

Analyses de co-inertie multiple : introduction

Résumé

L'Analyse de co-inertie multiple (ACOM, Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* : 44, 35-60) généralise à K tableaux l'analyse de co-inertie (Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294) pour deux tableaux de la même manière que l'analyse canonique généralisée de Carroll (A generalization of canonical analysis to three or more sets of variables. *Proceeding of the 76th Convention of the American Psychological Association* : 3, 227-228) généralise l'analyse canonique classique de Hotelling (Relations between two sets of variates. *Biometrika* : 28, 321-377). La fiche décrit l'utilisation du module KTA qui contient cette méthode.

Plan

1 — L'objectif poursuivi	2
2 — Analyses séparées et conjointe	4
2.1 — Données traitées.....	4
2.2 — Constitution du K-tableaux.....	8
2.2 — Les analyses séparées	9
2.3 — L'analyse conjointe.....	11
3 — Analyses coordonnées	13
3.1 — L'analyse de co-inertie multiple d'ordre 1	13
3.2 — Analyse d'ordre m.....	15
3.3 — Mise en œuvre	16
Références	23

D. Chessel & M. Hanafi

1 — L'objectif poursuivi

L'ACOM est une méthode d'ordination simultanée de K tableaux, comme STATIS, et en ce sens la fiche utilise le contenu des fiches précédente de ce fascicule. On dispose de K tableaux ayant en commun soit les lignes soit les colonnes. Nous supposons d'abord que ces tableaux ont des lignes afférentes à des objets communs. On dispose donc de K triplets statistiques $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$ ($1 \leq k \leq K$).

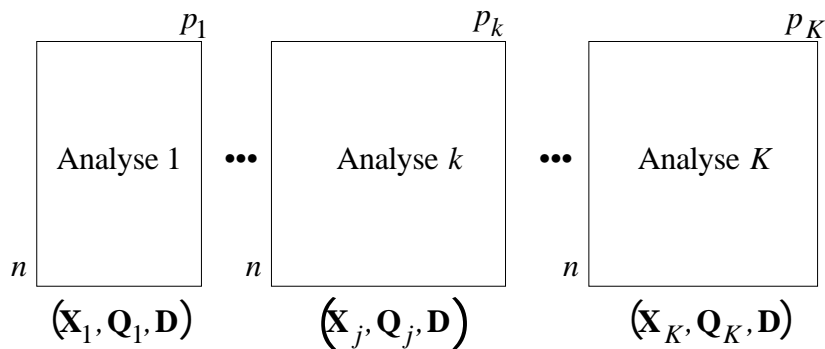


Figure 1 : définition générale d'un K-tableaux

Il apparaît immédiatement que le cas de 2 tableaux est un cas particulier. Les analyses particulières sont des ACP, des AFC, des ACM ou des variantes non classiques. La seule contrainte est qu'elles partagent une pondération des lignes commune consignée dans la matrice diagonale \mathbf{D} . Les notations sont celles du schéma de dualité¹. Nous supposons que l'analyse de co-inertie de deux quelconques des tableaux a un sens. Si K est grand, on sera rapidement débordé par le dépouillement de $K(K - 1)/2$ analyses et l'objectif principal est d'assurer l'analyse conjointe des K triplets statistiques en mettant en évidence la structure commune exprimée partiellement ou totalement par chacun des tableaux.

Pour décrire les méthodes actuellement disponibles Bove et Di Ciacco² distingue les *Three-way data sets* et les *Three-way data arrays*, c'est-à-dire les K -tableaux qui nous intéressent ici et les cubes de données. Pour les premiers ils indiquent que la base est formée de l'analyse canonique des corrélations (*Canonical Analysis*) dont tous les standards logiciels (BMDP, SAS, SPSS-X) donnent une procédure et l'ACP sur variables instrumentales (*Redundancy analysis*) qu'on peut exécuter dans SAS par la commande CANCORR. ADE-4 ignore la première pour l'ensemble des raisons qui ont conduit à privilégier en écologie la seconde. De toutes ces raisons la plus importante est le nombre des variables (espèces) qui figurent dans un tableau floro-faunistique. L'analyse canonique est strictement impossible dans la situation :

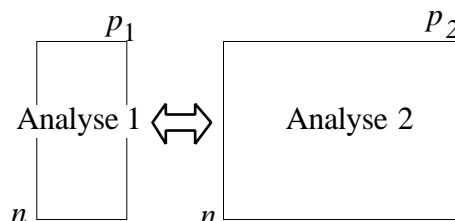


Figure 2 : définition générale d'un couple de tableaux

L'ACPVI, sa variante AFCVI ou ACC (*Canonical Correspondence Analysis*) est adaptée à la situation :

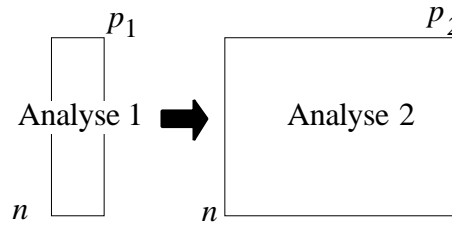


Figure 3 : Couple de tableaux supportant une analyse des redondances

où la flèche désigne l'opération de prédiction (prédire la faune par le milieu) ou l'opération de contrainte (analyser le tableau faunistique sous contrainte du tableau de variables mésologiques). L'analyse des redondances est strictement impossible dans la situation inverse :

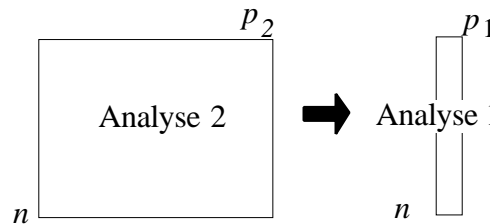


Figure 4 : Couple de tableau supportant une régression PLS

qu'on rencontre en chimiométrie (à gauche, le tableau des descripteurs de structure des molécules, à droite celui de l'activité de ces molécules, la flèche définissant la finalité des QSAR (*Quantitative Structure Activity Relationships*) et qui conduit à la régression PLS. Dans ces situations, les impossibilités énoncées sont mathématiques si le nombre de colonnes dépassent le nombre de lignes et pratiques si le nombre de lignes n'est pas bien plus grand que le nombre de colonnes. On entend par "bien plus grand" ce qui est nécessaire pour pouvoir interpréter les résultats.

L'analyse canonique est donc la solution au problème :

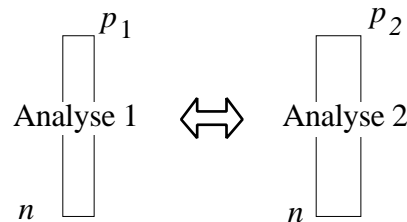
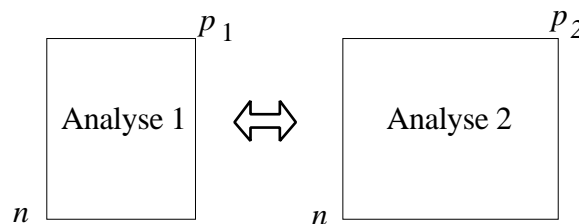
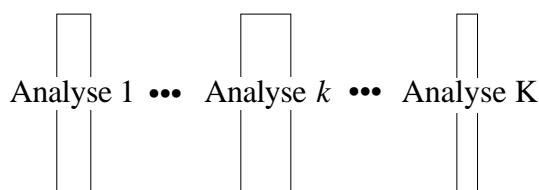


Figure 5 : Couple de tableau supportant une analyse canonique

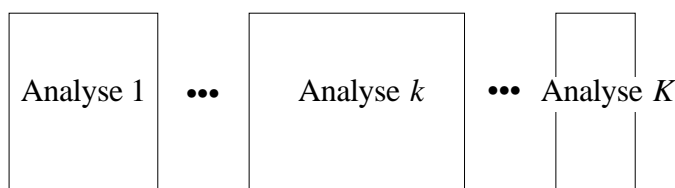
Cette question est tellement rare en écologie que nous l'avons laissée de côté au profit de l'analyse de co-inertie solution, sans contrainte, du problème de l'analyse symétrique :



Quand on passe de deux à K tableaux, on retrouve évidemment les mêmes difficultés. Dans tout ce qui suit, on considère les tableaux à égalité, aucun d'entre eux ne jouant un rôle particulier par rapport aux autres. L'analyse canonique généralisée est la solution du problème :



On cherche donc une solution équivalente adaptée aussi bien à la précédente qu'à la suivante, ou qu'à toute combinaison des deux :



Bove et Di Ciaccio cite dans ce créneau *deux méthodes importantes* respectivement STATIS³ et l'analyse factorielle multiple (AFMULT⁴), lesquelles apparaissent de même dans la revue de Rizzi & Vichi⁵ qui considèrent les K -tableaux issus des cubes de données. Nous avons fait une large part à la première (fascicule 6) et nous allons discuter des relations de nos propositions avec la seconde. L'objectif est d'ordonner simultanément K tableaux.

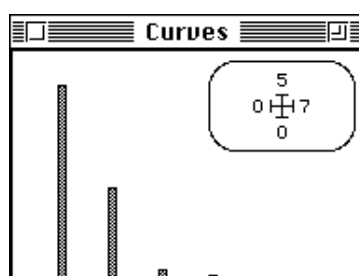
2 — Analyses séparées et conjointe

2.1 — Données traitées

Nous utiliserons d'abord les données formant l'exemple pratique utilisé par Sabatier⁶ pour une synthèse récente. On analyse ensemble 5 tableaux sur 23 districts. Le premier est sur la carte Meteo de ADE•Data.

