}_j)}}$$

```
Option operator averaging
Input file FaLog_CA
Number of rows: 95, columns: 33
```

```
----- Correlation matrix -----
[ 1] 1000
```



```
[ 2] 580 1000
[ 3] 409 387 1000
[ 4] 410 446 545 1000
[ 5] 524 514 411 419 1000
[ 6] 530 455 339 414 410 1000
```

Edition de la matrice des coefficients RV. Il n'y a pas de très fortes valeurs mais elles sont remarquablement constantes.

Ce coefficient *RV* peut être compris comme le rapport de la structure de co-inertie de deux tableaux (il existe une analyse de co-inertie appariée par les espèces pour chaque couple de dates) rapporté au produit de leur structure d'inertie (le rapport est compris entre 0 et 1). Ce coefficient RV est ainsi édité dans le listing de CoInertia : Coinertia analysis.

La matrice des RV est conservée dans le fichier ---.oaRV (oa nom générique pour Operator Averaging, car cette analyse fondamentalement cherche une structure moyenne d'opérateurs) :

```
File FaLog_CA.oa+RV contains the cosines between operators
It has 6 rows and 6 columns
```

Les tableaux ne sont pas comparables, car ils n'ont pas les mêmes lignes. Par contre les matrices de produits scalaires interspécifiques sont comparables (puisque les espèces sont les mêmes). On va donc construire une pseudo-matrice de covariances (au sens large, puisqu'on est ici dans une logique AFC) en combinant les 6 matrices observées. Le listing rappelle qu'on a choisi l'option diagonalisation de la matrice des RV :

```
Compromise with the cosines between operators
```

Le listing donne les valeurs propres de la matrice des RV :

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+3.2701E+00	+0.5450	+0.5450	02	+7.9391E-01	+0.1323	+0.6773
03	+6.0299E-01	+0.1005	+0.7778	04	+4.8605E-01	+0.0810	+0.8588
05	+4.6503E-01	+0.0775	+0.9363	06	+3.8197E-01	+0.0637	+1.0000

Cette phase est classiquement appelée **analyse de l'interstructure**. Les valeurs propres de la matrice des *RV* sont conservées dans le fichier ---.oa+EV :

```
File FaLog_CA.oa+EV contains the eigenvalues of the interstructure analysis
It has 6 rows and 1 columns
```

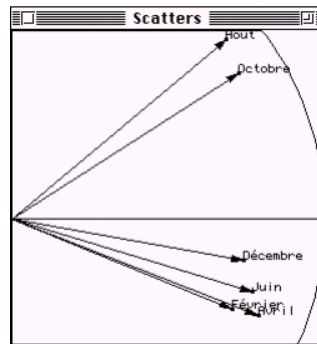
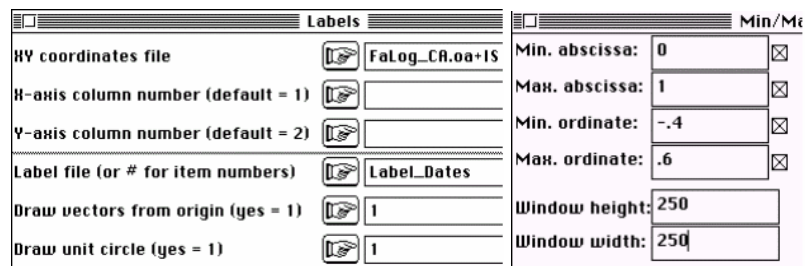


L'analyse vaut la peine si on distingue clairement une première valeur propre isolée. C'est bien le cas. Il s'agit d'une ACP non centrée dans l'espace des matrices de covariances.

Les vecteurs propres de la matrice des *RV* sont conservés dans le fichier ---.oa+IS :

```
File FaLog_CA.oa+IS contains the table scores of the interstructure
analysis
It has 6 rows and 6 columns
```

Ces vecteurs propres donnent une image euclidienne de la ressemblance entre tableaux mesurée par les *RV* (ScatterClass : Labels) :



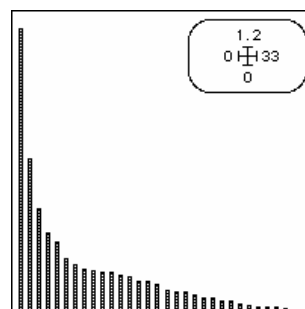
Le premier vecteur propre a 6 composantes qui sont toujours du même signe. Les positions sont assez homogènes et les 6 tableaux vont participer à la définition du compromis. Ces positions sont des poids affectés à chacune des matrices de covariance qui permettent de les combiner pour former une matrice de covariances dite matrice du compromis.

On note l'existence du compromis, en même temps que l'isolement des deux dates de la saison des pluies (août et octobre, 3 et 4). La méthode ne peut faire une **typologie de structure**, ce qui limite le terme d'analyse de l'interstructure, mais indique qu'une telle typologie de structure serait souhaitable.

La matrice de covariance synthétique est diagonalisée : on dit qu'on fait l'**analyse du compromis**. Les valeurs propres de cette diagonalisation sont éditées :

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.1299E+00	+0.2132	+0.2132	02	+6.1256E-01	+0.1156	+0.3289
03	+4.1679E-01	+0.0787	+0.4075	04	+3.2350E-01	+0.0611	+0.4686
...							
31	+1.4375E-02	+0.0027	+0.9993	32	+3.8901E-03	+0.0007	+1.0000
33	+0.0000E+00	+0.0000	+1.0000				

Ces valeurs propres sont affichées à l'écran et on demande de choisir un nombre réduit de facteurs. Il est légitime d'en conserver au moins 2 :

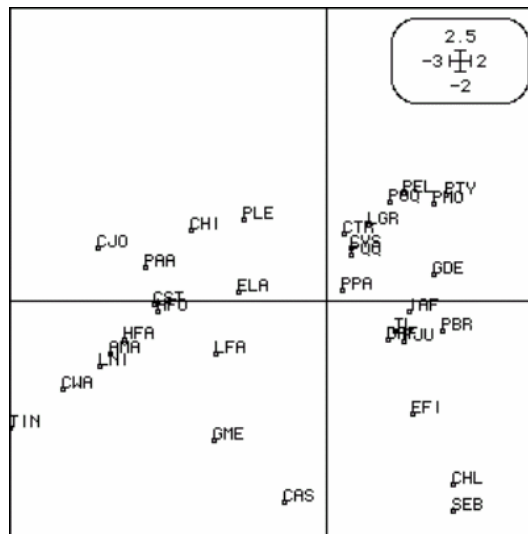


Les valeurs propres du compromis sont conservées dans un fichier ---.oa+vp :

File FaLog_CA.oa+vp contains the eigenvalues of the compromise analysis
It has 33 rows and 1 columns

On les juge comme dans une analyse ordinaire. Fondamentalement les poids des tableaux sont choisis pour que la somme des carrés des valeurs propres du compromis soit optimale, c'est-à-dire de manière à ce que ce graphe de valeurs propres soit le plus explicite possible. Le compromis est de dimension importante et on garde deux axes pour simplifier l'illustration. La diagonalisation donne des scores aux colonnes (ici espèces) et fait donc une typologie de référence des taxons :

File FaLog_CA.oa+co contains the column scores (analysis of the compromise)
It has 33 rows and 2 columns

