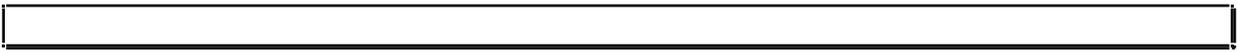


Manuel d'utilisation - V

Canonical

Révision - 03/06/98

Canonical : Canonical correlations Analysis.....	2
Canonical : CC coefficients.....	8
Canonical : Generalized Canonical Analysis.....	11
Canonical : RLS coefficients.....	23
Canonical : RV coefficients.....	30
Canonical : Test Sum_RV.....	40



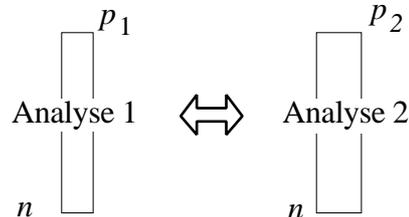
Canonical : Canonical correlations Analysis



Méthode d'analyse d'un couple de tableau : analyse canonique classique dite des corrélations.



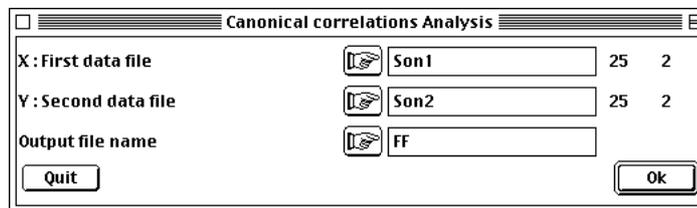
L'analyse canonique étudie un couple de tableaux relevant d'une ACP du type :



Elle élimine tous les effets associés aux moyennes, variances et covariances de chaque tableau pour donner des combinaisons linéaires des colonnes de chaque tableau de corrélation optimale.



L'option utilise une seule fenêtre de dialogue :



Nom du fichier binaire contenant le tableau X.

Nom du fichier binaire contenant le tableau Y.

Nom générique des fichiers de sortie (création).



Utiliser la carte Sons (données de Rao 1952¹ reprises dans Anderson 1958²).

First input file (X): Son1
Rows: 25 Columns: 2

----- Correlation matrix -----
[1] 1000
[2] 735 1000

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.7346E+00	+0.8673	+0.8673	02	+2.6544E-01	+0.1327	+1.0000

On retrouve la matrice de corrélations du premier tableau appelée dans Anderson (*op. cit.*) \mathbf{R}_{11} . Cette matrice de corrélation est diagonalisée pour des raisons techniques.

Second input file (Y): Son2
Rows: 25 Columns: 2

----- Correlation matrix -----
[1] 1000
[2] 839 1000

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.8393E+00	+0.9196	+0.9196	02	+1.6075E-01	+0.0804	+1.0000

On retrouve la matrice de corrélations du premier tableau appelée \mathbf{R}_{22} .

On calcule ensuite la matrice de corrélation croisée \mathbf{R}_{12} dont la transposée est \mathbf{R}_{21} (matrice des corrélations entre variables des deux tableaux) :

Generic output file name: FF
Crossed correlations rows = Y, col = X

```
-----
[ 1] 7.1075e-01 6.9316e-01
[ 2] 7.0398e-01 7.0855e-01
-----
```

```
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.2174E-01 +0.9954 +0.9954  | 02  +2.8880E-03 +0.0046 +1.0000  |
```

Dans Anderson, la matrice $\mathbf{R}_{21}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$ est diagonalisée. Dans l'option présente le calcul est conduit de manière différente mais on retrouve les mêmes valeurs propres. Anderson donne 0.621,816 et 0.002,900 (p. 304) ce qui s'accorde sur quatre décimales.

```
Canonical correlation coefficients
k= 1 rk = 7.8851e-01 rk2 = 6.2174e-01
k= 2 rk = 5.3740e-02 rk2 = 2.8880e-03
```

Squared canonical correlation coefficients are in the file FF.canvp

Les valeurs propres sont des carrés de corrélation et les racines des valeurs propres sont les corrélations canoniques (Anderson donne 0.788,553 et 0.053,852). Noter sa remarque : “ λ_2 is correct for only four or five significant figures” ce qui est vrai.

L'analyse fournit les combinaisons linéaires des variables normalisée du tableau 1 qu'on appelle variables canoniques du tableau X :

```
File FF.canl1 contains canonical variates (X)
Rows = 25 Col = 2
File :FF.canl1
```

Col.	Mini	Maxi
1	-1.798e+00	2.331e+00
2	-2.273e+00	2.061e+00

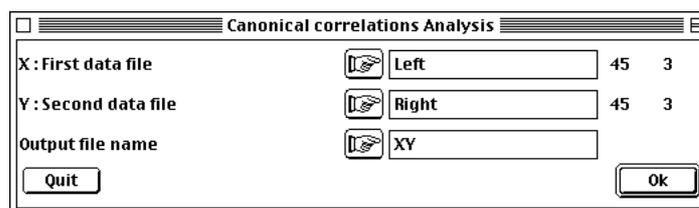
L'analyse fournit les combinaisons linéaires des variables normalisée du tableau 2 qu'on appelle variables canoniques du tableau Y :

```
File FF.canl2 contains canonical variates (Y)
Rows = 25 Col = 2
File :FF.canl2
```

Col.	Mini	Maxi
1	-1.546e+00	2.747e+00
2	-2.494e+00	2.347e+00

Le principe fondamental est que la première variable canonique de X et la première variable canonique de Y sont de corrélation maximale. Cette corrélation est la corrélation canonique d'ordre 1. “*The larger of the two canonical correlations, 0.789, is higher than any of the individual correlations of a variable of the first set with a variable of the other.*” L'analyse canonique est une des plus anciennes et des plus connues des méthodes multivariée. Elle est attribuée à Hotelling (1936)³. Pour s'en servir, il convient de citer l'ouvrage de référence de Gittins (1985)⁴, qu'on utilise ci-dessous pour vérifier les calculs.

Utiliser la carte Anglesey (Gittins p. 302) et retrouver les résultats de l'analyse pp. 132 et suivantes.



On obtient les corrélations du tableau X :

```
First input file (X): Left
```

Rows: 45 Columns: 3

----- Correlation matrix -----

```
[ 1] 1000
[ 2] 344 1000
[ 3] 378 775 1000
```

```
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +2.0282E+00 +0.6761 +0.6761 | 02  +7.4779E-01 +0.2493 +0.9253 |
03  +2.2398E-01 +0.0747 +1.0000
```

les corrélations du tableau Y :

Second input file (Y): Right

Rows: 45 Columns: 3

----- Correlation matrix -----

```
[ 1] 1000
[ 2] -225 1000
[ 3] -661 94 1000
```

```
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +1.7310E+00 +0.5770 +0.5770 | 02  +9.4349E-01 +0.3145 +0.8915 |
03  +3.2546E-01 +0.1085 +1.0000
```

les corrélations croisées :

Generic output file name: XY

Crossed correlations rows = Y, col = X

```
-----
[ 1] -2.2386e-01 -3.7493e-01 -3.5762e-01
[ 2] -1.3940e-01 -5.1734e-01 -4.6119e-01
[ 3] 3.0832e-01 3.6858e-01 5.8343e-01
```

```
-----
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +6.3143E-01 +0.7405 +0.7405 | 02  +2.1358E-01 +0.2505 +0.9910 |
03  +7.6691E-03 +0.0090 +1.0000
```

le test de dimensionalité de Bartlett (*ibidem*) :

Likelihood ratio tests of dimensionality

Barlett 1938, see Ch. 3.4.2 of Gittins, R. (1985) Canonical analysis, a review with applications in ecology. Springer-Verlag, Berlin. 1-351

```
-----
k= 0 Khi2 = 5.0466e+01 ddl = 9 proba = 1.7078e-07
k= 1 Khi2 = 1.0290e+01 ddl = 4 proba = 3.5427e-02
k= 2 Khi2 = 3.2719e-01 ddl = 1 proba = 5.7463e-01
```

les corrélations canoniques (tableau 6.1 p. 132) :

Canonical correlation coefficients

k= 1 rk = 7.9462e-01 rk2 = 6.3143e-01

k= 2 rk = 4.6214e-01 rk2 = 2.1358e-01

k= 3 rk = 8.7573e-02 rk2 = 7.6691e-03

Squared canonical correlation coefficients are in the file XY.canvp

les variables canoniques du premier tableau :

File XY.canl1 contains canonical variates (X)

Rows = 45 Col = 3

File :XY.canl1

Col.	Mini	Maxi
1	-5.741e-01	3.584e+00
2	-3.313e+00	2.207e+00
3	-2.095e+00	1.642e+00

les variables canoniques du second tableau :

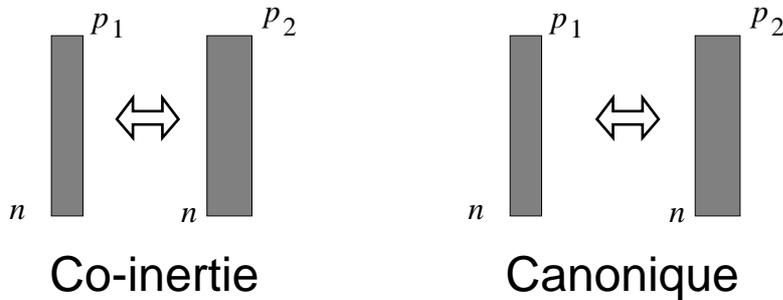
File XY.canl2 contains canonical variates (Y)