Data	0721	ATLAS Meteo (23-7) Températures (°C)	Précipitations (mm)
-6.6**1.1**9.7**22.9**131.5**139.5**1596.5			
-7.5**0.1**8.2**20.8**115		-6.6 1.1 9.7 22.9 131.5 139.5 1596.5	
-8.5**0.1**8.6**22**113**		-9999 0.1 8.2 20.8 115 145 1613	
-9.2**1.7**6.5**17.3**10		-8.5 -0.1 8.6 22 113 146.5 1738	
-8.1**0.2**7.5**21.6**110**		-9.2 -1.7 6.5 17.3 103 138 1630	
-7.3**0.5**8.6**21**113**		-8 -9999 7.5 21.6 110 129 1491.5	
-2.9**2.8**11.5;19.8;188			
-8.4;-1;8.3;21.		-6.6 1.1 9.7 22.9 131.5 139.5 1596.5	

Meteo a 23 lignes et colonnes. En faire l'ACP normée. Résumer les résultats (figure 6)



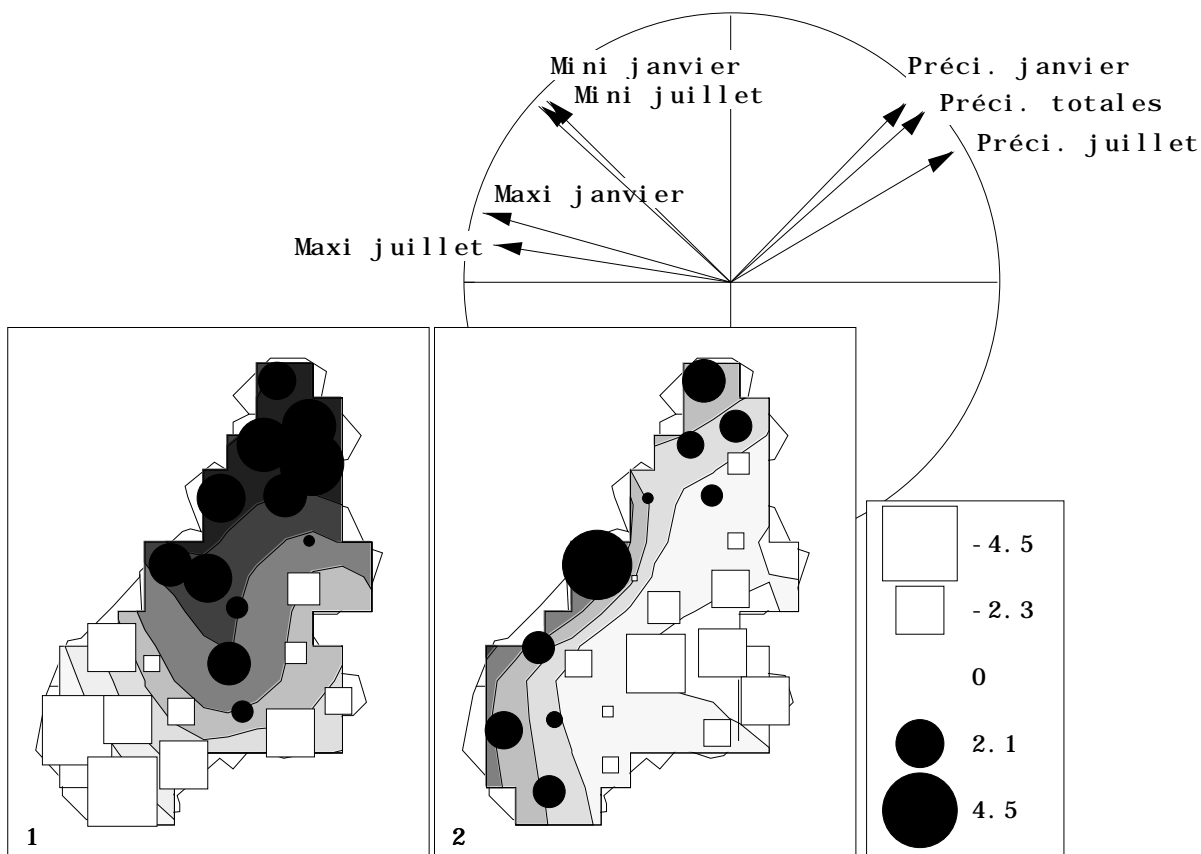


Figure 6: ACP normée de Meteo (premier tableau). Toutes les cartes avec courbes de niveaux seront tracées avec 8 voisins.

Détruire tous les fichiers créés à l'exception du tableau normalisé Meteo.cnta (23-7 normalisé par colonnes) et Meteo.cnpc (7-1 poids des colonnes).

Le second tableau est sur la carte Arbres :

Data	0723	ATLAS Arbres (23-9)
15,4		
201-3	202-2	423-1 119-1
16,6		
202-3	423-1	438-2 216-1 007-3 008-1
17,4		
202-3	423-2	423-3 007-2
18,5		
201-1	202-3	423-2 438-2 007-1
19,6		
201-2	202-2	423-2 438-2 007-1 119-2
20,5		
201-3	202-3	423-1 007-2 119-1
		007-Chêne pubescent (Quercus pubescens Wild.)
		008-Charme (Carpinus betulus L.)

Obtenir un fichier binaire Veg par EcoDataBin (module Basic) :

1	1	2	1	0	0	1	2	2	3	13	2	1	1	1	2	2	2	2	2
2	1	1	1	0	1	1	1	2	3	14	1	0	2	2	1	0	1	1	1
3	1	1	2	1	0	1	1	2	2	15	0	0	1	3	2	0	0	1	0
4	0	0	3	1	1	1	1	2	1	16	3	1	0	0	3	0	1	1	2
5	1	2	0	0	0	1	2	3	3	17	2	0	0	0	3	0	0	3	0
6	1	0	2	1	0	0	1	2	2	18	1	0	0	1	3	0	0	2	2
7	1	1	1	0	1	1	1	3	3	19	1	0	2	2	2	0	0	2	2
8	2	1	1	0	1	0	2	2	2	20	2	0	1	3	3	0	0	1	0
9	1	0	3	2	2	0	2	2	1	21	0	0	2	3	2	0	0	1	0
10	1	0	2	2	2	0	1	2	1	22	3	0	0	0	3	0	0	1	2
11	1	0	2	2	1	0	1	2	1	23	3	0	0	1	3	0	0	0	2
12	1	1	1	0	1	0	1	3	0										

Faire l'ACP centrée de Veg (figure 7).

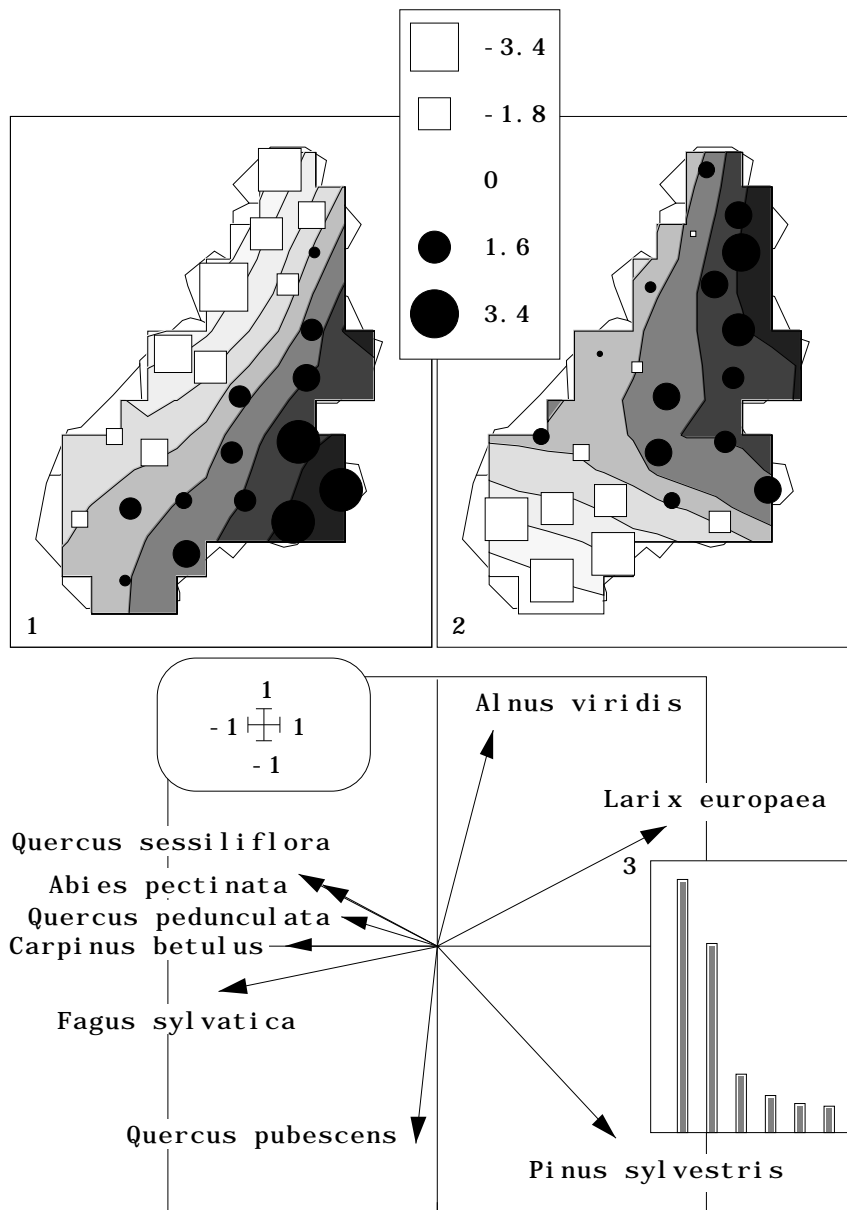


Figure 7 : ACP centrée de Veg (second tableau).