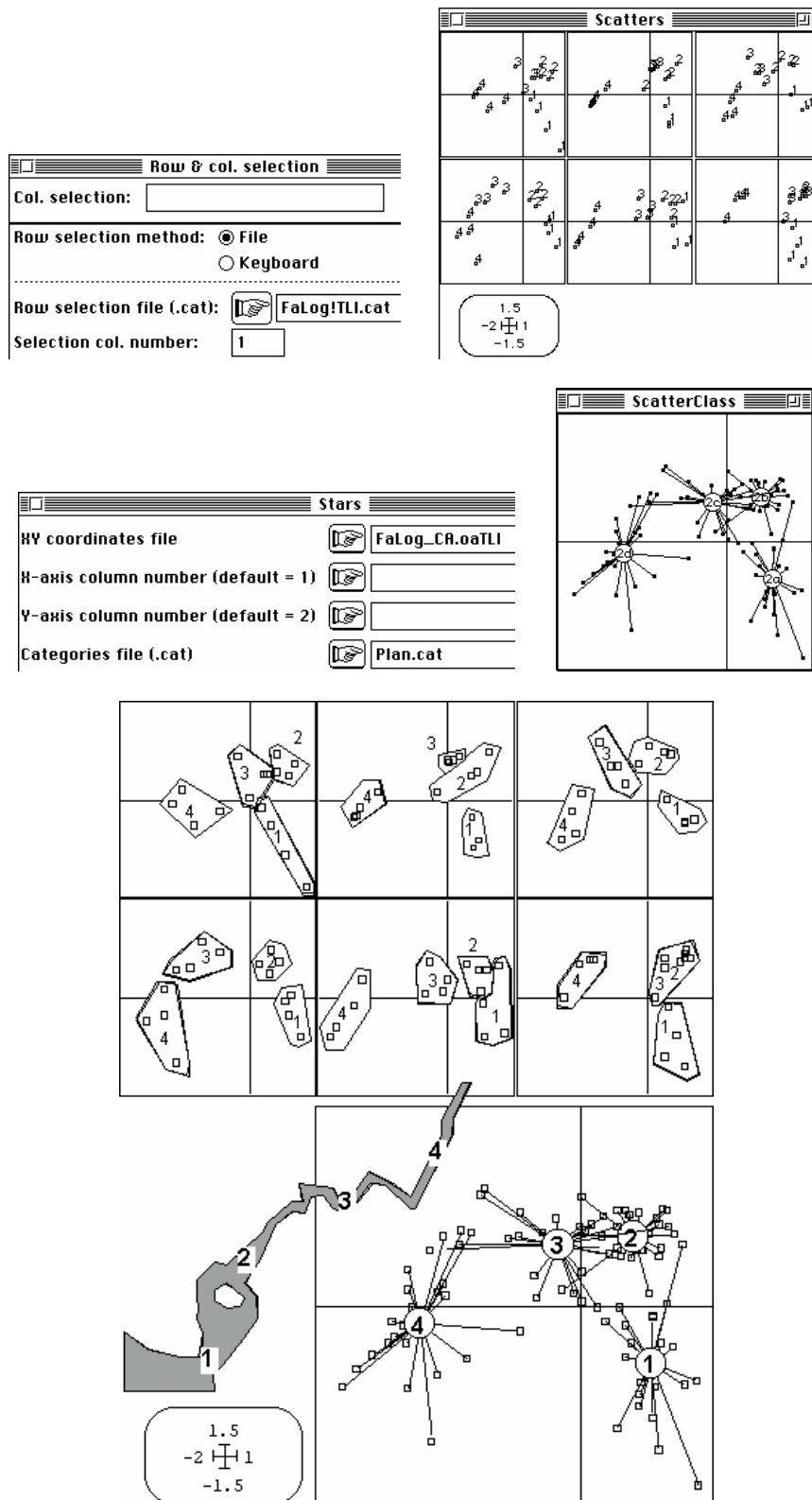


La méthode donne un repère de référence dans \mathbb{R}^{33} (la dimension commune) qui s'interprète du point de vue des espèces (dans la logique d'une analyse en coordonnées principales, voir Distances : Principal Coordinates). Dans ce repère on peut projeter les 6 nuages de points :

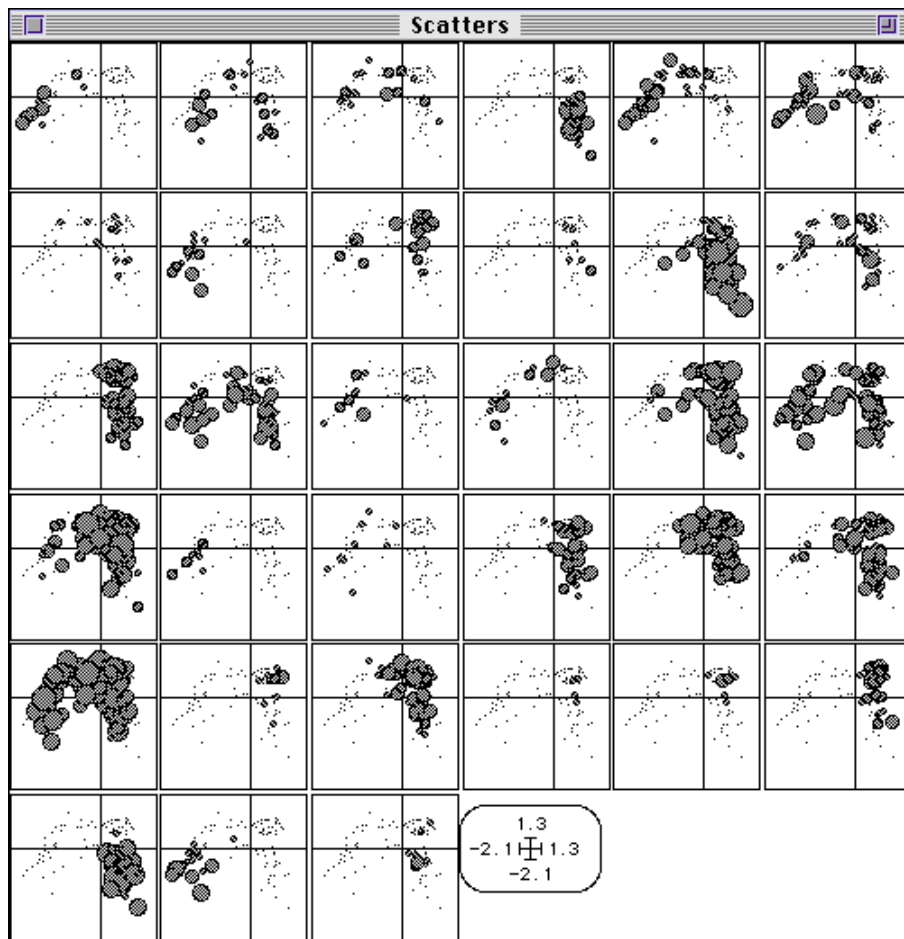
File FaLog_CA.oaTLI contains the standard column scores with lambda norm
It has 95 rows and 2 columns

Dans Scatters : Labels puis ScatterClass : Stars :

Labels	
#Y coordinates file	<input type="text" value="FaLog_CA.oaTLI"/>
#X-axis column number (default = 1)	<input type="text"/>
#Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="text" value="Label_Sta"/>



Typiquement, la structure spatiale des stations (observées par 4 répétitions) donne une structure faunistique globale reproduite de manière variable. Le gradient attendu est exprimé de manière optimale au cœur de la saison humide (mois n° 4).