Rows = 45 Col = 3

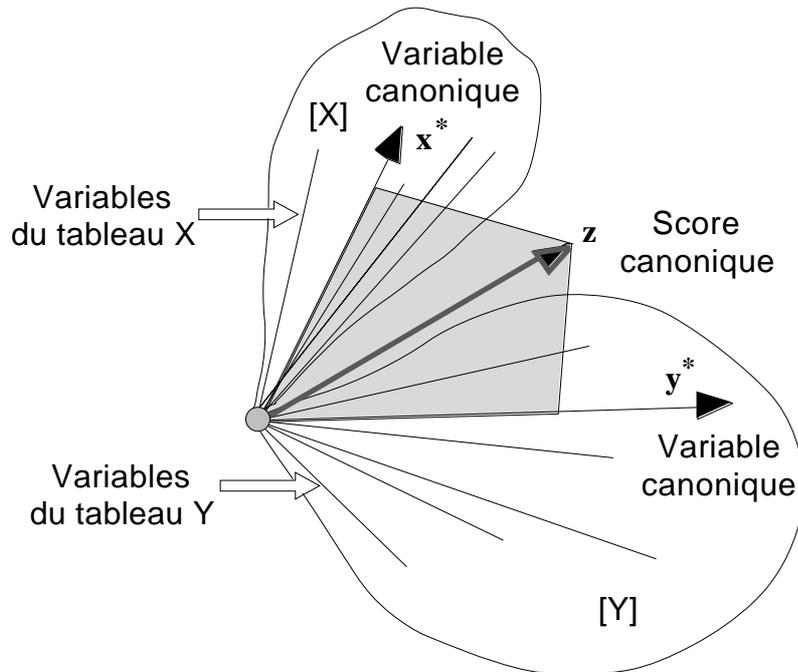
File :XY.canl2

Col.	Mini	Maxi
1	-1.168e+00	2.530e+00
2	-2.469e+00	2.589e+00
3	-2.155e+00	1.784e+00

Les aides à l'interprétation disponibles s'appuient sur une vision géométrique dans l'espace des variables. Fondamentalement un couplage de tableaux par l'analyse de co-inertie s'appuie sur le concept des deux tableaux comme deux nuages de n lignes dans deux espaces différents, alors que l'analyse canonique s'appuie sur le concept des deux tableaux comme deux nuages de colonnes dans le même espace :



On a retenu le schéma de principe (voir fiche Thema 4.A) :



Toutes les corrélations calculées décrivent les angles entre les vecteurs. On obtient les scores canoniques :

File XY.canll contains canonical scores (Bissectors)
 Rows = 45 Col = 3

File :XY.canll

Col.	Mini	Maxi
1	-8.672e-01	3.167e+00
2	-2.960e+00	1.795e+00
3	-2.532e+00	2.046e+00

les corrélations entre variables de X et scores canoniques :

File XY.cavl contains correlations between variables (X) and canonical scores

Rows = 3 Col = 3
File :XY.cav1

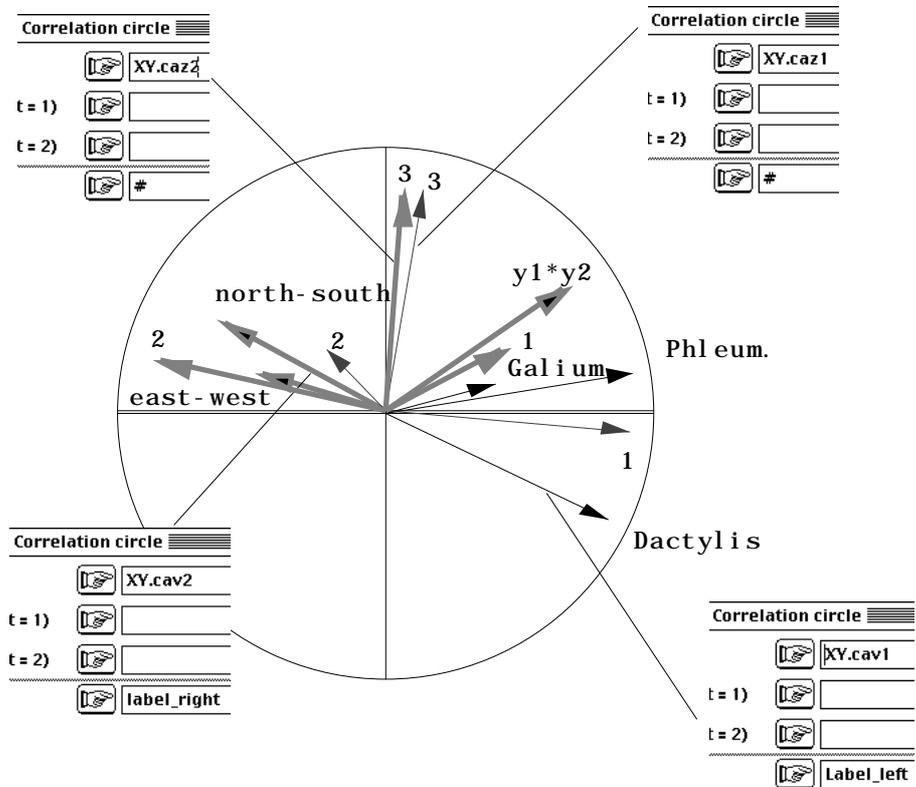
Col.	Mini	Maxi
1	4.101e-01	9.298e-01
2	-4.099e-01	1.496e-01
3	-6.582e-01	5.711e-02

les corrélations entre variables de Y et scores canoniques :

File XY.cav2 contains correlations between variables (Y) and canonical scores

Rows = 3 Col = 3
File :XY.cav2

Col.	Mini	Maxi
1	-5.916e-01	6.601e-01
2	1.399e-01	4.566e-01
3	-5.041e-01	6.343e-01



les corrélations entre composantes principales de X et scores canoniques :

File XY.caz1 contains correlations between principal components (X) and canonical scores

Rows = 3 Col = 3
File :XY.caz1

Col.	Mini	Maxi
1	-2.120e-01	9.114e-01
2	-8.041e-02	8.209e-01
3	-6.919e-01	1.714e-01

les corrélations entre composantes principales de Y et scores canoniques :

File XY.caz2 contains correlations between principal components (Y) and canonical scores

Rows = 3 Col = 3
File :XY.caz2

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

1	-8.374e-01	4.380e-01
2	1.827e-01	8.033e-01
3	-6.233e-01	2.474e-01

On peut superposer les cercles de corrélations associés (ci-dessus).

On pourra retrouver encore les corrélations *intra-sets* entre variables de X et variables canoniques de X (entre variables de Y et variables canoniques de Y) et les corrélations *inter-sets* entre variables de X et variables canoniques de Y (entre variables de Y et variables canoniques de X) par l'option MatAlg: Diagonal Inner product $C=X'DY$. Par exemple (Table 6.2 p. 134) :

Input file for X matrix

Option for X matrix (default=none)

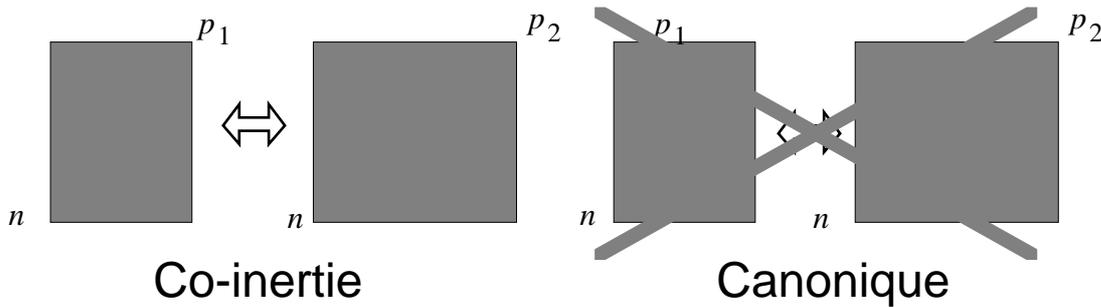
Input file for Y matrix

Option for Y matrix (default=none)

[1]	4.3297e-01	8.7720e-01	9.8153e-01
[2]	1.2620e-01	-4.7935e-01	1.7494e-01
[3]	-8.9253e-01	-2.7139e-02	7.7451e-02

⚠ Si on remplace un des tableaux par une transformation linéaire de rang plein (en particulier, si on remplace les variables centrées par les variables normées, ou les variables par les coordonnées de l'ACP normée du tableau) on obtient le même résultat.

🖼 Retenir le schéma associé aux dimensions des tableaux :



🔄 L'extension aux cas de plus de deux tableaux est disponible dans Canonical: Generalized Canonical Analysis.



- 1 Rao, C.R. (1952) *Advanced Statistical Methods in Biometric Research*. John Wiley and Sons, New York, p.245.
- 2 Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York, p. 58 et 303.
- 3 Hotelling, H. (1936) Relations between two sets of variates. *Biometrika* : 28, 321-377.
- 4 Gittins, R. (1985) *Canonical analysis, a review with applications in ecology*. Springer-Verlag, Berlin. 1-351.