On reconnaît déjà la cohérence partielle des deux analyses et leur originalité. Détruire tous les fichiers à l'exception de Veg.cpta (23-9, centré par colonnes) et Veg.cppc (9-1 poids des colonnes). Pour illustrer les capacités de gestion des modules d'ADE-4, nous allons ajouter aux deux premiers tableaux 3 tableaux d'une analyse des correspondances floues (carte Oiseaux de la pile ADE•Data) :

Data	0727	Tableau faunistique multi-groupes (23-8+6+5)	
0221100002012020020	↑	1-Fauvette orphée <i>Sylvia hortensis</i> (Gm.)	8
0221100002112120120	□	2-Fauvette des jardins <i>Sylvia borin</i> (Boddar)	6
0222100002113120120	□	3-Fauvette à tête noire <i>Sylvia atricapilla</i>	5
0321000001112110121	□	4-Fauvette babillarde <i>Sylvia curruca</i> (L.)	
0321200001112111121	□	5-Fauvette grisette <i>Sylvia communis</i> (Lath.)	

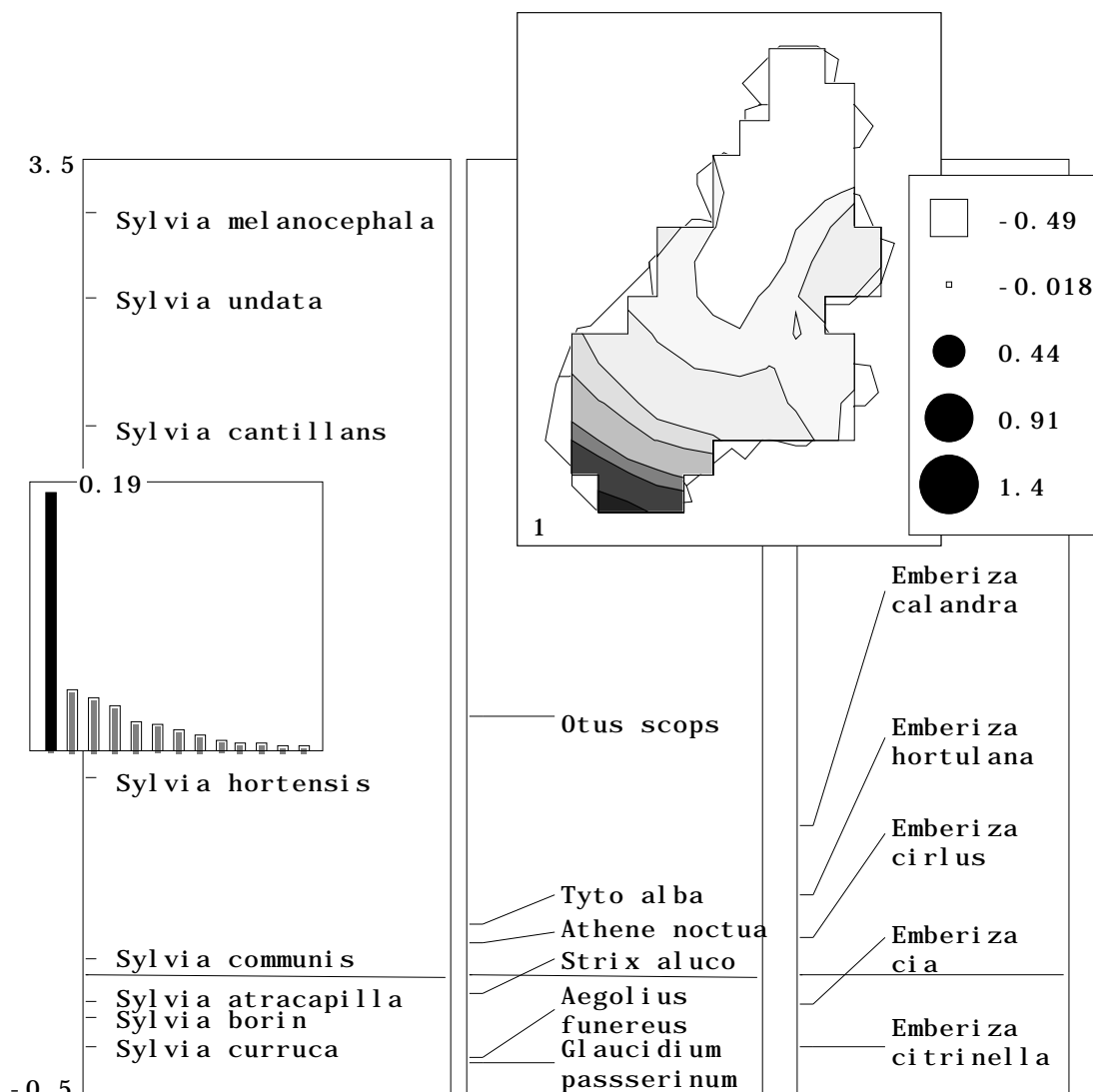


Figure 8 : ACM flou du tableau Avi (Tableaux 3 à 5).

Récupérer le fichier binaire Avi (23 districts et 19 espèces) et le fichier AviAtlas.num contenant sur une seule colonne le nombre d'espèces par blocs. Lire Avi par FuzzyVar :

Read Fuzzy File			
Fuzzy variables: input file (---)	<input type="button" value="..."/>	Avi	23 19
Category indication file	<input type="button" value="..."/>	AviAtlas.num	3 1
Output file name (default = ---F)	<input type="button" value="..."/>		

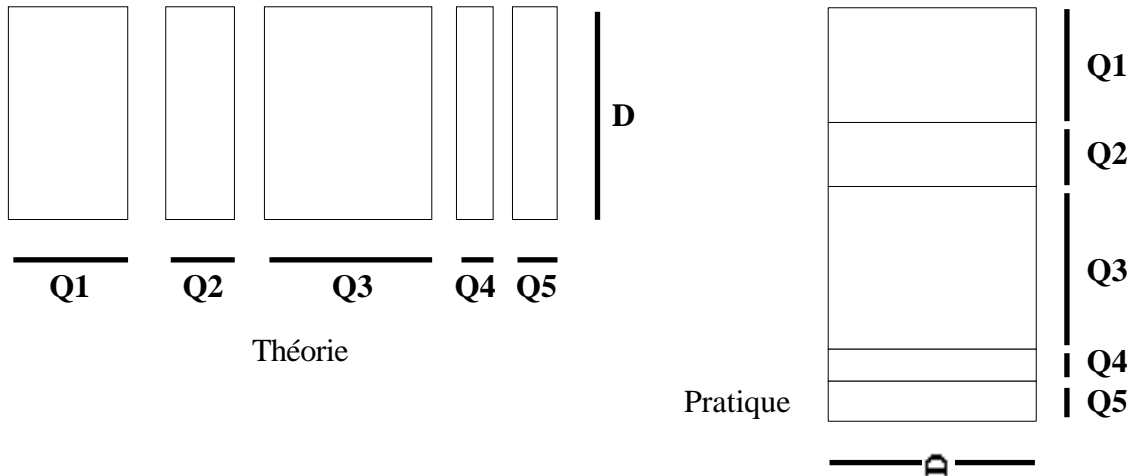
Faire l'AFC floue :

Fuzzy Correspondence Analysis	
.fuz type file	<input type="button" value="..."/> AviF.fuz
Option: Row weighting file	<input type="button" value="..."/>

Un premier facteur très important cache peut-être des nuances attachées à l'un ou l'autre des sous-tableaux (interprétation dans la figure 8).

2.2 — Constitution du K-tableaux

Il est possible de constituer entièrement à la main un K-tableaux. Il suffit de connaître les contraintes d'entrée dans le module KTA. Nous avons déjà vu que l'association des tableaux se fait en pile, alors que le schéma théorique les associe côte à côte :



Pour vouloir mélanger ACP normée (tableau 1), ACP centrée (tableau 2) et AFC (tableaux 3, 4 et 5) nous devons constituer les fichiers de la partie droite du graphe ci-dessus, à savoir un tableau superposant les 5 parties et deux fichiers de poids. Dans le tableau, nous mettrons l'un au dessus de l'autre le transposé de *Meteo.cnta* (7-23), le transposé de *Veg.cpta* (9-23) et le transposé de *Avi.flta* (19-23) ce dernier contenant 3 tableaux à centrage AFC (8-23, 6-23 et 5-23). Transposer les éléments et associer dans un fichier *F0* (35-23) par le module *FilesUtil*.

La pondération *D* des lignes des tableaux initiaux est uniforme : la pondération des colonnes du K-tableaux sera donc uniforme et sera créée par le module *KTabUtil* (c'est une option standard). Reste donc à créer la pondération des lignes du K-tableaux qui doit impérativement être formée de l'empilement de vraies pondération (somme des valeurs égale à 1 par bloc). Ce n'est le cas d'aucune des pondérations de colonnes initiales. En effet dans une ACP standard les poids des colonnes sont unitaire et dans une ACM floue les poids des colonnes sont de somme unité sur toutes les modalités de toutes les variables. Mais le module *KTabUtil* refera cette opération de recalibrage des pondération par blocs. Superposer donc les trois fichiers de pondération *Meteo.cnpc*, *Veg.cppc* et *Avi.flpc* dans un fichier *Fpoili* (35-1) :

PasteFiles-SameCol			
Input BIN file 1		Meteo.cnpc	7 1
Input BIN file 2 (optional)		Veg.cppc	9 1
Input BIN file 3 (optional)		Avi.flpc	19 1
Output file		Fpoili	
<input type="button" value="Quit"/>		<input type="button" value="Ok"/>	

Il suffit alors de créer le fichier binaire de constitution des blocs de lignes *Fbloc* (5-1) qui contient les valeurs 7, 9, 8, 6 et 5, nombres d'éléments associés à chaque bloc. Lancer le module *KTabUtil* :

puis :

InitKTab			
Matrix input file		F0	35 23
Row indicator (Default = 1 class)		Fbloc	5 1
Col indicator (Default = 1 class)			
Option: output file name		K	