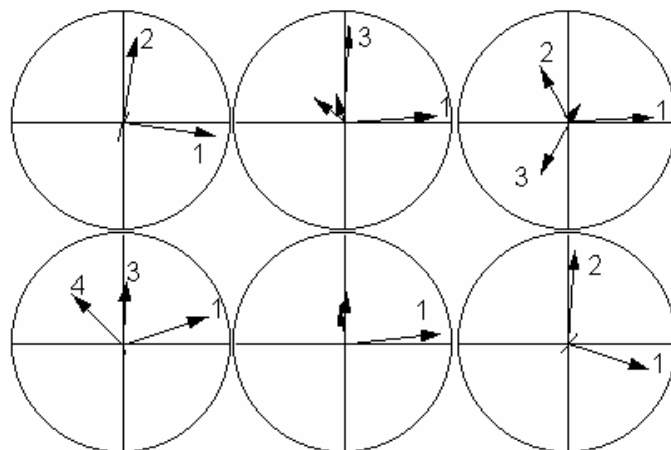
Retour aux données

Sur le plan de référence, on peut projeter les axes principaux de chacun des nuages :

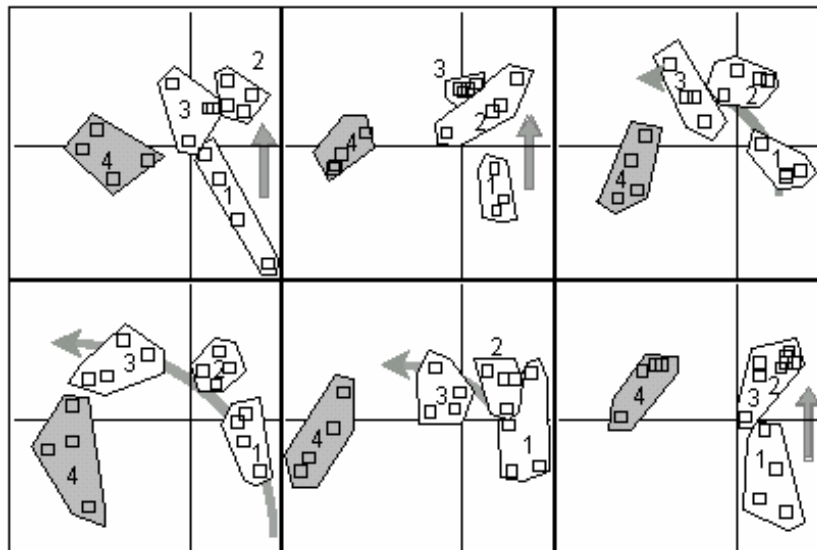
File FaLog_CA.ooT4a contains the axis scores of the separate analyses
 It has 24 rows and 2 columns
 It is to be used with --T4a.label and --T4a.cat files

Labels	
HV coordinates file	FaLog_CA.ooT4a
X-axis column number (default = 1)	
Y-axis column number (default = 2)	
Label file (or # for item numbers)	FaLog!T4a.label

Row & col. selection	
Col. selection:	
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	FaLog!T4a.cat
Selection col. number:	1



L'axe 1 du compromis est l'axe 1 de chacune des analyses séparées, ce qui indique que la partie permanente de la structure est l'originalité de la station 4. L'axe 2 du compromis est approximativement les axes 2 des dates 1 et 6 (saison sèche) et l'axe 3 de la date 2. En saison humide la structure des tableaux n'est pas exprimée et ce qui s'exprime dans le plan du compromis est très différent. On enrichit l'interprétation :



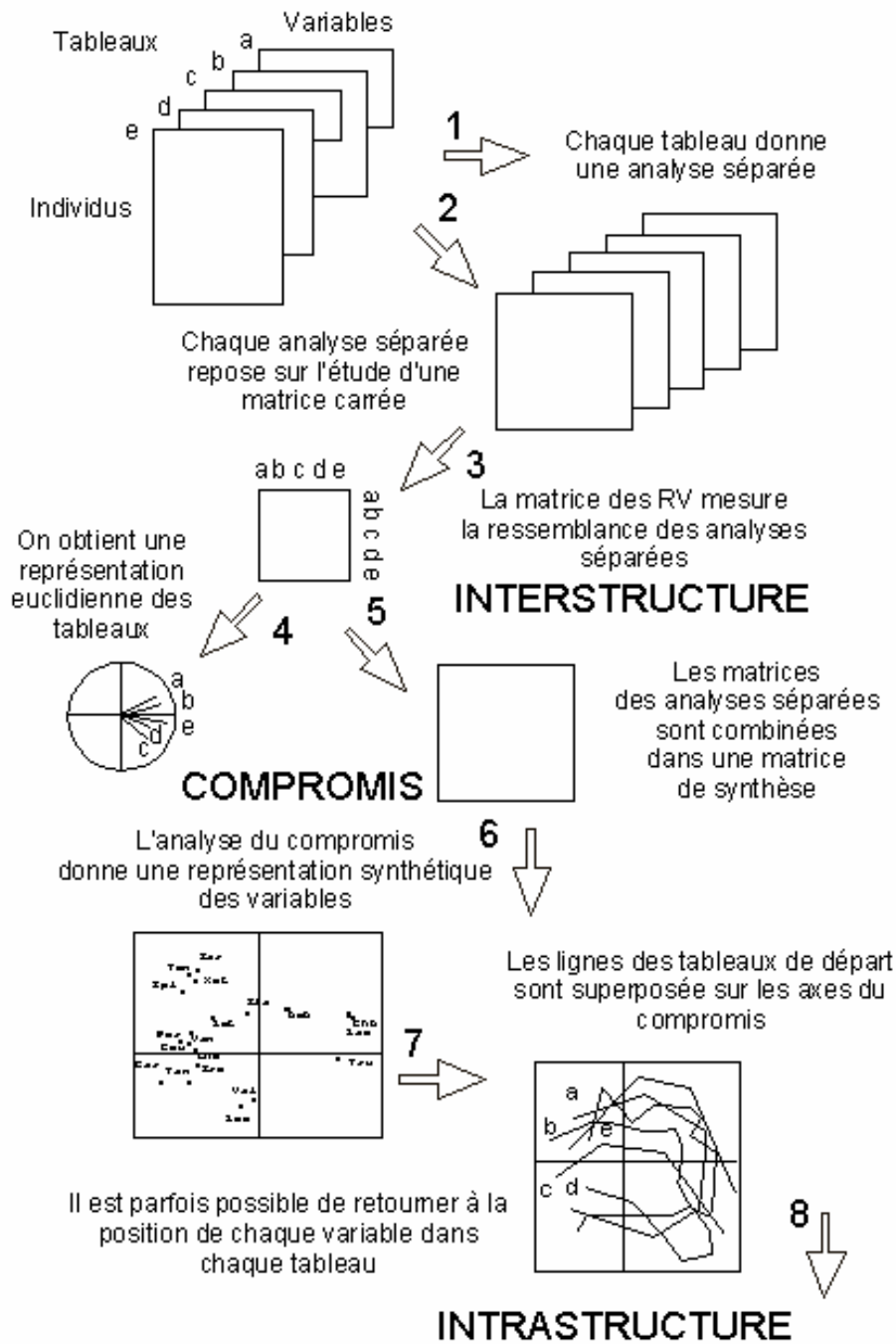
L'analyse s'achève par :