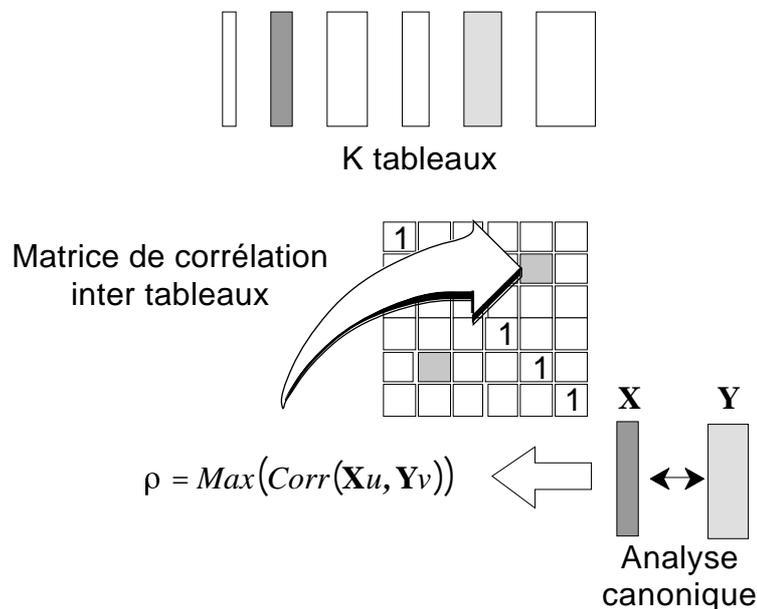
Canonical : CC coefficients



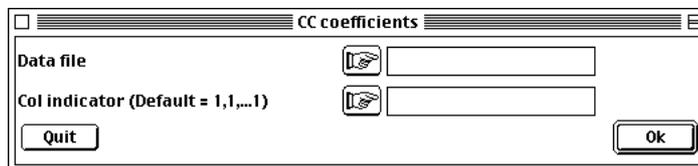
Typologie de tableaux sur corrélations canoniques.



Dans un K -tableaux, on cherche à définir ressemblance et dissemblances entre tableaux. Ici la ressemblance est mesurée par la corrélation canonique, c'est-à-dire la racine de la première valeur propre de l'analyse canonique de deux tableaux. La différence est mesurée par la distance déduite de cette mesure de corrélation.



L'option utilise une seule fenêtre de dialogue :



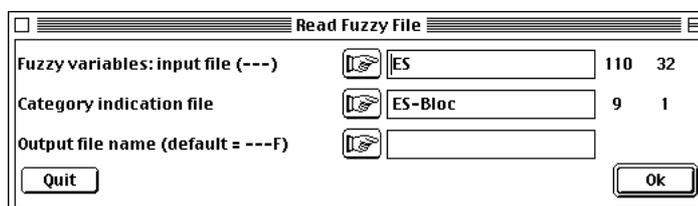
Nom du fichier binaire d'entrée (n lignes-individus et p colonnes-variables). L'option ne fait aucun centrage préalable et s'utilise normalement sur un tableau centré ou normalisé *a priori*.

Nom du fichier binaire indicateur de blocs de variables. Ce fichier doit être compatible avec le précédent. Il a une seule colonne et autant de lignes qu'il y a de tableaux (K). A la ligne k , on y trouve le nombre de variables (p_k) du tableau k . On doit avoir :

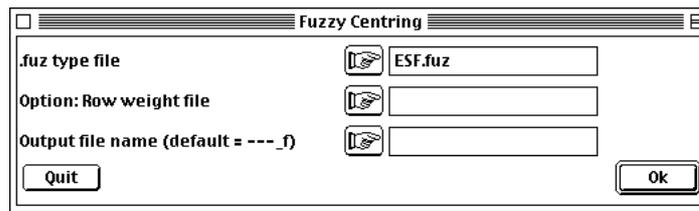
$$\sum_{k=1}^K p_k = p$$



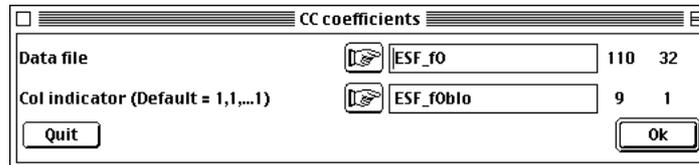
Utiliser la carte Coléoptères. Le tableau ES de variables floues est lu par FuzzyVar: Read Fuzzy File :



Centrer simplement le tableau par FuzzyVar: Fuzzy Centring :



Chaque variable floue définit un tableau centré par colonne.



Les paramètres d'entrée sont :

```
Input file: ESF_f0
-> Rows: 110, columns: 32
-> 9 blocs: 4/5/4/2/4/3/5/3/2/
```

Le rang de chaque tableau est calculé par ACP séparées. Chaque variable floue à m modalités définit un tableau de rang $m-1$:

```
Block: 1 Dim: 110 - 4 Rank: 3
Block: 2 Dim: 110 - 5 Rank: 4
Block: 3 Dim: 110 - 4 Rank: 3
Block: 4 Dim: 110 - 2 Rank: 1
Block: 5 Dim: 110 - 4 Rank: 3
Block: 6 Dim: 110 - 3 Rank: 2
Block: 7 Dim: 110 - 5 Rank: 4
Block: 8 Dim: 110 - 3 Rank: 2
Block: 9 Dim: 110 - 2 Rank: 1
```

Chaque couple de deux tableaux fournit une analyse canonique dont on garde le coefficient de corrélation canonique de rang 1. Ces coefficients forment une matrice symétrique de corrélations entre tableaux. Cette matrice est éditée :

```
Canonical correlation coefficients
Hotelling 1936
----- Correlation matrix -----
[ 1] 1000
[ 2] 519 1000
[ 3] 630 852 1000
[ 4] 480 735 652 1000
[ 5] 458 830 747 780 1000
[ 6] 311 375 377 117 282 1000
[ 7] 477 546 632 392 499 994 1000
[ 8] 385 699 578 696 713 135 360 1000
[ 9] 274 826 628 659 621 163 416 809 1000
-----
```

Elle est conservée dans un fichier :

```
ESF_f0_CC1 is a binary file with 9 rows and 9 columns
Content: Canonical correlation (sqrt(lambda1))
```

La matrice qui précède contient les corrélations canoniques comprises entre 0 et 1. Si K est le nombre de tableau :

$$\mathbf{R} = [\rho_{ij}]_{1 \leq i, j \leq K} \quad \mathbf{D} = [d_{ij}]_{1 \leq i, j \leq K}$$

avec $d_{ij} = \sqrt{2(1 - \rho_{ij})}$.

```
ESF_f0_CC.dist is a binary file with 9 rows and 9 columns
Content: Among array distances = sqrt(2(1-r))
Non euclidean distance
```

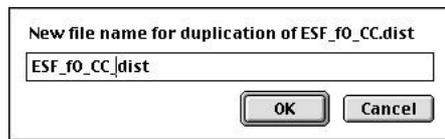
```
cancor squared
----- Correlation matrix -----
[ 1] 1000
[ 2] 270 1000
[ 3] 397 726 1000
[ 4] 230 540 425 1000
[ 5] 209 689 558 608 1000
[ 6] 97 141 142 14 79 1000
[ 7] 228 298 400 154 249 987 1000
[ 8] 148 488 334 484 508 18 129 1000
[ 9] 75 682 394 434 385 26 173 654 1000
-----
```

ESF_fo_CC2 is a binary file with 9 rows and 9 columns
 Content: Squared canonical correlation (lambda1)

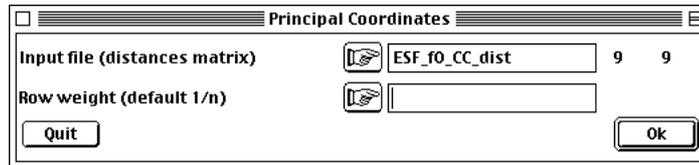
On garde et on édite enfin la matrice des carrés de corrélations canoniques.



La matrice de distances ainsi obtenue n'est pas euclidienne. Renommer le fichier qui la contient pour éliminer le point d'extension :



Dans Distances :



File ESF_fo_CC_dist.pp contains the matrix $a_{ij} - a_{i.} - a_{.j} + a_{..}$
 with $a_{ij} = -d_{2ij}^2/2$
 --- It has 9 rows and 9 columns

Num	Eigenval.	Num.	Eigenval.	Num.	Eigenval.	Num.	Eigenval.
001	1.868e-01	002	8.474e-02	003	4.700e-02	004	4.175e-02
005	2.581e-02	006	1.635e-02	007	6.323e-03	008	2.007e-17
009	-4.900e-03						

Il y a une valeur propre négative. Comme les indices RV et RLS donnent des distances euclidiennes, cette analyse en coordonnées principales ne s'impose pas comme un outil utile.



Cette option permet une approche rapide des liens entre tableaux mais la méthode STATIS de Canonical: RV coefficients est beaucoup plus efficace.

Canonical : Generalized Canonical Analysis



Méthodes K -tableaux, généralisant l'analyse canonique inventée par Carrol (1968) ¹. Le programme utilise l'étude de Tenenhaus (1984) et intègre les aides à l'interprétation définies par Casin et Turlot (1986) ³. Analyse canonique généralisée (ACG) et analyse de co-inertie multiple (ACOM) entretiennent les mêmes relations que l'analyse canonique (AC) et l'analyse de co-inertie (ACO).



K tableaux d'ACP donnent K sous-espaces de variables (ensembles de combinaisons linéaires dans lequel on trouve les composantes principales de chaque analyse séparée). On cherche dans le sous-espace k une variable canonique \mathbf{z}_k et une variable de synthèse (score canonique) \mathbf{z} de manière à optimiser :

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



L'option utilise une seule fenêtre de dialogue :

Nom du fichier binaire d'entrée (n lignes-individus et p colonnes-variables).