CentringKtab			
---.ktp input file		K!.ktp	
Matrix input file		F0	35 23
Row weighting (default 1/nk)		Fpoili	35 1
Column weighting (default 1/pj)			
Centring option (no default)		5	
Option: output file name		A	

```
Option: InitKTab
Input file: F0
Row indicator file: Fbloc
Col indicator file:
Output ASCII file: K!.ktp
-----
Qualitative variables file: K!TLI
Number of rows: 35, variables: 3, categories: 49
Auxiliary ASCII output file K!TLI.label: labelling file
-----
Qualitative variables file: K!TCc
Number of rows: 115, variables: 3, categories: 51
Auxiliary ASCII output file K!TCc.label: labelling file
-----
Qualitative variables file: K!T4a
Number of rows: 20, variables: 2, categories: 9
Auxiliary ASCII output file K!T4a.label: labelling file
-----
```

Voir le signification du contenu de ces fichiers dans la fiche 1 du fascicule 6 page 22 et 23.

```
Option CentringKtab
Input file F0
Number of rows: 35, columns: 23
Row weights (Sum inside block = 1) are selected from file
Fpoili
File A.ktpl contains the row weights
It has 35 rows and 1 column
Each column has uniform weight (Sum inside block = 1)
File A.ktpc contains the column weights
It has 23 rows and 1 column
File A.ktta contains the non modified table
It has 35 rows and 23 columns
```

2.2 — Les analyses séparées

Pour refaire les 5 analyses séparées, utiliser KTA :

```

----- Separate analyses -----
---.kttta input file      [A.kttta] 35 23

```

```

Row bloc: 1 - Col bloc: 1 - Total inertia:      1
Num. Eigenval.  R. Iner.  R. Sum  | Num. Eigenval.  R. Iner.  R. Sum |
01  +6.0950E-01 +0.6095 +0.6095  | 02  +2.9334E-01 +0.2933 +0.9028 |
...

```

```

-----
Row bloc: 2 - Col bloc: 1 - Total inertia: 0.747322
Num. Eigenval.  R. Iner.  R. Sum  | Num. Eigenval.  R. Iner.  R. Sum |
01  +3.1281E-01 +0.4186 +0.4186  | 02  +2.3209E-01 +0.3106 +0.7291 |
...

```

```

-----
Row bloc: 3 - Col bloc: 1 - Total inertia: 0.541902
Num. Eigenval.  R. Iner.  R. Sum  | Num. Eigenval.  R. Iner.  R. Sum |
01  +3.4072E-01 +0.6287 +0.6287  | 02  +7.2803E-02 +0.1343 +0.7631 |
...

```

```

-----
Row bloc: 4 - Col bloc: 1 - Total inertia: 0.350237
Num. Eigenval.  R. Iner.  R. Sum  | Num. Eigenval.  R. Iner.  R. Sum |
01  +1.7028E-01 +0.4862 +0.4862  | 02  +9.6333E-02 +0.2751 +0.7612 |
...

```

```

-----
Row bloc: 5 - Col bloc: 1 - Total inertia: 0.247036
Num. Eigenval.  R. Iner.  R. Sum  | Num. Eigenval.  R. Iner.  R. Sum |
01  +1.4786E-01 +0.5986 +0.5986  | 02  +5.1476E-02 +0.2084 +0.8069 |
...

```

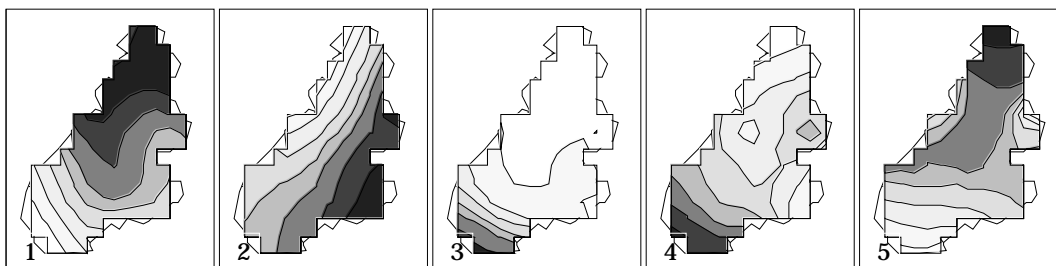
Eigenvalues are stored by column into file A.savp
Row : 9 - Col (tables): 5

File A.saTlli contains row scores in separate analyses
It has 35 rows and 4 columns
It is to be used with --Tll.label and --Tll.cat files

File A.saTCco contains column scores in separate analyses
It has 115 rows and 4 columns
It is to be used with --TCc.label and --TCc.cat files

File A.sapa contains total inertia and rank for each separate analyses
It has 5 rows and 2 columns

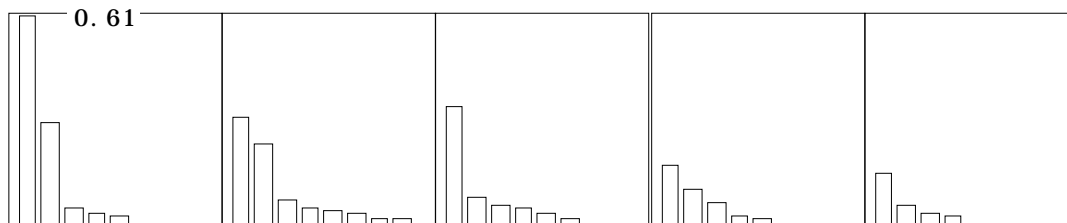
Sans aucune coordination d'aucune sorte, nous avons cinq jeux de coordonnées factorielles. La première colonne de A.saco contient superposée les premières coordonnées (115 = 23*5) des lignes des 5 analyses de départ. Extraire la colonne (fichier Auxil 115-1), changer les dimensions (fichier Auxil 5-23), transposer (fichier Auxil 23-5) et cartographier :



La première carte est celle de la figure 6 (ACP normée de Meteo), la seconde celle de la figure 7 (ACP centrée de Veg), les trois autres celles des coordonnées de rang 1 de

trois AFC sur des tableaux faunistiques (relevés en lignes, colonnes en espèces, passage en pourcentage par relevés pour obtenir la pondération uniforme et la cohérence avec les deux analyses précédentes).

Les valeurs propres des cinq analyses sont :



Se pose ici la question de l'hétérogénéité de l'inertie totale des tableaux associés (1, 0.75, .54, .35, .25). Noter que la pondération uniforme ($1/p$) des colonnes de l'ACP normée passe l'inertie de 7 (nombre de variables) à 1 (sans changer les coordonnées des relevés). Ici, ces inerties sont d'un ordre de grandeur comparable, mais pourrait dans d'autres cas varier dans la proportion, par exemple, de 1 à 1000. Se pose par conséquent la question de l'hétérogénéité des premières valeurs propres, indicateurs numériques de l'importance de la structure de chaque tableau (0.61, 0.31, 0.34, 0.17, 0.15). Ici elles sont du même ordre de grandeur mais pourrait varier fortement d'un tableau à l'autre.

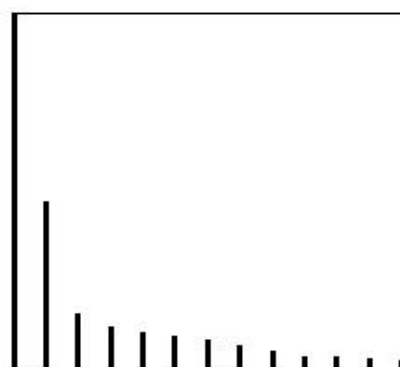
2.3 — L'analyse conjointe

Le K -tableaux pondéré définit, sans plus de difficultés, un triplet statistique ordinaire. Il est donc naturel d'en faire l'analyse. C'est Dagnélie⁷ qui eut le premier l'idée d'associer les espèces et les variables mésologiques dans un même tableau pour en faire l'analyse conjointe. Dans DDUtil :