Typological value indices
 Rows = row number of each table
 Weights = Weights of operators in the compromise
 NS norm2 = Square norm (Hilbert-Schmidt) of operators
 Cos2 = Square cosine between operator and approximated compromise

Number	Rows	Weights	NS norm2	Cos2
1	16	4.362e-01	4.008e-01	0.418
2	15	4.266e-01	6.108e-01	0.461
3	16	3.805e-01	7.186e-01	0.327
4	16	4.012e-01	1.088e+00	0.394
5	16	4.109e-01	7.988e-01	0.372
6	16	3.914e-01	8.644e-01	0.330

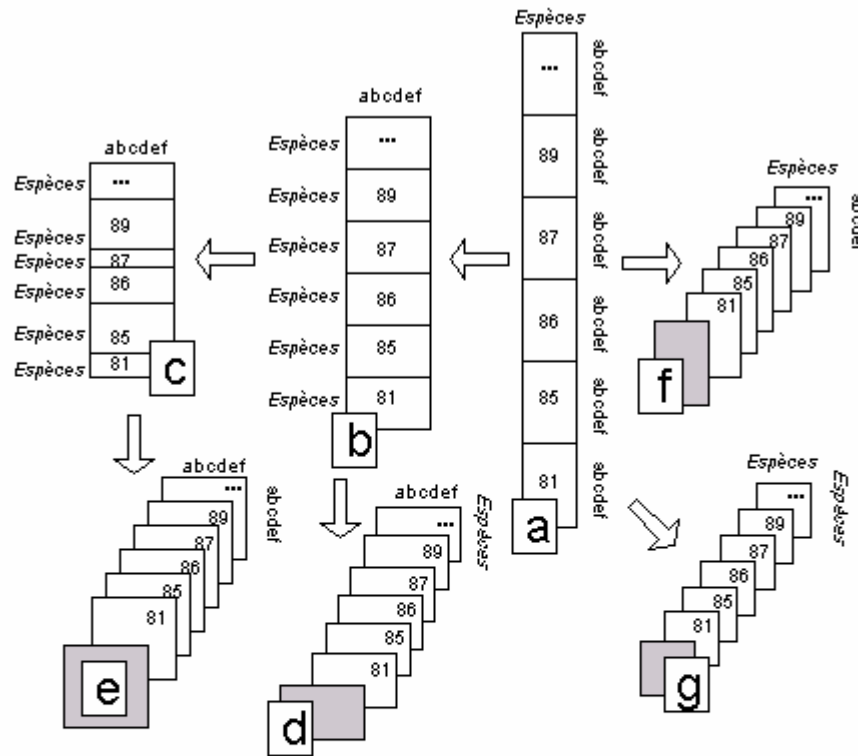
On a le nombre de points par tableau, son poids dans la constitution du compromis (déjà observé comme très stable), la norme d'Hilbert-Schmit de l'opérateur associé qui varie dans une proportion de 1 à 2.5 et est optimum en saison humide (ce qui indique que la typologie forte de ces dates n'est pas exprimée totalement dans le compromis, du moins dans ses deux premiers axes) et le cosinus carré entre l'opérateur à une date donnée et l'opérateur compromis reconstitué par les axes conservés (donc ici la reconstitution de rang 2). Il n'y a pas de fortes valeurs (maximum à l'unité) mais une grande stabilité. Dans la structure du compromis une part de l'information sur la structure de certaines dates est perdue. Le compromis est une structure moyenne calculée puis analysée. Cette typologie moyenne est ensuite reconstruite par chaque tableau et cette reconstruction varie d'une date à l'autre. L'analyse simple de ce type de tableau fait évidemment l'impasse sur de telles observations. STATIS sur les **VQ** est du type intraclasse mais la recherche des axes communs est beaucoup plus sophistiquée et plus solide.

A retenir le point d'orientation :



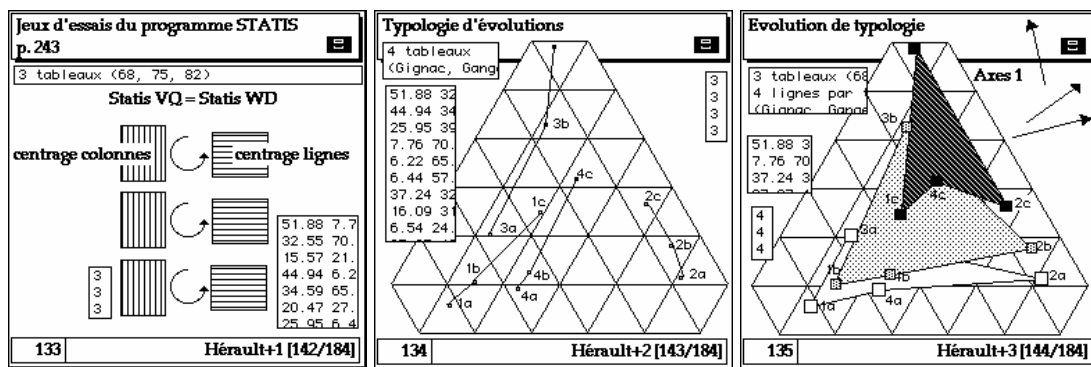
L'intrastructure a un sens précis, par exemple dans ²⁶.

Il existe enfin des cas où les STATIS sur les tableaux et STATIS sur les opérateurs sont en concurrence. On peut très bien avoir des tableaux totalement appariés et utiliser une méthode portant sur les opérateurs. On multiplie malheureusement les choix possibles :



- a - K-tableaux des données saisies
- b - K-tableaux transposé
- c - K-tableaux réduit
- d - Compromis de tableaux SE
- e - Compromis d'opérateurs SS
- f - Compromis de tableaux ES
- g - Compromis d'opérateurs EE

Une toute petite illustration montre qu'il faut vraiment orienter la recherche :