Nom du fichier binaire indicateur de blocs de variables. Ce fichier doit être compatible avec le précédent. Il a une seule colonne et autant de lignes qu'il y a de tableaux (K). A la ligne k , on y trouve le nombre de variables (p_k) du tableau k . On doit avoir :

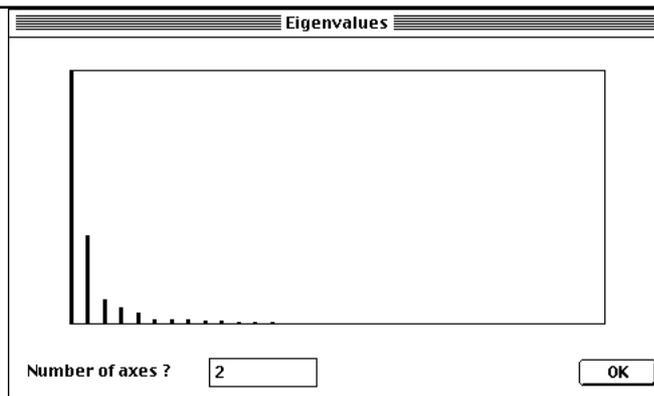
$$\sum_{k=1}^K p_k = p$$

Nom générique des fichiers de sortie (création).



Utiliser la carte Coléoptères. Le tableau ES de variables floues est lu par FuzzyVar: Read Fuzzy File :

Centrer simplement le tableau par PCA: Fuzzy PCA :



Chaque variable floue définit un tableau centré par colonne.

Generalized Canonical Analysis

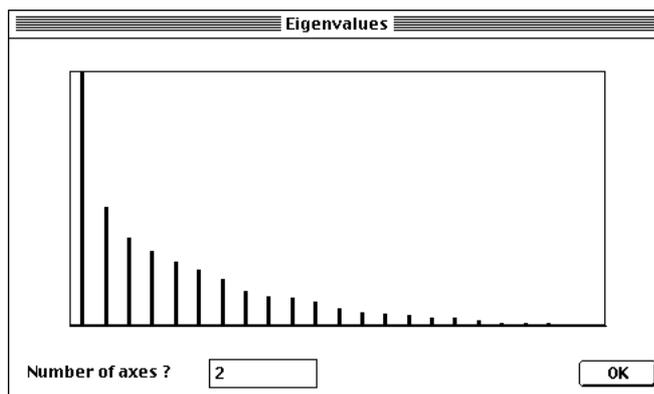
Data file: A.cpta 110 32

Col indicator (Default = 1,1,...,1): ES-Bloc 9 1

Output file name: AA

Buttons: Quit, Ok

Il s'affiche un graphe de valeurs propres :



Input file: A.cpta
 -> Rows: 110, columns: 32
 -> 9 blocs: 4/5/4/2/4/3/5/3/2/

Block: 1 Dim: 110- 4 Rank: 3
 Block: 2 Dim: 110- 5 Rank: 4
 Block: 3 Dim: 110- 4 Rank: 3
 Block: 4 Dim: 110- 2 Rank: 1
 Block: 5 Dim: 110- 4 Rank: 3
 Block: 6 Dim: 110- 3 Rank: 2
 Block: 7 Dim: 110- 5 Rank: 4
 Block: 8 Dim: 110- 3 Rank: 2
 Block: 9 Dim: 110- 2 Rank: 1

Noter d'abord que le programme admet des tableaux dont les colonnes ne sont pas indépendantes. Pour une variable floue à m modalités, le rang du tableau est au plus égal à $m-1$. De 32 modalités, on passe à $32-9 = 23$ dimensions. Chaque tableau est remplacé par ses composantes principales par diagonalisation de sa matrice de covariances. Suit la diagonalisation de la matrice de covariances du tableau transformé :

Num	Eigenval.	Num	Eigenval.	Num	Eigenval.	Num	Eigenval.
001	6.537e+00	002	3.041e+00	003	2.272e+00	004	1.907e+00
005	1.649e+00	006	1.434e+00	007	1.182e+00	008	8.722e-01
009	7.432e-01	010	7.027e-01	011	6.109e-01	012	4.228e-01
013	3.233e-01	014	3.018e-01	015	2.503e-01	016	2.020e-01
017	1.850e-01	018	1.357e-01	019	7.017e-02	020	6.687e-02
021	5.939e-02	022	2.877e-02	023	2.944e-03		

Ces valeurs propres sont les valeurs du critère optimisé :

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

Cela signifie qu'on peut trouver une variable de synthèse (score canonique) et une variable par tableau (variables canoniques) qui donne une somme de carrés de corrélation de 6.54.

Eigenvalues in file AA_vpro - Rows: 23 -Col: 1

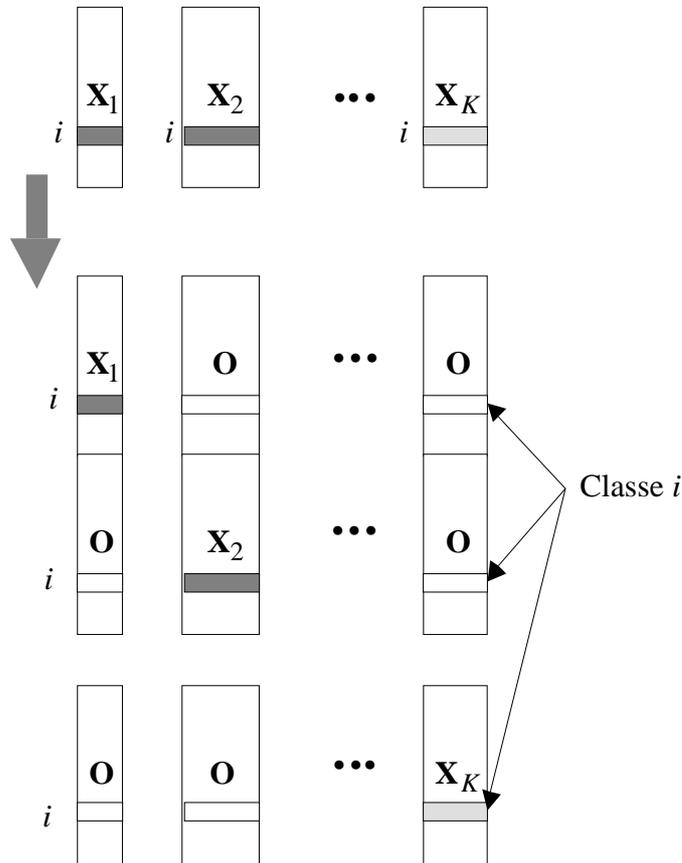
Les scores canoniques sont disponibles dans :

Canonical scores in file AA_casc - Rows: 110 -Col: 2

File :AA_casc

Col.	Mini	Maxi
1	-8.579e-01	2.296e+00
2	-2.538e+00	1.672e+00

Le programme conserve ensuite les aides à l'interprétation de Casin et Turlot (1986)³. La structure de cette information dérive de l'approche :



La ligne i du tableau k est une cellule et les différentes cellules associées au même individu définissent une classe. L'ensemble des cellules reçoit pour chaque score canonique retenu un nouveau code numérique appelé score des cellules. Ces scores de cellules sont de variance unité, de corrélations nulles entre elles, et optimisent la variance des moyennes par classes. L'ACG est vue alors comme une véritable analyse discriminante :

Cell scores are in the file AA_cell

Rows: 990 - Col: 2

990 = tab number [9] x row number [110]

To be used with labels in AA_celq.lab and the next files

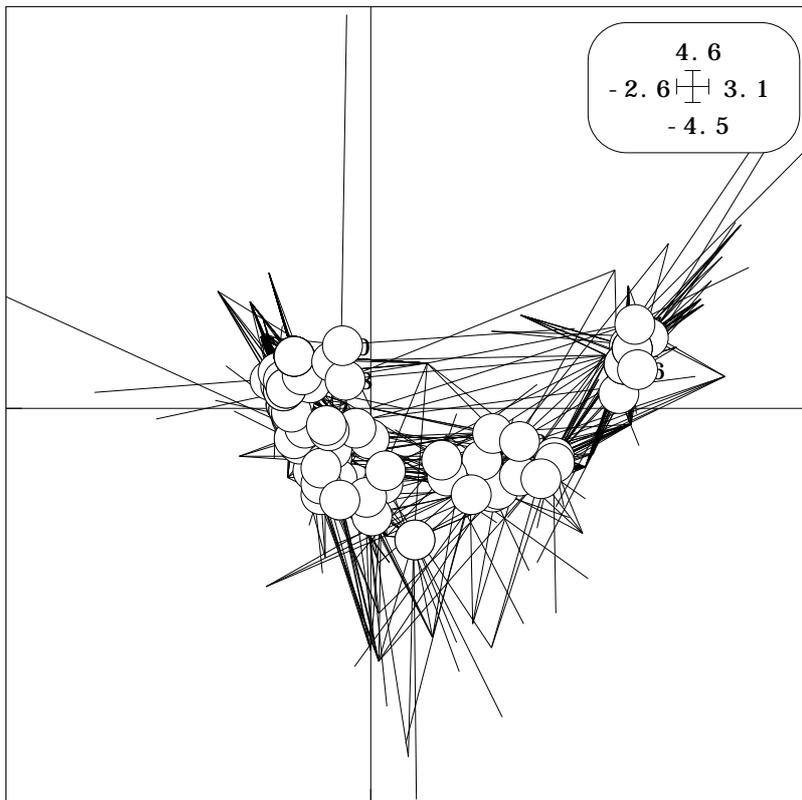
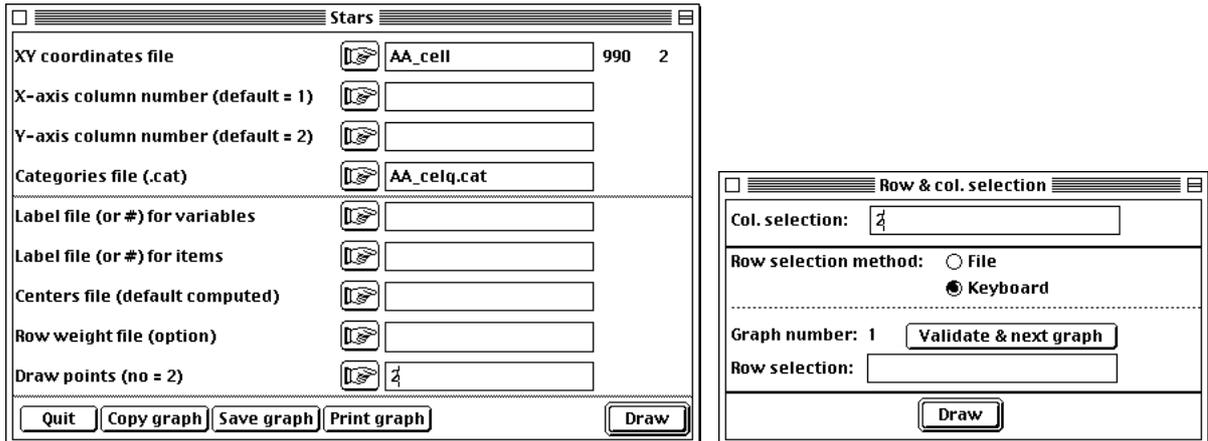
Les moyennes par classes (individus) sont dupliquées dans :