On obtient :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: A.ktta
--- Number of rows: 35, columns: 23
-----
Total inertia: 2.8865
```



Number of axes ? <

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+1.2386E+00	+0.4291	+0.4291	02	+5.8030E-01	+0.2010	+0.6301
03	+1.9426E-01	+0.0673	+0.6974	04	+1.4877E-01	+0.0515	+0.7490
...							

File A.ktvp contains the eigenvalues and relative inertia for each axis

--- It has 23 rows and 2 columns

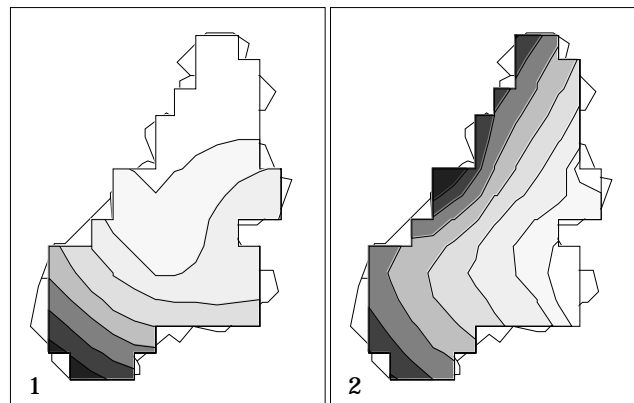
File A.ktco contains the column scores

--- It has 23 rows and 2 columns

File A.ktli contains the row scores

--- It has 35 rows and 2 columns

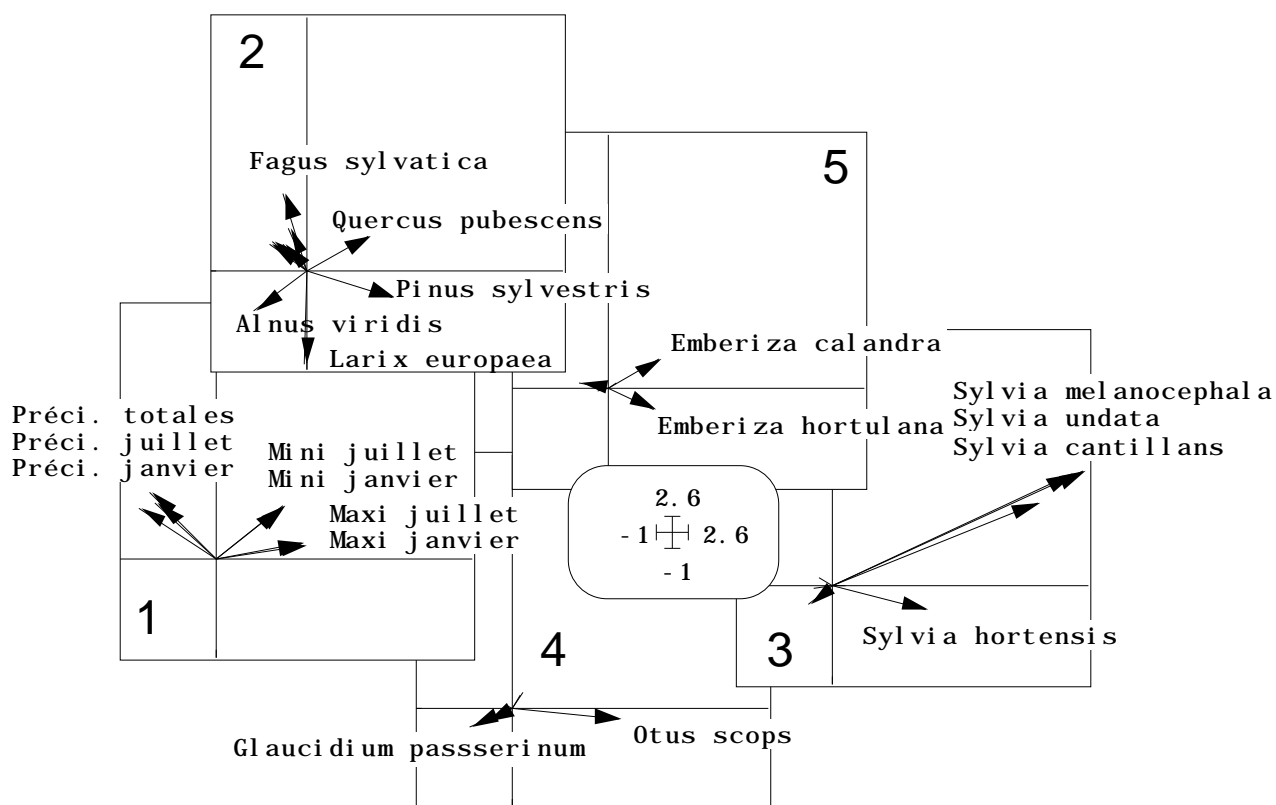
L'inertie totale est la somme des inerties séparées. Il y a deux facteurs indiscutables, ou l'on retrouve les gradients Nord-Sud (influence méditerranéenne) et Est-Ouest (influence océanique ou rôle de l'altitude) :



A cette carte unique des lignes correspond cinq cartes des colonnes des tableaux initiaux :

Labels	
XY coordinates file	<input type="button" value="👉"/> A.ktli 35 2
X-axis column number (default = 1)	<input type="button" value="👉"/> <input type="text"/>
Y-axis column number (default = 2)	<input type="button" value="👉"/> <input type="text"/>
Label file (or # for item numbers)	<input type="button" value="👉"/> Code_Var_5Tab
Draw vectors from origin (yes = 1)	<input type="button" value="👉"/> 1

Row selection file (.cat):	<input type="button" value="👉"/> K!TLL.cat
Selection col. number:	<input type="text" value="1"/>



A une correction (ici mineure) près l'AFMULT est simplement cette analyse conjointe. On aurait exactement l'AFMULT en divisant chacun des tableaux par la racine de sa première valeur propre pour uniformiser le rôle de chacun d'entre eux. On a une vision d'ensemble des variables de tous les tableaux au prix d'une perte de ce qui se passe dans chacun d'entre eux. L'approche par STATIS des tableaux 3, 4 et 5 est manifestement plus riche d'interprétation (Fiche 2 du fascicule 5, p. 26 et suivantes). Dans l'AFMULT l'appartenance d'une variable à un groupe, sans être sans effet dans la correction initiale, ne joue cependant pas un rôle explicite. Nous voulons décrire la structure de plusieurs tableaux en les comparant, donc coordonner l'analyse de plusieurs triplets statistiques. STATIS propose une solution possédant cependant un défaut non négligeable, par l'absence de représentation convaincante des objets communs pour chaque tableau élémentaire.

3 — Analyses coordonnées

L'Analyse de co-inertie multiple⁸ (ACOM) généralise à K tableaux l'analyse de co-inertie⁹ pour deux tableaux de la même manière que l'analyse canonique généralisée de Carrol¹⁰ généralise l'analyse canonique classique de Hotteling¹¹.

3.1 — L'analyse de co-inertie multiple d'ordre 1

Soit $(\mathbf{X}_j, \mathbf{Q}_j, \mathbf{D})$ le triplet associé au tableau j . Faire l'analyse d'ordre 1 de ce tableau, c'est chercher un vecteur \mathbf{u}_j^1 , normé de \mathbb{R}^{p_j} pour le produit scalaire \mathbf{Q}_j , qui maximise la variance projetée soit $\|\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^1\|_{\mathbf{D}}^2$. C'est en même temps chercher un vecteur \mathbf{v}_j^1 , normé de \mathbb{R}^n pour le produit scalaire \mathbf{D} , qui maximise l'inertie projetée soit

$\left| \mathbf{X}_j^t \mathbf{D} \mathbf{v}_j^1 \right|_{\mathbf{Q}_j}^2$. Les analyses séparées ne s'occupant pas de coordination, chacune d'entre elle obtient sa propre solution par la diagonalisation de la matrice à p_j lignes et p_j colonnes $\mathbf{X}_j^t \mathbf{D} \mathbf{X}_j \mathbf{Q}_j$ ou la matrice à n lignes et n colonnes $\mathbf{X}_j \mathbf{Q}_j \mathbf{X}_j^t \mathbf{D}$. STATIS qui s'occupe essentiellement de structure moyenne cherche une moyenne de ces matrices $\mathbf{X}_j \mathbf{Q}_j \mathbf{X}_j^t \mathbf{D}$ qui sont de mêmes dimensions et la diagonalise pour faire une analyse fictive (compromis).

L'ACG, analyse canonique généralisée de Carrol (op.cit.) étudiée et étendue par Kettinger¹², qui ne s'occupe absolument pas de structures de tableaux, mais seulement de corrélations entre paquet de variables, cherche aussi des vecteurs \mathbf{u}_j^1 , normé de \mathbb{R}^{p_j} pour le produit scalaire \mathbf{Q}_j , mais un seul vecteur \mathbf{v}^1 , normé de \mathbb{R}^n pour le produit scalaire \mathbf{D} , pour maximiser :

$$f(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K, \mathbf{v}) = \prod_{j=1}^K r^2(\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j, \mathbf{v})$$

où $r(.,.)$ désigne la corrélation linéaire entre les variables entre parenthèses.

L'analyse canonique des corrélations (ACC) qui est un cas particulier à deux tableaux, cherche donc deux vecteurs d'un côté et un troisième de l'autre pour maximiser :

$$f(\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}) = r^2(\mathbf{X}_1 \mathbf{Q}_1 \mathbf{u}_1, \mathbf{v}) + r^2(\mathbf{X}_2 \mathbf{Q}_2 \mathbf{u}_2, \mathbf{v})$$

ce qu'on obtient en optimisant simplement¹³ :

$$f(\mathbf{u}_1, \mathbf{u}_2) = r^2(\mathbf{X}_1 \mathbf{Q}_1 \mathbf{u}_1, \mathbf{X}_2 \mathbf{Q}_2 \mathbf{u}_2)$$