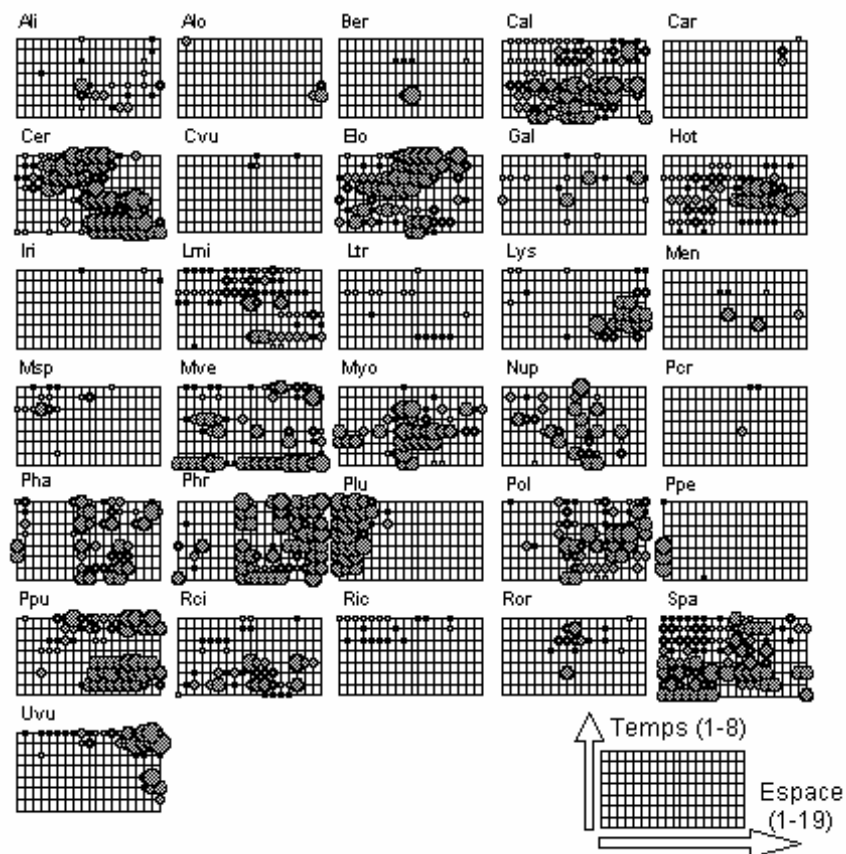
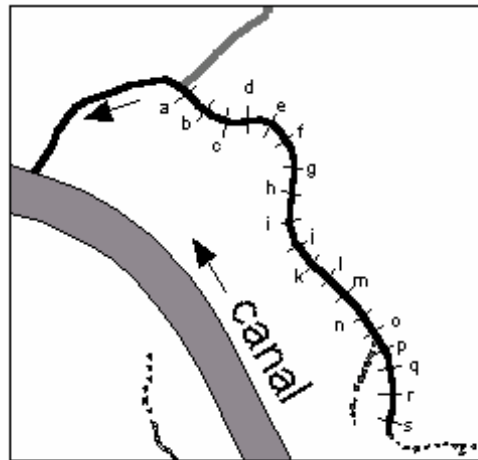


4.4. Analyses de co-inertie multiple

La méthode STATIS a pour fondement de construire un compromis puis de l'analyser pour le décomposer en une série d'axes dont on garde les principaux.

On peut faire l'inverse : choisir progressivement des axes pour construire un compromis. Les deux conduisent en général à des résultats voisins par des moyens

théoriques complètement différents. Dans la deuxième famille, on retrouve l'analyse factorielle multiple déjà rencontrée et l'analyse de co-inertie multiple (voir ²⁷ où on trouve une comparaison des trois méthodes).



Cube de données espèces-espace-temps ²⁸

Dans les méthodes de couplage on voit bien que deux tableaux appariés par les lignes sont soit deux ensembles de variables dans le même espace (analyse canonique) soit deux ensembles de points dans deux espaces différents (analyse de co-inertie) et sont les deux. Mais on a perdu la possibilité de tout faire à la fois comme pour un tableau.

STATIS résout la difficulté en passant dans les espaces d'opérateurs. La difficulté est sévère quand lignes et colonnes n'ont pas clairement le statut d'individus ou de variables.

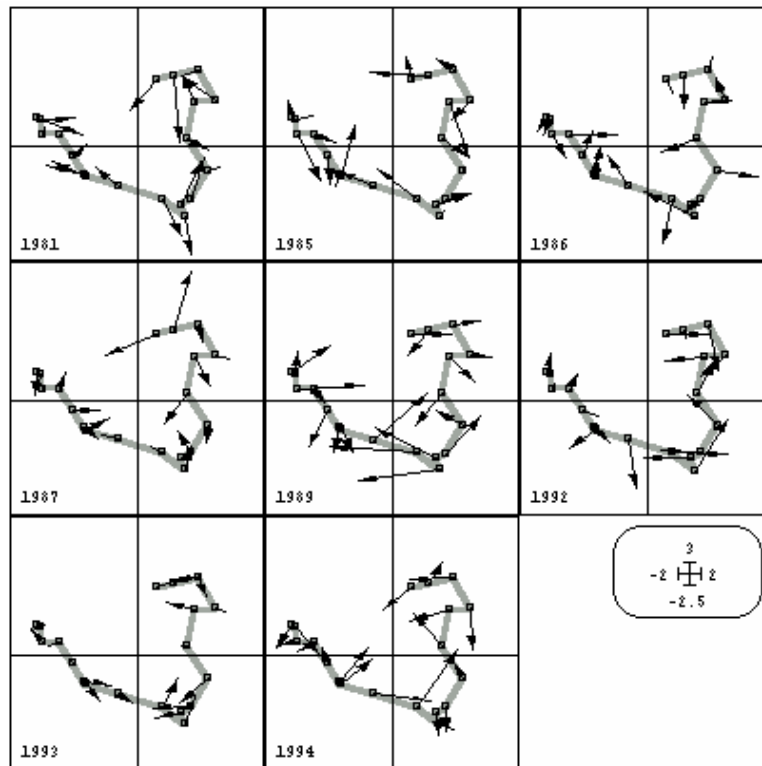
Elle est moins grande quand on peut dire sans hésitations que les données sont formées de n points-individus qui sont vus au moyen de K paquets de p_1, \dots, p_K descripteurs. Deux cas se présentent. Dans le premier, les n points servent à faire une typologie de variables et ceci K fois. Dans le second, les p_1, \dots, p_K descripteurs servent à faire une typologie des individus et ceci dans K conditions. Par exemple, les p_1, \dots, p_K descripteurs sont des espèces rangées par groupes faunistiques. Les stations peuvent être utilisées pour faire simultanément K typologies d'espèces par groupe ou les K groupes peuvent servir à faire K typologies de stations. Ceci est d'autant plus délicat que K groupes de stations (par date, par région, par méthode, ...) peuvent donner des tableaux sur le même groupe faunistique. Ce groupe sert-il à faire K typologies des mêmes stations ou ces stations servent-elles à faire K typologies du même groupe ?

Si *a priori* un objectif ne l'emporte pas sur l'autre, préférer l'AFM. On pourrait dire que c'est la plus généraliste et la moins contraignante.

STATIS répond en priorité à la question de la reproductibilité d'une structure. Si on pense *à priori* « Il n'y a qu'une seule typologie et ce qui m'intéresse c'est moins cette structure que je connais que la qualité de sa reproduction », les méthodes STATIS sont centrales. C'est pourquoi, il ne faut pas nécessairement utiliser un STATIS sur les X parce que les tableaux ont mêmes individus et mêmes variables. Même dans ce cas, si l'important c'est la reproduction de la structure entre les lignes il faut choisir un STATIS sur les opérateurs dont les éléments communs sont les lignes et si l'important c'est la reproduction de la structure entre les colonnes il faut choisir un STATIS sur les opérateurs dont les éléments communs sont les colonnes. *Les repères sur la forme des données qui permettent de rentrer dans un programme sont à proscrire.*