Means of the cell scores duplicated in file AA_celm
 Rows: 990 - Col: 2
 990 = tab number [9] x row number [110]

Lorsqu'on attribue les scores aux cellules pour faire une analyse discriminante, les classes (individus) sont des moyennes. Ces moyennes sont directement liées aux scores canoniques. On a proportionnalité simple mais les scores canoniques sont de variances 1 tandis que les moyennes sont de variances v_k . Ces variances v_k sont directement liées aux valeurs propres de l'analyse par la relation :

$$v_k = \frac{\lambda_k}{K}$$

La remarquable innovation des auteurs cités conduit à une figure de synthèse du type :



On a utilisé les variables qualitatives nécessaires implantées dans les fichiers :

Description of a coded matrix

Qualitative variables file: AA_celq
 Number of rows: 990, variables: 2, categories: 119

Description of categories:

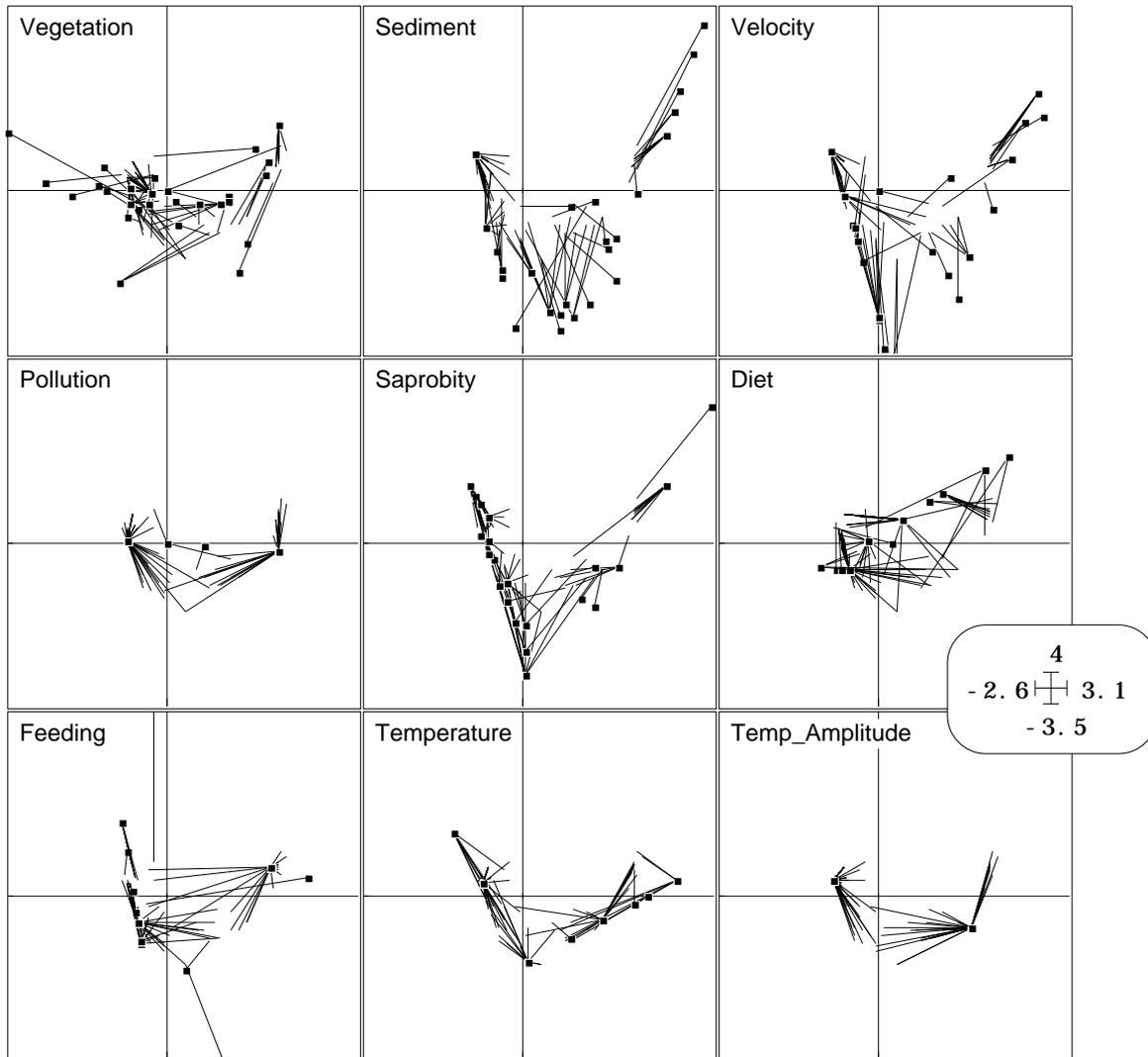
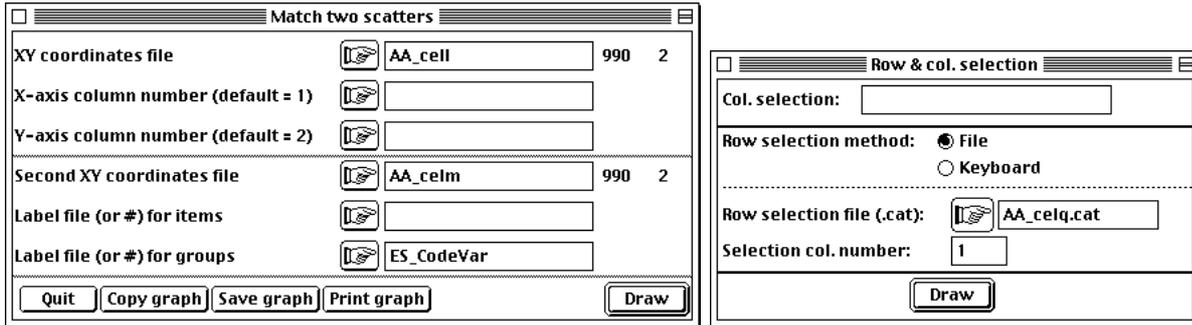
```

-----
Variable number 1 has 9 categories
-----
[  1]Category:  111 Num:  110 Freq.:  0.1111
...
[  9]Category:  111 Num:  110 Freq.:  0.1111

Variable number 2 has 110 categories
-----
[ 10]Category:  111 Num:    9 Freq.:  0.009091
...
[119]Category:  111 Num:    9 Freq.:  0.009091

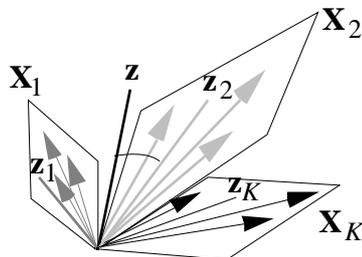
```

Sur la figure précédente chaque ligne est la classe de ses représentations par chaque tableau. Pour rendre lisible cette figure d'analyse discriminante, on peut utiliser :



 Auxiliary ASCII output file AA_celq.123: labels (two characters) for 119 modalities
 It contains one label for each modality
 It has 119 rows (modalities) and labels 1,2, ..., 9

Le programme reprend l'interprétation traditionnelle de l'ACG qui porte sur les variables tandis que les propositions de Casin et Turlot permettaient de voir les individus. Le score canonique fait avec chaque sous-espace un angle décrit par le carré de son cosinus :