L'analyse de co-inertie, approche géométrique générale de l'analyse inter-batterie de Tucker¹⁴, qui s'occupe de co-structure, c'est-à-dire de structure coordonnée de tableaux, cherche encore deux vecteurs pour optimiser :

$$g(\mathbf{u}_1, \mathbf{u}_2) = (\mathbf{X}_1 \mathbf{Q}_1 \mathbf{u}_1 | \mathbf{X}_2 \mathbf{Q}_2 \mathbf{u}_2)_{\mathbf{D}}^2$$

où le produit scalaire définit la covariance entre les deux systèmes de coordonnées des projections. Ceci ne donne pas la solution du problème des trois vecteurs associés, maximiser :

$$g(\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}) = (\mathbf{X}_1 \mathbf{Q}_1 \mathbf{u}_1 | \mathbf{v})_{\mathbf{D}}^2 + (\mathbf{X}_2 \mathbf{Q}_2 \mathbf{u}_2 | \mathbf{v})_{\mathbf{D}}^2$$

On en tire cependant l'idée d'un critère de co-inertie généralisée sous la forme de la fonction :

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

Dans cette approche, cependant l'inertie d'un tableau pouvant l'emporter sur les autres, nous introduisons une pondération des analyses (q_j est le poids du triplet j) en utilisant le critère d'optimisation de :

$$f_q(\mathbf{u}_1^1, \mathbf{u}_2^1, \dots, \mathbf{u}_K^1, \mathbf{v}_1) = \sum_{j=1}^K q_j \left(\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^1 | \mathbf{v}_1 \right)_{\mathbf{D}}^2 \quad (1)$$

L'analyse de co-inertie multiple (ACOM) de rang 1 de K triplets statistiques sera donc la recherche de k vecteurs \mathbf{u}_j^1 , normé de \mathbb{R}^{p_j} pour le produit scalaire \mathbf{Q}_j et d'un vecteur \mathbf{v}_1 , normé de \mathbb{R}^n pour le produit scalaire \mathbf{D} , qui maximise la fonction (1). On démontre (op. cit.) que la solution est la première coordonnée normée de l'AFMULT du K -tableaux pour les lignes et la première coordonnée de la même analyse *normalisée par blocs* pour les colonnes. Ceci n'est vrai que pour le premier axe et là s'arrête la relation entre l'ACOM et l'AFMULT.

3.2 — Analyse d'ordre m

Faire l'analyse d'ordre 2 du tableau j initial, c'est après avoir trouver le vecteur \mathbf{u}_j^1 chercher un vecteur \mathbf{u}_j^2 , normé de \mathbb{R}^{p_j} pour le produit scalaire \mathbf{Q}_j et orthogonal à \mathbf{u}_j^1 pour ce produit scalaire ($(\mathbf{u}_j^1 | \mathbf{u}_j^2)_{\mathbf{Q}_j} = 0$), qui maximise la nouvelle variance projetée soit $\|\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^2\|_{\mathbf{D}}^2$. C'est en même temps chercher un vecteur \mathbf{v}_j^2 , normé de \mathbb{R}^n pour le produit scalaire \mathbf{D} et orthogonal à \mathbf{v}_j^1 pour ce produit scalaire, qui maximise l'inertie projetée soit $\|\mathbf{X}_j^t \mathbf{D} \mathbf{v}_j^2\|_{\mathbf{Q}_j}^2$. L'orthogonalité dans les deux espaces est coordonnée par le schéma de dualité et on obtient d'un seul coup des systèmes de vecteurs orthogonaux donc des systèmes de coordonnées non corrélées.

Ces remarques laisseront l'utilisateur indifférent mais non le lecteur attentif aux principes de fonctionnement. En effet, pour étendre l'ACOM au niveau suivant, on cherchera à maximiser :

$$f_q(\mathbf{u}_1^2, \mathbf{u}_2^2, \dots, \mathbf{u}_K^2, \mathbf{v}_2) = \sum_{j=1}^K q_j \left(\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^2 | \mathbf{v}_2 \right)_{\mathbf{D}}^2 \quad (1)$$

Il est important de savoir quelles contraintes seront employées puisque la dissymétrie engendrée par l'appariement des tableaux ne peut garantir en même temps l'orthogonalité des \mathbf{u}_j et la non corrélation des $\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j$ caractéristique des analyses simples.

On conservera une vue très proche des analyses simples en utilisant dans chaque espace :

$$(\mathbf{u}_j^1 | \mathbf{u}_j^2)_{\mathbf{Q}_j} = 0$$

et la contrainte commune $(\mathbf{v}_2 | \mathbf{v}_2)_{\mathbf{D}} = 0$. Ainsi de proche en proche, on appellera ACOM d'ordre m la recherche de K nouveaux vecteurs normés dans les espaces \mathbb{R}^{p_j} et un

nouveau vecteur de \mathbb{R}^n maximisant la même fonction sous contrainte de normalité et d'orthogonalité dans chacun des $K + 1$ espaces considérés. L'existence d'une solution et sa méthode de calcul sont dans l'article cité. On utilise la recherche d'un axe 1 d'une AFMULT sur un tableau recentré à chaque itération et donc une procédure très voisine des programmes anglo-saxons d'analyse des données étendue à K tableaux (voir par exemple ¹⁵). On obtient des bases orthonormées de sous-espaces de \mathbb{R}^{p_j} laissant la possibilité de projeter les nuages des points-lignes de chaque tableau sans obtenir toutefois la non corrélation des coordonnées de ces projections. La projection des nuages coordonnées par le critère de l'ACOM autorise cependant l'exploration simultanée de la géométrie de K nuages de n points.

3.3 — Mise en œuvre

L'option ACOM du module :

permet trois options de pondération :