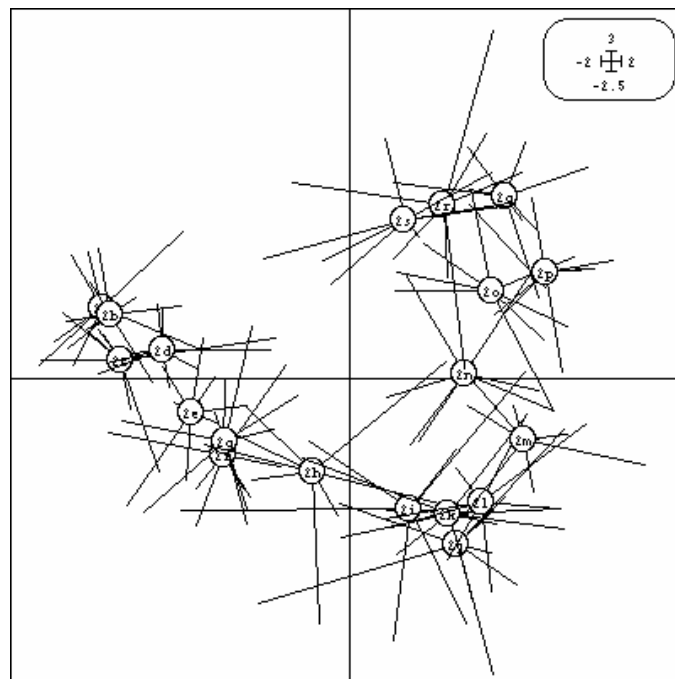
L'AFM est préférable si les individus sont des échantillons (on pourrait en prendre d'autres) qui servent à discuter des variables. Ce n'est pas par hasard si les exemples de tous les articles qui illustrent la méthode (carte Afmult) portent sur des échantillons de dégustation. Si on pense *à priori* « Il n'y a une typologie moyenne et ce qui m'intéresse c'est moins cette structure que ce qu'elle induit sur les variables », l'AFM est centrale. C'est la seule qui possède une double capacité de représenter les paquets de variables par tableau et les points par leur position sur chacun des tableaux de manière optimum. A préférer si les n points servent à faire une typologie de variables dans K groupes indépendants.

L'ACOM est optimum si on veut examiner les déformations d'une typologie de points faites en plusieurs occasions. Elle prend chaque tableau comme un nuage de n points de \mathbb{R}^{p_k} et le projette sur des axes comme dans une analyse simple. Mais les systèmes d'axes dans les K espaces sont choisis pour optimiser la covariance des coordonnées avec des scores de synthèse (un nuage de référence construit dimension par dimension). Le premier facteur de l'ACOM est celui de l'AFM mais les suivants sont différents. Une des figures importantes de l'ACOM est du type :



Logique d'analyse de co-inertie multiple

Chaque tableau essaie de représenter un même nuage de points (données de la carte Fatale). Les points origines des flèches sont les mêmes dans chaque fenêtre. Les extrémité des flèches sont les positions réelles pour le tableau correspondant. Comparer avec la même information :



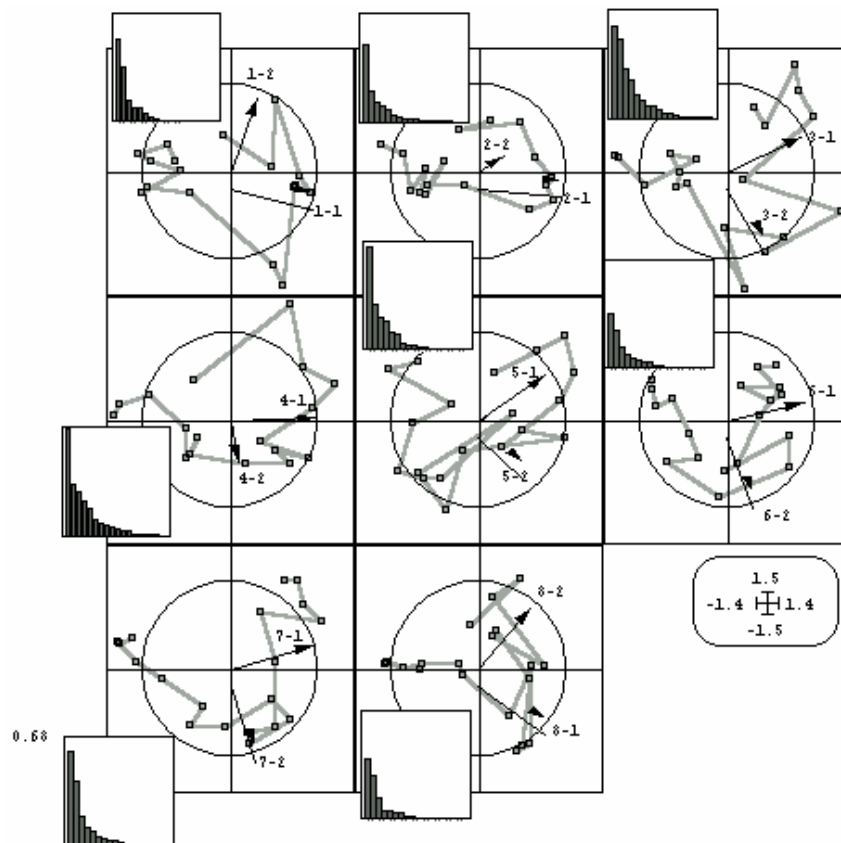
Logique d'analyse factorielle multiple

Le principe de l'ACOM tient dans l'optimisation de

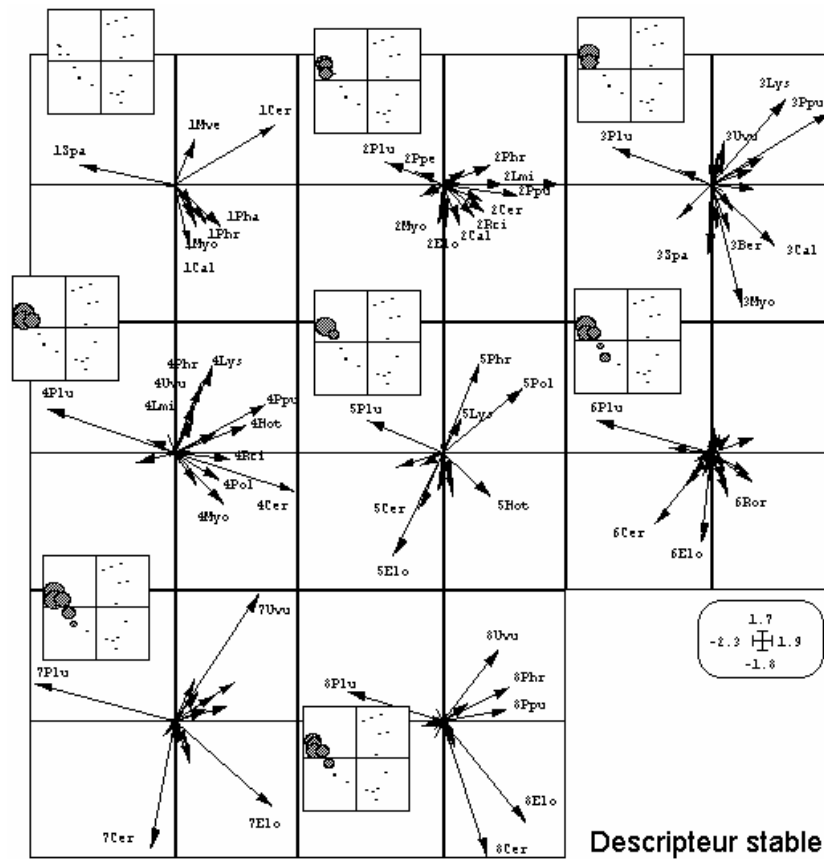
$$f(\mathbf{u}_1^1, \mathbf{u}_2^1, \dots, \mathbf{u}_K^1, \mathbf{v}_1) = \sum_{j=1}^K q_j \langle \mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^1 | \mathbf{v}_1 \rangle_{\mathbf{D}}^2$$