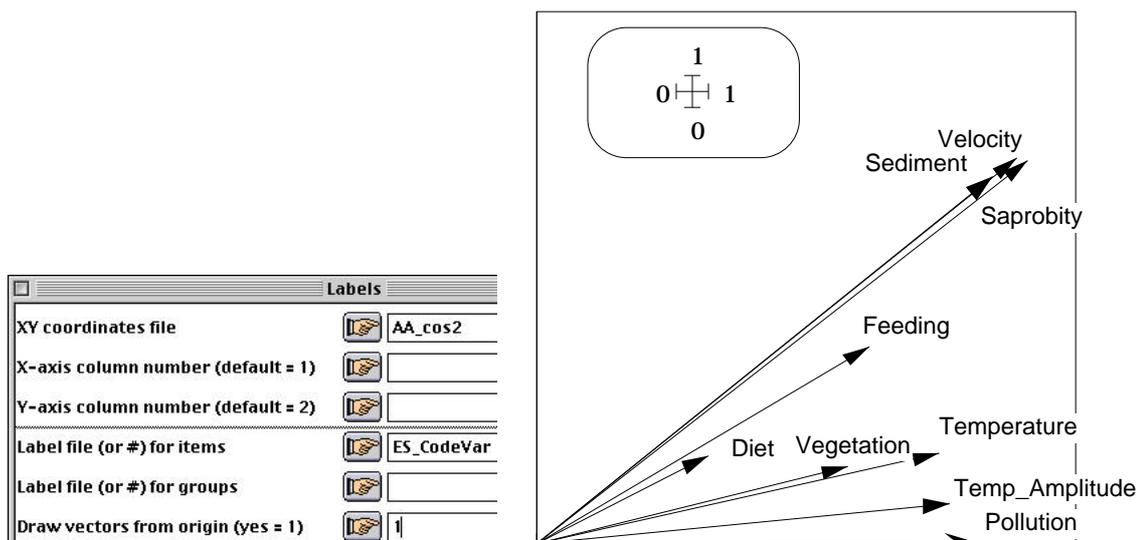


Cos 2 between scores and subspaces

	fac 1	fac 2
Tab 1	5.818e-01	1.416e-01
Tab 2	9.185e-01	7.171e-01
Tab 3	8.977e-01	7.239e-01
Tab 4	8.144e-01	4.086e-03
Tab 5	8.519e-01	6.866e-01
Tab 6	3.236e-01	1.607e-01
Tab 7	6.219e-01	3.684e-01
Tab 8	7.534e-01	1.670e-01
Tab 9	7.735e-01	7.187e-02
Total	6.537e+00	3.041e+00

These values are in the file AA_cos2
 Rows: 9 - Col: 2

Ces valeurs permettent une représentation des tableaux :

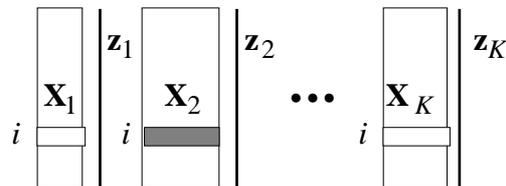


Les valeurs propres étant des sommes de \cos^2 se décompose en pourcentage :

Eigenvalue ratios 1/10000

	fac 1	fac 2
Tab 1	890	466
Tab 2	1405	2358
Tab 3	1373	2380
Tab 4	1246	13
Tab 5	1303	2258
Tab 6	495	528
Tab 7	951	1211
Tab 8	1153	549
Tab 9	1183	236
Total	10000	10000

Les scores canoniques se projettent sur les sous-espaces de variables. La projection sur un sous-espace est portée par une variables normalisée qui s'appelle la variable canonique associée à ce sous-espace :



On conserve ces variables canoniques pour faire des graphes canoniques qui expriment directement le critère optimisé.

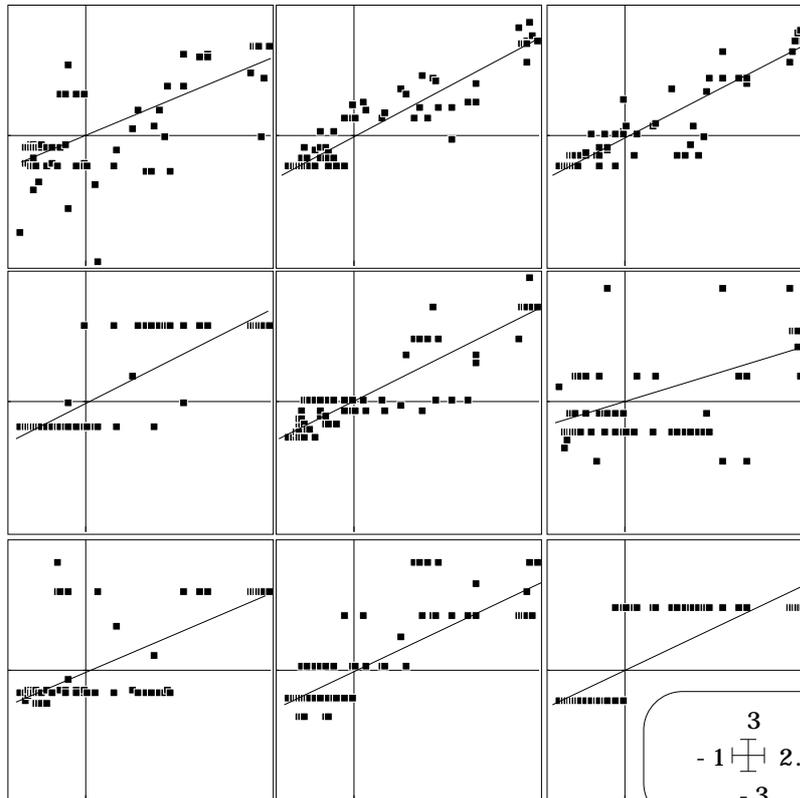
Polynomials

X file (default = 1, 2, 3, ..., n)

X file column number (default = 1)

Y file (no default)

Order of polynomial (default = 1) ?



En abscisse, on prend le score canonique de rang 1. En ordonnée, on prend les 9 variables canoniques de rang 1. Chacune d'entre elle est une combinaison linéaire normalisée des variables du tableau correspondant. On résume ainsi chaque tableau par une variable, l'ensemble de ces variables étant collectivement le plus lié possible à une variable de synthèse.

On comprend alors que l'ACP normée est le cas particulier avec une variable par tableau et que l'ACM est le cas particulier avec un paquet d'indicatrices d'une variable qualitative par tableau. Donc l'analyse de Hill & Smith est le cas particulier avec une variable qualitative ou quantitative par tableau. De ce point de vue, l'ACG est l'extension aux variables floues de l'ACP normée des variables quantitatives et de l'ACM des variables qualitatives.

L'ACG permet de mélanger dans un même point de vue les qualitatives, les quantitatives et les floues.

```
Canonical variables number 1 in file AA_cav1
```

```
Rows(samples) : 110 - Col(Tables) : 9
```

```
File :AA_cav1
```

Col.	Mini	Maxi
1	-2.891e+00	2.025e+00
2	-6.953e-01	2.617e+00

```
...
```

9	-6.831e-01	1.464e+00
---	------------	-----------

```
Canonical variables number 2 in file AA_cav2
```

```
Rows(samples) : 110 - Col(Tables) : 9
```

```
File :AA_cav2
```

Col.	Mini	Maxi
1	-3.115e+00	2.171e+00
2	-2.058e+00	2.440e+00

```
...
```

9	-1.464e+00	6.831e-01
---	------------	-----------

On remarquera que les aides à l'interprétation permettent de voir les variables canoniques soit comme des scores numériques (on représente alors la valeur de chaque individu) soit comme des vecteurs :

```
Cosinus between canonical scores and canonical variables
```

```
File AA_corr - Rows: 18 -Col: 2
```

```
Sort by block number [1,9] and score number [1,2]
```

```
To be used with labels in AA_corq.lab and the next files
```

Les variables qualitatives nécessaires à leur manipulation sont implantées :

```
Cosinus between canonical scores and canonical variables
```

```
File AA_corr - Rows: 18 -Col: 2
```

```
Sort by block number [1,9] and score number [1,2]
```

```
To be used with labels in AA_corq.lab and the next files
```

```
-----  
| Description of a coded matrix |  
-----
```

```
Qualitative variables file: AA_corq
```

```
Number of rows: 18, variables: 2, categories: 11
```

```
Description of categories:
```

```
-----  
Variable number 1 has 9 categories
```

```
-----  
[ 1]Category: 3 Num: 2 Freq.: 0.1111
```

```
...
```

```
[ 9]Category: 3 Num: 2 Freq.: 0.1111
```

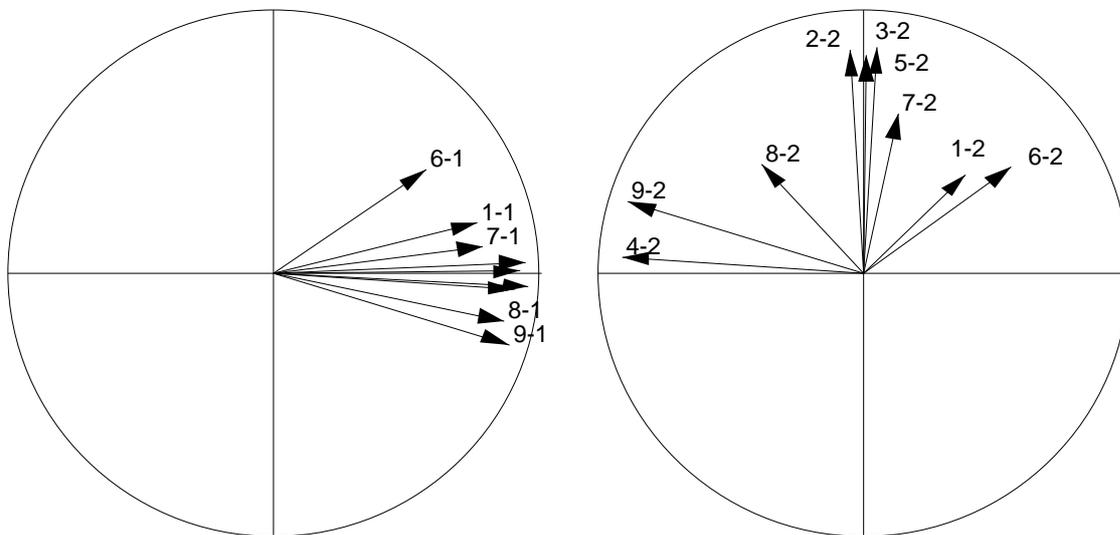
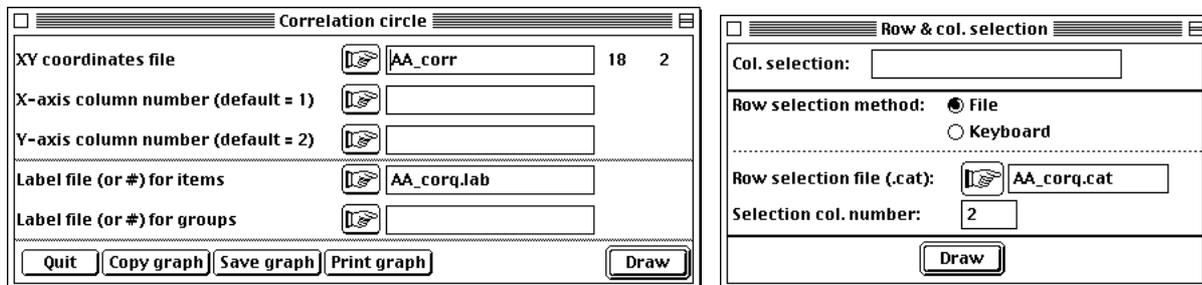
```
-----  
Variable number 2 has 2 categories
```

```
-----  
[ 10]Category: 3 Num: 9 Freq.: 0.5
```