L'option 2 utilise le poids 1 pour chaque tableau et n'a donc pas d'effet. L'option 1 pondère le tableau par l'inverse de son inertie totale et a donc pour fonction de ramener l'inertie totale de chaque analyse à l'unité. Ce sera sans effet si les inerties sont égales (c'est le cas de l'association de K ACP normées avec pondération uniforme des colonnes). L'option 3 pondère chaque analyse par l'inverse de sa première valeur propre et a pour effet de ramener la valeur propre de rang 1 de chaque triplet à l'unité. C'est le choix fait dans l'AFMULT. Si les inerties des tableaux sont très différentes mais que les types de répartition entre valeurs propres dans chaque analyse sont voisines, utiliser l'option 1. Si les tableaux sont du même type, en particulier s'ils supportent une AFC de Foucart (voir fiche 3 du fascicule 6), utiliser le type 2. Si les niveaux de structure entre tableaux sont très différents ou si les structures des tableaux sont peu marquées, comme dans les ACM, utiliser l'option 3. On choisit ici l'option 1 par défaut.

Generalised Ktab-PCA analysis
Input file A

Row number: 35, Column number: 23, Block number: 5
 Selected weighting option -> Total inertia from separate analysis

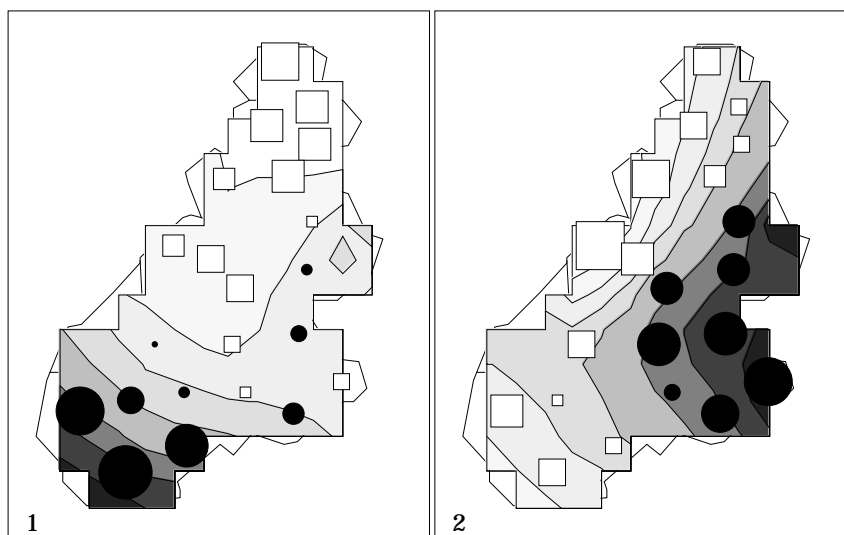
vp level 1 =
1. 8721e-01
vp level 2 =
6. 8555e-02

On trouve ici les maxima successifs de la fonction optimisée.

File A.--C1 contains column scores (unit norm)
 It has 23 rows and 2 columns
 File : A.--C1

----- Minimum/Maximum:
 Col.: 1 Mini = -1.2761 Maxi = 2.68
 Col.: 2 Mini = -1.9438 Maxi = 1.9355

Le fichier A.--C1 contient dans sa première colonne le vecteur v_1 . La deuxième colonne est le vecteur v_2 qui joue au pas suivant le même rôle que le vecteur v_1 .



Le couple de vecteurs donnent donc une carte factorielle fictive à deux coordonnées indépendantes et de variances unitaires, qui n'est pas une projection, mais une image de référence (ci-dessus) qu'on cherchera à reproduire par projection dans chaque espace de chaque tableau.

File A.--TC1rep contains column scores (unit norm)
 Duplication of the previous file
 It has 115 rows and 2 columns
 It is to be used with --TCc.label and --TCc.cat files
 File : A.--TC1rep

----- Minimum/Maximum:
 Col.: 1 Mini = -1.2761 Maxi = 2.68
 Col.: 2 Mini = -1.9438 Maxi = 1.9355

Dans le fichier A.TC1, ces coordonnées sont répétées par bloc pour superposition avec les suivantes.

File A.--TL1 contains row scores (unit norm for each block)
 It has 35 rows and 2 columns
 File : A.--TL1

----- Minimum/Maximum:
 Col.: 1 Mini = -1.1935 Maxi = 5.1906
 Col.: 2 Mini = -2.4962 Maxi = 3.5757

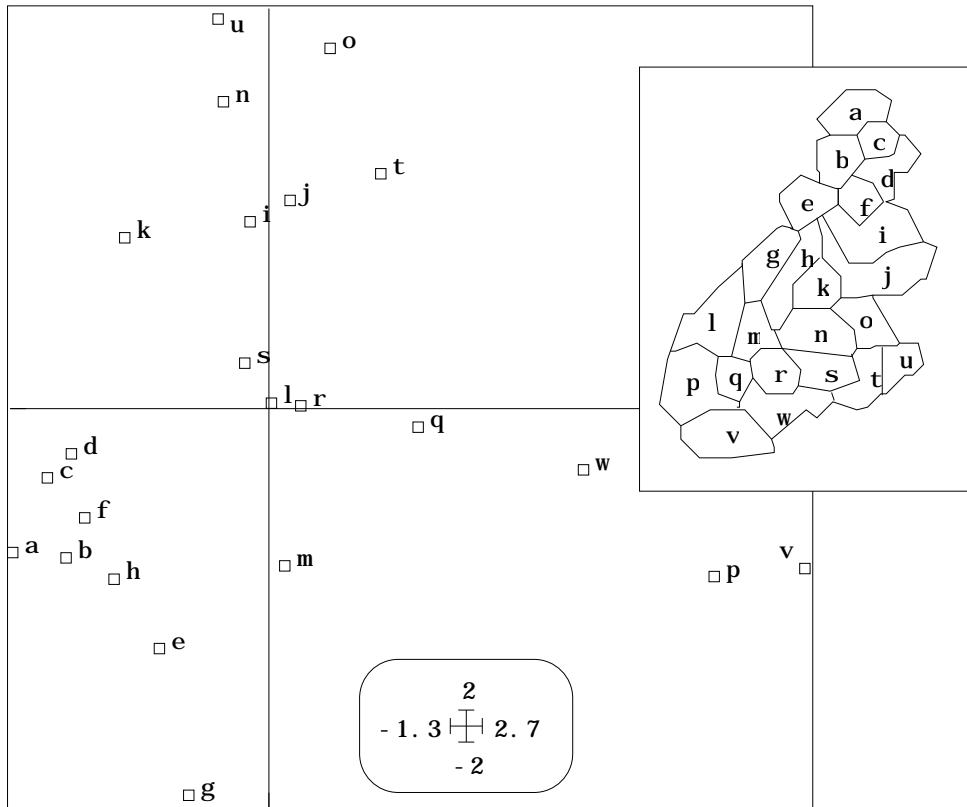


Figure 9 : Composantes principales normées de synthèse (vecteurs v).

Les lignes du K-tableaux dans sa disposition pratique sont les colonnes des tableaux initiaux. On trouve dans le fichier A.--TL1 les vecteurs u_j^1 et u_j^2 en colonne 1 et 2, superposés par tableau. Ce sont des poids canoniques de constitution des coordonnées pour les tableaux d'ACP (à la variante des poids uniforme près). Ils donnent le rôle des variables de chaque analyse dans la co-inertie multiple :

Labels			
XY coordinates file		A.--TL1	35 2
X-axis column number (default = 1)			
Y-axis column number (default = 2)			
Label file (or # for item numbers)		Code_Var_5Tab	
Draw vectors from origin (yes = 1)		1	

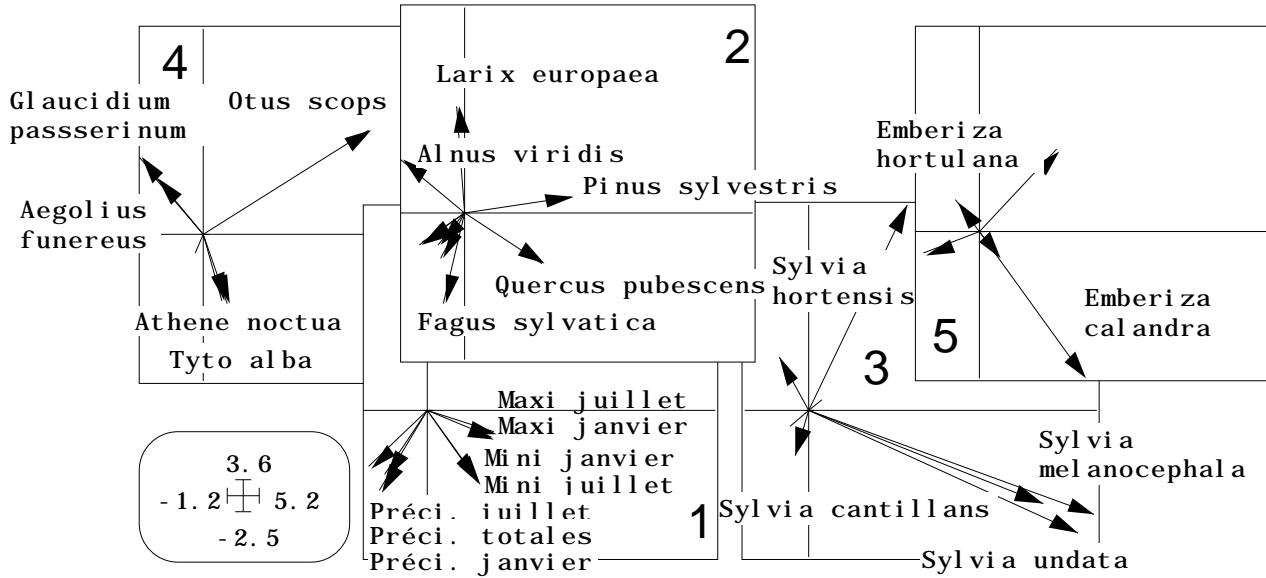
File A.--TCco contains column scores
 from projections of separate clouds
 It has 115 rows and 2 columns
 It is to be used with --TCC.label and --TCC.cat files
 File : A.--TCco

----- Minimum/Maximum:
 Col.: 1 Mini = -1.4036 Maxi = 2.0897
 Col.: 2 Mini = -1.5283 Maxi = 1.0366

File A.--TC1 contains normalized column scores
 from projections of separate clouds
 It has 115 rows and 2 columns

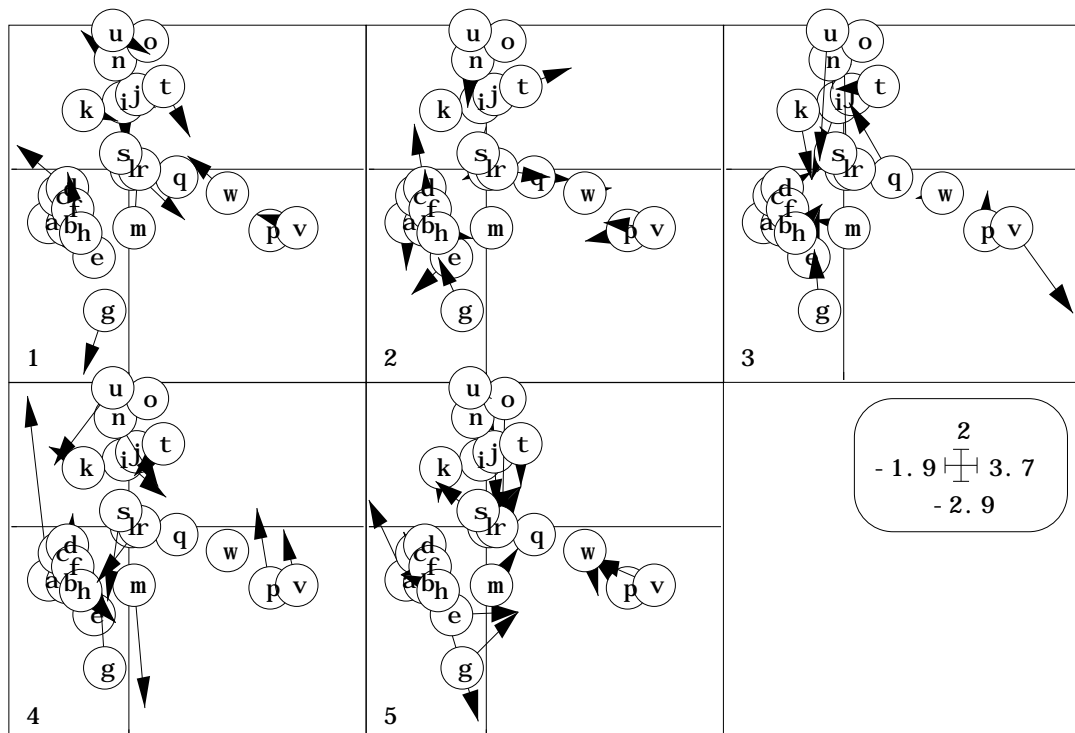
It is to be used with --TCc.label and --TCc.cat files
 It is to be matched with A.--TC1rep files
 File : A.--TC1

----- Minimum/Maximum:
 Col.: 1 Mini = -1.8994 Maxi = 3.6227
 Col.: 2 Mini = -2.843 Maxi = 1.9283



Dans chaque espace on a projeté le nuage des lignes (modèle théorique) d'un tableau sur les axes de coinertie multiple u_j^1 .

Ces coordonnées (fichier A.--TCco) sont ramenées à la variance unité pour comparaison à l'image de référence sont dans le fichier A.--TCc (notation générale mémotechnique des K -tableaux). Fondamentalement on résout ici le problème de la représentation simultanée des lignes des tableaux.



On a utilisé Scatters :

Match two scatters			
HY coordinates file		A.--TC1rep	115 2
H-axis column number (default = 1)			
Y-axis column number (default = 2)			
Second HY coordinates file		A.--TC1	115 2
Label file (or #) for rows		Code_rel	
Label file (or #) for groups		#	

avec la séparation des figures par tableau :

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

Les cinq analyses simultanées ainsi effectuées sont comparables aux cinq analyses séparées par un tableau du type co-inertie. On trouve successivement dans le listing le nombre de lignes du sous-tableau (les lignes du *K-tableau* pratique sont les colonnes du tableau de départ), le nombre de colonnes (les lignes de départ), le rang du tableau et son poids. Ce poids est celui qui intervient dans la formule (1). Le fichier A.sapa contient les inerties totales et les rangs des analyses de départ :