On trouve un axe principal \mathbf{Q}_j -normé \mathbf{u}_j^1 dans l'espace \mathbb{R}^{p_j} , on projette dessus les n points du nuage de lignes associés au tableau \mathbf{X}_j , ce qui donne un système de coordonnées $\mathbf{L}_j = \mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^1$. On trouve simultanément un score normé des individus \mathbf{v}_1 qui donne avec le système de coordonnées une covariance $\langle \mathbf{L}_j | \mathbf{v}_1 \rangle_{\mathbf{D}}$ et ce qui est maximisé est la somme des carrés des covariances des systèmes de coordonnées avec le score de synthèse. Cette somme est pondérée par q_j pour égaliser le rôle des tableaux, par exemple par l'inverse de la première valeur propre de l'analyse simple comme en AFM. Ce qui sépare ensuite l'ACOM de l'AFM est qu'on recommence en prenant comme contrainte d'orthogonalité $\langle \mathbf{u}_j^1 | \mathbf{u}_j^2 \rangle_{\mathbf{Q}_j} = 0$ dans chacun des espaces et $\langle \mathbf{v}_2 | \mathbf{v}_2 \rangle_{\mathbf{D}}$ (non corrélation entre les scores de synthèse).

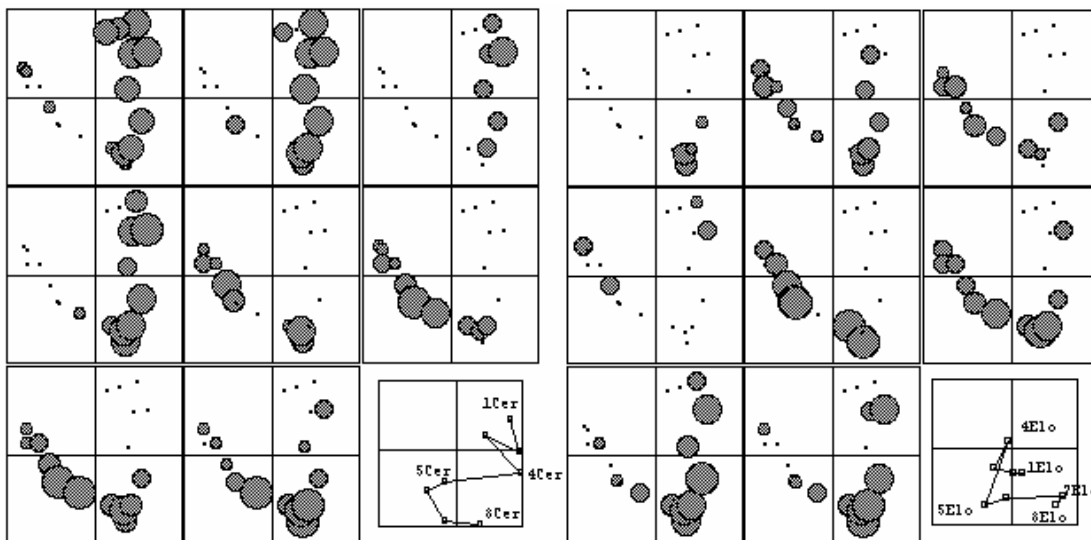
L'ACOM, comme la co-inertie, permet donc de représenter chacune des structures dans son propre espace par projection sur une base d'axes de co-inertie, donc de faire K analyses coordonnées :



Ce sont ces coordonnées normalisées qui sont utilisées avec les scores de synthèse pour donner la figure qui précède. Les variables sont, elles, projetées sur le plan des variables de synthèse ($\mathbf{X}'_j \mathbf{D} \mathbf{v}_1, \mathbf{X}'_j \mathbf{D} \mathbf{v}_2, \dots$). Ici les variables sont les mêmes d'un tableau à l'autre, mais il n'y a eu contrainte de cohérence que sur les individus. Certaines espèces sont des descripteurs stables :



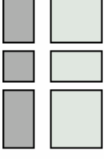

Certaines espèces au contraire ont une coordonnée instable ce qui rend compte de l'instabilité de leur dispersion :



Descripteurs instables

On voit donc qu'un fait majeur des données aurait été perdu par une contrainte d'appariement double : la structure spatiale est stable et s'exprime de manière continue, la légende de cette structure spatiale est instable dans l'ensemble des descripteurs. On sait donc caractériser une typologie stable à définition variable (A est toujours différent de B mais parfois pour une raison, parfois pour une autre). *Ce qu'on ne sait pas encore faire c'est une typologie de typologie.*

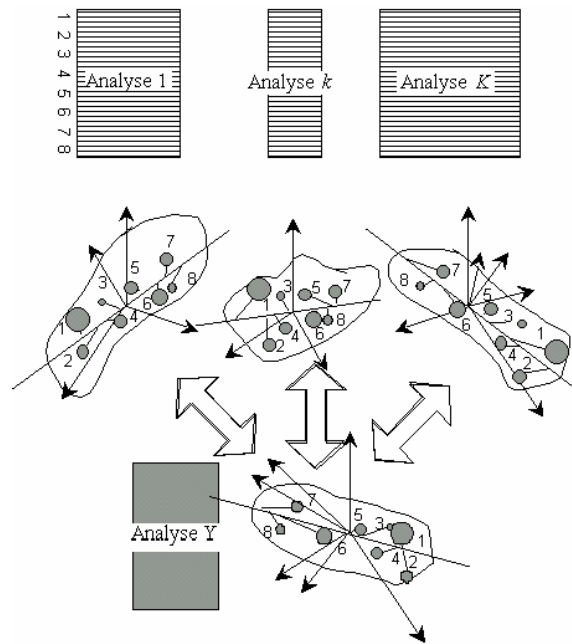
4.5. Extensions

	STATIS	AFMULT	ACOM
Carte de synthèse des lignes	++	+	+
Reproductibilité de la carte de synthèse	-	+ logique inter	++ logique intra
Capacités typologiques	+ logique opérateurs	+ logique variables	+ logique individus
Reconstitution simultanée	non	non	non
Typologie de structures	perspectives	non	non
Extensions mêmes variables	oui	possible	oui
Versions AFC	oui	oui	oui
Simplicité	-	++	-
Versions naïves	oui poids uniformes	oui ACP conjointe	non
Extension 	oui STATICO	non	non
Extension 	non	non	oui

Le tableau contient des points de comparaison entre méthodes.

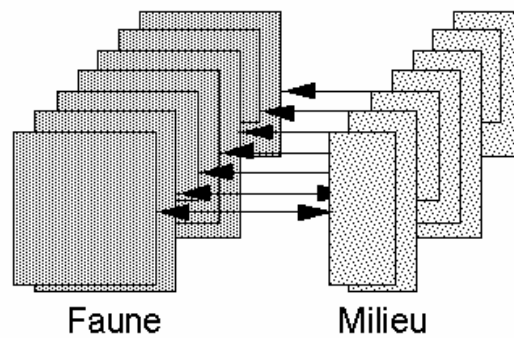
Pour approfondir la partie technique voir ²⁹, pour une autre méthode de cette famille ³⁰, pour une revue de méthodes ³¹, pour l'équivalent en statistique gaussienne ³², pour l'étude de K matrices de distances ³³.