[11]Category: 3 Num: 9 Freq.: 0.5

 Auxiliary ASCII output file AA_corq.123: labels (two characters) for 11 modalities
 It contains one label for each modality
 It has 11 rows (modalities) and labels 1,2, ..., 9

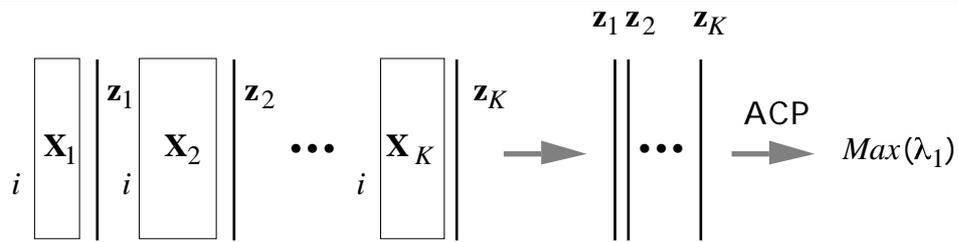
On peut donc représenter la projection sur le plan des scores canoniques des variables canoniques :



On voit sur cette figure un des défauts majeurs de l'analyse. La seconde variable canonique du tableau 4 est très liée au score canonique de rang 1 (ce qui n'a aucun sens). C'est pourtant normal, le tableau 4 est de rang 1. La seconde variable canonique ne peut être que la première (au signe près). La contrainte d'orthogonalité porte sur les scores canoniques (variables de synthèse) mais rien n'est imposé pour les variables canoniques qui peuvent être parfaitement redondantes.

Pour ne voir que le lien entre une composante canonique et les variables canoniques qui lui sont associées, le graphe canonique s'impose. Notons en outre la propriété fondamentale (Kettering 1971⁴ p. 435). Le score canonique de rang 1 est la première composante principale de l'ensemble des variables canoniques de rang 1, comme le score canonique de rang 2 est la première composante principale de l'ensemble des variables canoniques de rang 2. Les figures ci-dessus en rendent parfaitement compte.

La première valeur propre de l'analyse $\sum_{k=1}^K \text{corr}^2(\mathbf{z}, \mathbf{z}_k)$ est donc 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 variables normée par sous-espace dont l'ensemble donne une ACP de première valeur propre optimale :



Kettering (*op. cit.*) attribue ce principe à Vinograd (1950)⁵ et les premières approches numériques à Horst (1961)⁶. C'est Carroll qui a introduit les variables de synthèse (*auxiliary sample variates*) que nous avons appelé scores canoniques pour rendre compte des liaisons entre les variables canoniques (souvent appelées *canonical variates*).

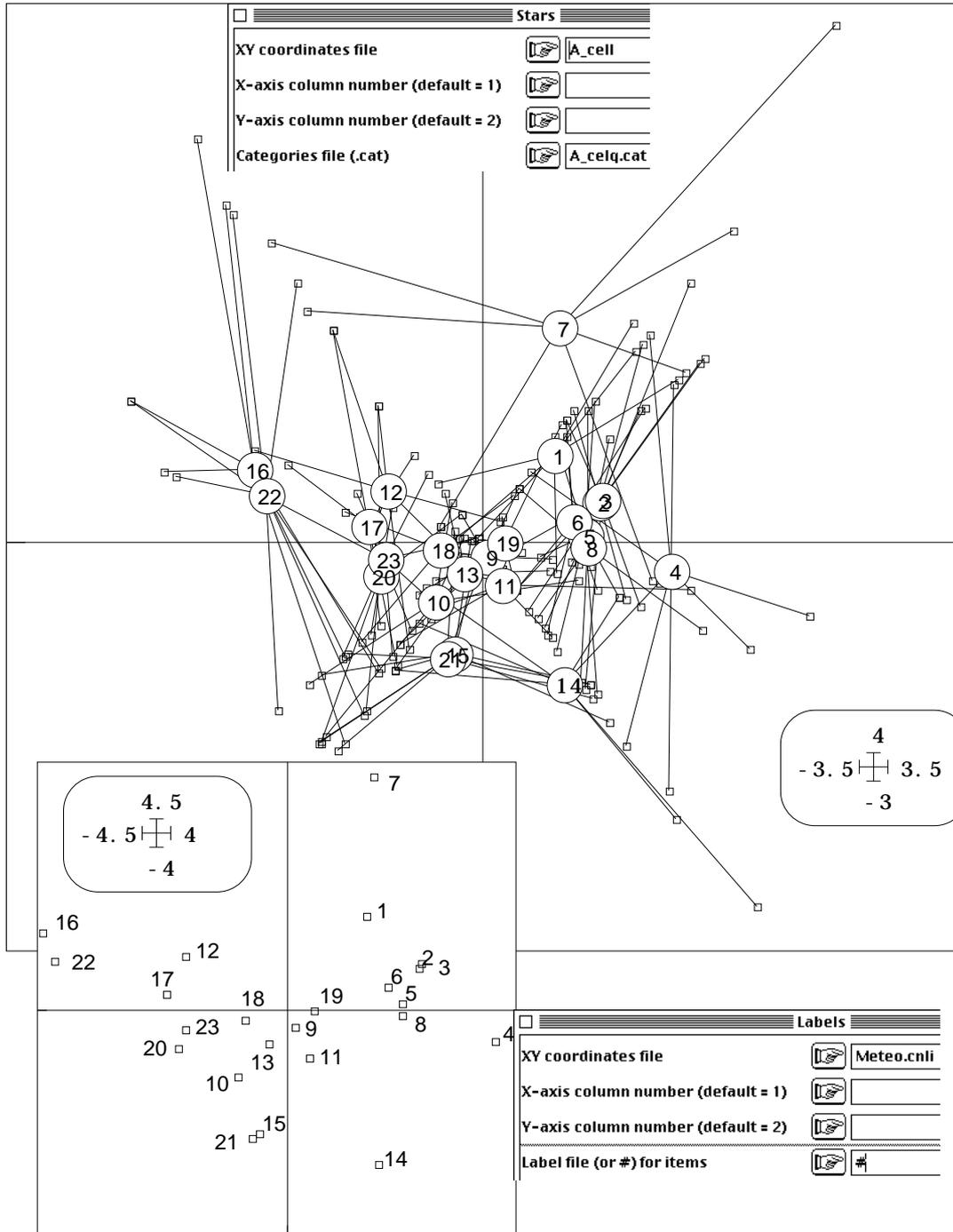


Figure 1 : Comparaison d'une ACP normée et d'une Analyse Canonique Généralisée avec une variable par tableau.

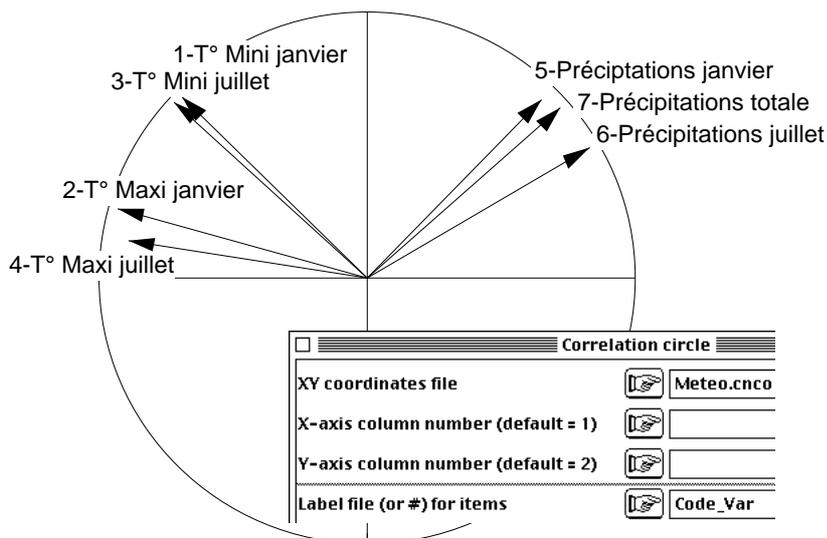
Ces remarques montrent que la matrice des corrélations entre variables canoniques de même rang peut servir à l'interprétation comme toute opérations liées à l'ACP normée du fichier XXX_cav1.