```

Input file: A.sapa
Row: 5 Col: 2
 1 | 1.0000 | 7.0000 |
 2 | 0.7473 | 9.0000 |
 3 | 0.5419 | 7.0000 |
 4 | 0.3502 | 5.0000 |
 5 | 0.2470 | 4.0000 |

```

On pondère les tableaux par l'inverse de leur inertie (respectivement 1, 1.3381, 1.846, 2.8555 et 4.0486). La somme vaut 11.0858 et on ramène les poids à l'unité ($1/11.0858 = 0.090$ noté 90 (en 1/1000)) :

```

Table n°: 1 * Rows: 7 * Col: 23 * Rank: 7
          * Weight (x1000): 90 * RV (x1000): 867

```

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	6.095e-01	6.095e-01	6.091e-01	6.091e-01	4.993e-01	8.199e-01
2	2.933e-01	9.028e-01	2.890e-01	8.980e-01	2.081e-01	7.200e-01

Les valeurs propres des analyses de départ sont dans A.savp :

```

Input file: A.savp
Row: 9 Col: 5
 1 | 0.6095 | 0.3128 | 0.3407 | 0.1703 | 0.1479 |
 2 | 0.2933 | 0.2321 | 0.0728 | 0.0963 | 0.0515 |
 3 | 0.0425 | 0.0688 | 0.0473 | 0.0570 | 0.0256 |
 4 | 0.0284 | 0.0420 | 0.0417 | 0.0159 | 0.0221 |
 5 | 0.0164 | 0.0327 | 0.0247 | 0.0107 | 0.0000 |
 6 | 0.0062 | 0.0291 | 0.0089 | 0.0000 | 0.0000 |
 7 | 0.0037 | 0.0152 | 0.0058 | 0.0000 | 0.0000 |
 8 | 0.0000 | 0.0101 | 0.0000 | 0.0000 | 0.0000 |
 9 | 0.0000 | 0.0046 | 0.0000 | 0.0000 | 0.0000 |

```

Elles sont éditées sous la rubrique Iner et Iner+ (dans la forme cumulée progressivement sur les axes d'inertie des analyses simples). A côté sont indiquées les variances et les variances cumulées des coordonnées sur les axes de co-inertie multiple. Ceci permet de comparer les analyses séparées et les analyses coordonnées. Sur le plan 1-2 de l'ACP normée de Meteo on projette une inertie de 0.9028 (c'est le maximum possible avec un plan) et sur le plan de co-inertie associée à Meteo on projette une inertie de 0.8980 (99.5 % de l'optimum).

On trouve encore RV (x1000), le coefficient RV entre le tableau d'origine et le tableau de synthèse formé par \mathbf{v}_1 et \mathbf{v}_2 . On trouve enfin les produits scalaires $(\mathbf{X}_1 \mathbf{Q}_1 \mathbf{u}_1 | \mathbf{v}_1)_{\mathbf{D}}^2$ et $(\mathbf{X}_1 \mathbf{Q}_1 \mathbf{u}_2 | \mathbf{v}_2)_{\mathbf{D}}^2$ et les carrés de corrélation associés (Cos2). La lecture de ces quantités pour chaque tableau montre l'adéquation des structures à 2 dimensions entre météorologie et végétation et la seule correspondance sur l'axe 1 entre les cinq tableaux.

Table n°: 2 * Rows: 9 * Col: 23 * Rank: 9
* Weight (x1000): 120 * RV (x1000): 914

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	3.128e-01	3.128e-01	2.484e-01	2.484e-01	2.013e-01	8.103e-01
2	2.321e-01	5.449e-01	2.933e-01	5.417e-01	2.533e-01	8.636e-01

Table n°: 3 * Rows: 8 * Col: 23 * Rank: 7
* Weight (x1000): 166 * RV (x1000): 759

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	3.407e-01	3.407e-01	3.327e-01	3.327e-01	2.666e-01	8.012e-01
2	7.280e-02	4.135e-01	4.979e-02	3.825e-01	2.201e-02	4.420e-01

Table n°: 4 * Rows: 6 * Col: 23 * Rank: 5
* Weight (x1000): 257 * RV (x1000): 754

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	1.703e-01	1.703e-01	1.659e-01	1.659e-01	1.367e-01	8.238e-01
2	9.633e-02	2.666e-01	8.616e-02	2.521e-01	2.752e-02	3.194e-01

Table n°: 5 * Rows: 5 * Col: 23 * Rank: 4
* Weight (x1000): 365 * RV (x1000): 757

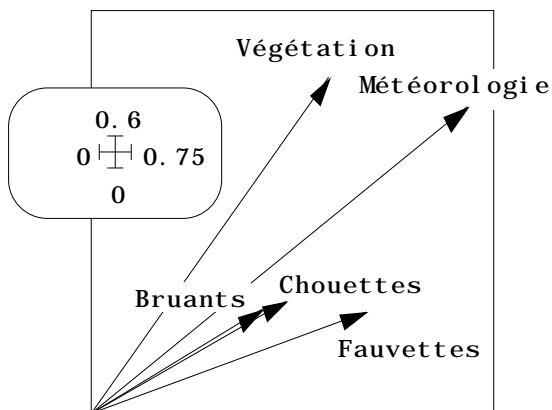
N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	1.479e-01	1.479e-01	1.453e-01	1.453e-01	1.049e-01	7.218e-01
2	5.148e-02	1.993e-01	5.250e-02	1.978e-01	2.319e-02	4.417e-01

La dernière propriété, et non la moindre, de l'ACOM, est enfin de faire de la typologie de tableaux après avoir fait de la typologie coordonnée des lignes et des colonnes de chaque tableau. En effet on peut utiliser comme coordonnée de rang 1 d'un tableau la coordonnée de rang 1 de l'ACP du tableau des $\mathbf{X}_k \mathbf{Q}_k \mathbf{u}_k^1$ (composantes principales synthétiques) affecté des norme \mathbf{D} et $Diag(q_1, \dots, q_K)$, ACP dont la composante principale est \mathbf{v}_1 . De même, on peut utiliser comme coordonnée de rang 2 d'un tableau la coordonnée de rang 1 de l'ACP du tableau des $\mathbf{X}_k \mathbf{Q}_k \mathbf{u}_k^2$ (composantes principales synthétiques) affecté des norme \mathbf{D} et $Diag(q_1, \dots, q_K)$, ACP dont la composante principale est \mathbf{v}_2 . Ces coordonnées de tableaux sont dans A.--tab.

File A.--tab contains table scores
from projections of principal components (XkDkuk)

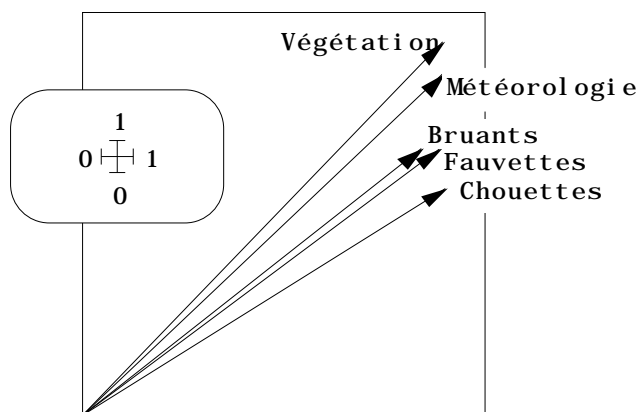
It has 5 rows and 2 columns
 File : A.--tab
 ----- Minimum/Maximum:
 Col.: 1 Mini = 0.3239 Maxi = 0.70665
 Col.: 2 Mini = 0.14835 Maxi = 0.50327

Le résultat est surprenant de pertinence:



Ces coordonnées sont toujours positives et l'on obtient une extension multidimensionnelle de la notion de l'axe 1 d'interstructure de STATIS. La version normalisée du graphe est associée au fichier A.--tab1 :

File A.--tab1 contains table scores
 from projections of principal components (Normed)
 It has 5 rows and 2 columns
 File : A.--tab1
 ----- Minimum/Maximum:
 Col.: 1 Mini = 0.84961 Maxi = 0.90762
 Col.: 2 Mini = 0.56512 Maxi = 0.92933



Au total, l'option ACOM ne présente aucune difficulté d'interprétation. L'examen d'une comparaison approfondie avec STATIS sur des problèmes concrets s'impose. Avec l'AFMULT, la relation est claire : les deux méthodes partagent la recherche de la première valeur propre puis leur logique interne les conduit à des démarches différentes. L'AFMULT est à l'ACOM, en partie, ce que l'ACP de deux tableaux juxtaposés et pondérés est à l'analyse de coïncidence.

Références

¹ Escoufier, Y. (1987) The duality diagramm : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute , Serie G .Springer Verlag, Berlin. 139-156.

² Bove, G. & Di Ciaccio, A. (1994) A user-oriented overview of multiway methods and software. *Computational Statistics and Data Analysis* : 18, 15-37.

³ Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.

Escoufier, Y. (1977) Operators related to a data matrix. In : *Recent developments in Statistics*. Barra, J.R. & Coll. (Eds.) North-Holland. 125-131.

Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis* : 18, 97-119.

⁴ Escoufier, B. & Pages, J. (1984) L'analyse factorielle multiple : une méthode de comparaison de groupes de variables. In : *Data Analysis and Informatics III*. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 41-55.

Escoufier, B. & Pagès, J. (1994) Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis* : 18, 121-140.

⁵ Rizzi, A. & Vichi, M. (1995) Representation, synthesis, variability and data preprocessing of a three-way data set. *Computational Statistics and Data Analysis* : 19, 203-222.

⁶ Sabatier, R. (1993) Critères et contraintes pour l'ordination simultanée de K tableaux. In : *Biométrie et Environnement*. Lebreton, J.D. & Asselain , B. (Eds.) Masson, Paris. 101-121.

⁷ Dagnélie, P. (1965) L'étude des communautés végétales par l'analyse statistique des liaisons entre les espèces et les variables écologiques : principes fondamentaux, un exemple. *Biometrics* : 21, 345-361 & 890-907.

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

⁹ Dolédec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.

¹⁰ Carrol, J.D. (1968) A generalization of canonical analysis to three or more sets of variables. *Proceeding of the 76th Convention of the American Psychological Association* : 3, 227-228.

¹¹ Hotelling, H. (1936) Relations between two sets of variates. *Biometrika* : 28, 321-377.

¹² Kettering, R.J. (1971) Canonical analysis of several sets of variables. *Biometrika* : 58, 433-451.

¹³ Saporta, G. (1975) *Liaisons entre plusieurs ensembles de variables et codage de données qualitatives*. Thèse de 3^o cycle, Université Pierre et Marie Curie, Paris VI. 1-102.

¹⁴ Tucker, L.R. . (1958) An inter-battery method of factor analysis. *Psychometrika* : 23, 2, 111-136.

¹⁵ Palmer, M. (1993) Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* : 74, 2215-2230.