On peut enfin citer deux développements récents. Le premier permet l'analyse de K tableaux conjointement avec un tableau de référence (typiquement K groupes faunistiques et un tableau de milieu sur les mêmes stations). Il est proposé dans ³⁴ (disponible dans K+1: Concordance analysis) et s'appuie sur le principe :

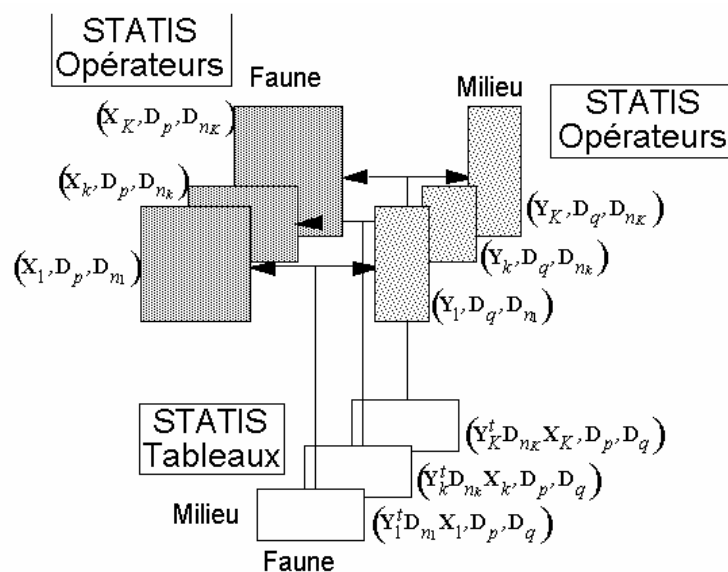


C'est une analyse de co-inertie multiple sur des tableaux croisés.

Le second est proposé par ³⁵ et permet d'aborder les données du type :



C'est un STATIS sur les tableaux croisés.



On l'utilise avec STATIS: Table averaging avec la fiche Thema58.

5. Références

- 1 Friday, L.E. (1987) The diversity of macroinvertebrate and macrophyte communities in ponds. *Freshwater Biology* : 18, 87-104.
- 2 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.
- 3 Blondel, J. & Farre, H. (1988) The convergent trajectories of bird communities along ecological successions in european forests. *Öcologia* (Berlin) : 75, 83-93.
- 4 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.
- 5 Hill, M.O. & Smith, A.J.E. (1976) Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* : 25, 249-255.
- 6 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.
- 7 Kettnering, R.J. (1971) Canonical analysis of several sets of variables. *Biometrika* : 58, 433-451
- 8 Vinograd, B. (1950) Canonical positive definite matrices under internal linear transformations. *Proceedings of the American Mathematical Society* : 1, 159-161.
- 9 Horst, P. (1961) Relations among m sets of variables. *Psychometrika* : 26, 129-149.
- 10 Casin, Ph. & Turlot, J.C. (1986) Une présentation de l'analyse canonique généralisée dans l'espace des individus. *Revue de Statistique Appliquée* : XXXV, 3, 65-75.
- 11 Saporta, G. (1975) *Liaisons entre plusieurs ensembles de variables et codage de données qualitatives*. Thèse de 3^e cycle, Université Pierre et Marie Curie, Paris VI. 1-102.
- 12 Escofier, B. & Pagès, J. (1990) *Analyses factorielles simples et multiples : objectifs, méthodes et interprétation*. Dunod, Paris. 1-267.
Escofier, B. & Pagès, J. (1994) Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis* : 18, 121-140.
- 13 Guinand, B. (1996) Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biological Journal of the Linnean Society* : 58, 173-195.
- 14 Foucart, T. (1978) Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des données* : 2, 67-84.
- 15 Bouroche, J.M. (1975) *Analyse des données ternaires: la double analyse en composantes principales*. Thèse de 3^e cycle, Université de Paris VI. 1-57 + annexes.
- 16 Thioulouse, J. & Chessel, D. (1987) Les analyses multi-tableaux en écologie factorielle. I De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. *Acta Œcologica, Œcologia Generalis* : 8, 4, 463-480.
- 17 Kroonenberg, P.M. (1989) The analysis of multiple tables in factorial ecology. III Three-mode principal component analysis: "analyse triadique complète". *Acta Œcologica, Œcologia Generalis* : 10, 3, 245-256.
- 18 Dolédec, S. (1988) Les analyses multi-tableaux en écologie factorielle. II Stratification longitudinale de l'Ardèche à partir des descripteurs physico-chimiques. *Acta Œcologica, Œcologia Generalis* : 9, 119-135.

- 19 Blanc, L., Chessel, D. & Dolédec, S. (1998) Etude de la stabilité temporelle des structures spatiales par Analyse d'une série de tableaux faunistiques totalement appariés. *Bulletin Français de la Pêche et de la Pisciculture* : 348, 1-21.
- 20 Kroonenberg, P.M. (1983) *Three-mode principal component analysis*. DSWO Press, Leiden. 1-380.
- 21 Franc, A. (1992) *Etude algébrique des multitableaux : apports de l'algèbre tensorielle*. Thèse de doctorat, Université de Montpellier II. 1-361.
- 22 Coppi, R. & Bolasco, S. Eds. (1989) *Multway Data Analysis*. Elsevier Science Publishers B.V., North-Holland. 1-552.
- 23 Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.
- 24 Lavit, Ch. (1988) *Analyse conjointe de tableaux quantitatifs*. Masson, Paris. 1-240. Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis* : 18, 97-119.
- 25 Baran, E. (1995) *Dynamique spatio-temporelle des peuplements de Poissons estuariens en Guinée (Afrique de l'Ouest)*. Thèse de Doctorat, Université de Bretagne Occidentale. 1-204 + annexes.
- 26 Gaertner, J.C., Chessel, D. & Bertrand, J. (1998) Stability of spatial structures of demersal assemblages: a multitable approach. *Aquatic Living Resources* : 11, 75-85.
- 27 Chessel, D. & Hanafi, M. (1996) Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* : 44, 35-60.
- 28 Bornette, G., Amoros, C. & Collilieux, G. (1994) Role of seepage supply on the aquatic vegetation dynamics in former river channels: prediction testing using a hydroelectric construction. *Environmental Management* : 18, 223-234.
- 29 Dazy, F. & Le Barzic, J.F. (1996) *L'analyse des données évolutives*. Technip, Paris. 227 p.
- 30 Casin, Ph. (1996) L'analyse en composantes principales généralisée. *Revue de Statistique Appliquée* : 44, 63-81.
- 31 Coppi, R. (1996) Computer programs in multiway data analysis. In : *SoftStat 95. Advances in statistical software* 5. Faulbaum, F. & Bandilla, W. (Eds.) Lucius & Lucius. 37-46.
- 32 Flury, B.N. (1984) Common principal components in k groups. *Journal of the American Statistical Association* : 79, 892-898.
- 33 Gower, J.C. (1975) Generalized procustes analysis. *Psychometrika* : 40, 33-51.
- 34 Lafosse, R. & Hanafi, M. (1997) Concordance d'un tableau avec K tableaux : définition de K+1 uples synthétiques. *Revue de Statistique Appliquée* : 45, 111-126.
- 35 Simier, M., Blanc, L., Pellegrin, F. & Nandris, D. (1998) Approche simultanée de K couples de tableaux : Application à l'étude des relations pathologie végétale - environnement. *Revue de Statistique Appliquée* : 47, 31-46.