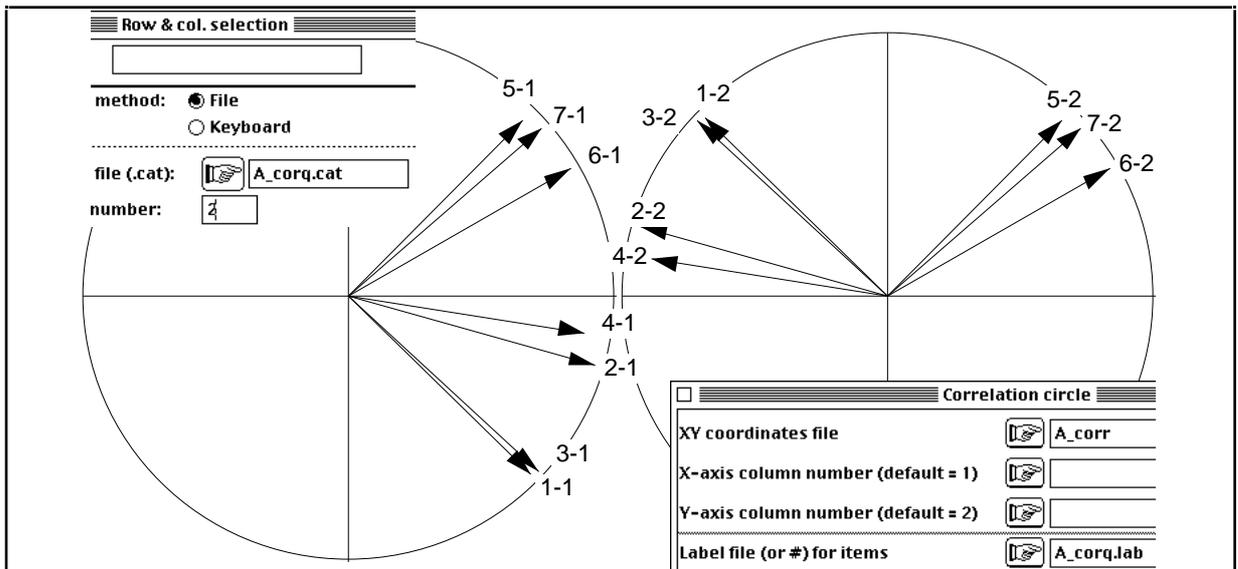
⚠ L'analyse canonique généralisée ne semble pas avoir été utilisée en écologie.

♻ L'ACP normée est une analyse canonique généralisée et comme telle apparaît sous un nouveau jour. Utiliser la carte Meteo. Normaliser les données :

Faire l'ACP normée :

Les deux cartes sont exactement les mêmes (figure 1). L'une des deux permet de positionner chaque valeur prise par chaque individu sur chacune des variables. Étonnant, non ? Ce qui signifie qu'on peut mesurer la cohérence de chacun des individus sur un ensemble de critères.





L'ACG fait ici de chaque variable un tableau. On s'en sert, en général, pour faire de chaque tableau une variable.

Toutes les illustrations confortent l'idée de Tenenhaus (1984 *op. cit.* p.83) selon laquelle cette méthode devrait plutôt s'appeler *analyse en composantes principales généralisée*.



- 1 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.
- 2 Tenenhaus, M. (1984) L'analyse canonique généralisée de variables numériques, nominales ou ordinales par des méthodes de codage optimal. In : *Data Analysis and Informatics*, III. Diday, E. & Coll. (Eds.) Elsevier Science Publishers B.V., North-Holland. 71-84.
- 3 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.
- 4 Kettering, R.J. (1971) Canonical analysis of several sets of variables. *Biometrika* : 58, 433-451.
- 5 Vinograd, B. (1950) Canonical positive definite matrices under internal linear transformations. *Proceedings of the American Mathematical Society* : 1, 159-161.
- 6 Horst, P. (1961) Relations among m sets of variables. *Psychometrika* : 26, 129-149.

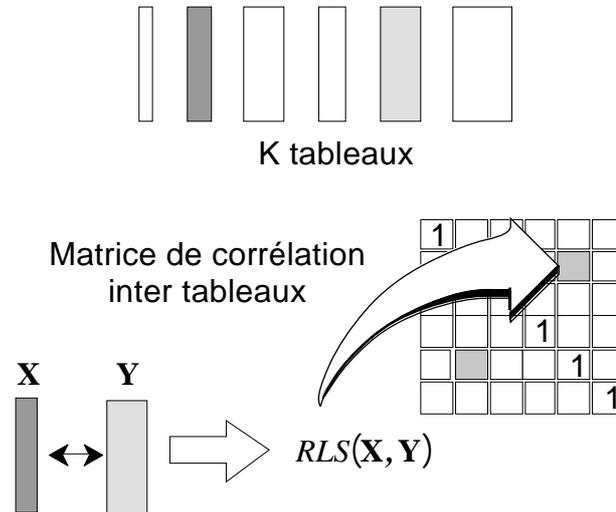
Canonical : RLS coefficients



Typologie de tableaux sur coefficients RLS.



Dans un K -tableaux, on cherche à définir ressemblance et dissemblances entre tableaux. Ici la ressemblance est mesurée par le coefficient RLS directement associé à l'analyse de co-inertie de deux tableaux. La différence est mesurée par la distance déduite de cette mesure de corrélation.



Le coefficient utilisé est défini par :

$$RLS(\mathbf{X}, \mathbf{Y}) = \frac{\text{Trace} \frac{1}{n} \mathbf{X}\mathbf{X}^t \frac{1}{n} \mathbf{Y}\mathbf{Y}^t}{\sqrt{\text{Trace} \frac{1}{n} \mathbf{X}\mathbf{X}^t \text{Trace} \frac{1}{n} \mathbf{Y}\mathbf{Y}^t}}$$

Il ne figure pas dans la revue de Lazraq et Cléroux (1988)¹, est attribué à Lingoes & Schönemann (1974)² par Lazraq & Coll. (1992)³ et étudié par Kiers & Coll. (1994)⁴. Il est directement connecté à l'analyse de co-inertie entre \mathbf{X} et \mathbf{Y} . On peut le réécrire sous la forme :

$$RLS(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{k=1}^r \text{Cov}(\mathbf{X}\mathbf{u}_k, \mathbf{Y}\mathbf{v}_k)}{\sqrt{\sum_{i=1}^{r_x} \lambda_i \frac{1}{n} \mathbf{X}^t \mathbf{X} \sum_{j=1}^{r_y} \lambda_j \frac{1}{n} \mathbf{Y}^t \mathbf{Y}}}$$

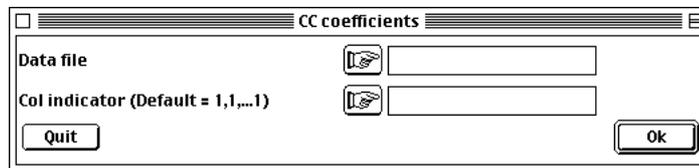
où $\text{Cov}(\mathbf{X}\mathbf{u}_k, \mathbf{Y}\mathbf{v}_k)$ est la covariance des coordonnées sur le couple d'axes principaux de co-inertie de rang k , c'est-à-dire la racine de la valeur propre de rang k de l'analyse de co-inertie. Il est directement lié au coefficient RV qui s'écrit :

$$RV(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{k=1}^r \text{Cov}^2(\mathbf{X}\mathbf{u}_k, \mathbf{Y}\mathbf{v}_k)}{\sqrt{\sum_{i=1}^{r_x} \lambda_i^2 \frac{1}{n} \mathbf{X}^t \mathbf{X} \sum_{j=1}^{r_y} \lambda_j^2 \frac{1}{n} \mathbf{Y}^t \mathbf{Y}}}$$

où $\lambda_m^2(\mathbf{A})$ désigne le carré de la valeur propre de \mathbf{A} de rang m .

RV est un produit scalaire, ce qui n'est pas vrai pour RLS mais la signification expérimentale de ce dernier ne fait pas de doute et c'est une mesure d'association entre tableaux très intéressante. RLS fait intervenir la notion de valeur absolue d'opérateurs symétriques qui est mathématiquement très solide (ce n'est pas un *bricolage* numérique !). Le carré de RLS intervient directement en Analyse canonique généralisée (Kiers & Coll *op. cit.*) et sa position est centrale.

 L'option utilise une seule fenêtre de dialogue :



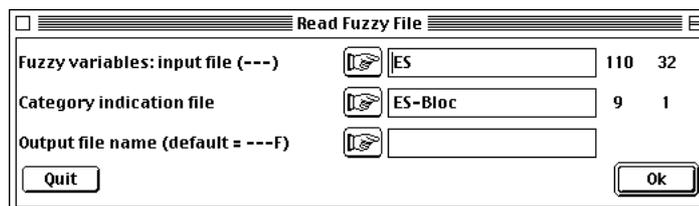
 Nom du fichier binaire d'entrée (n lignes-individus et p colonnes-variables). L'option ne fait aucun centrage préalable et s'utilise normalement sur un tableau centré ou normalisé *a priori*.

 Nom du fichier binaire indicateur de blocs de variables. Ce fichier doit être compatible avec le précédent. Il a une seule colonne et autant de lignes qu'il y a de tableaux (K). A la ligne k , on y trouve le nombre de variables (p_k) du tableau k . On doit avoir :

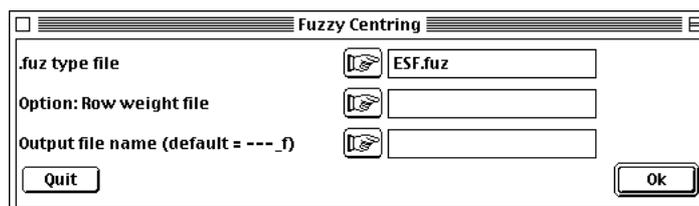
$$\sum_{k=1}^K p_k = p$$



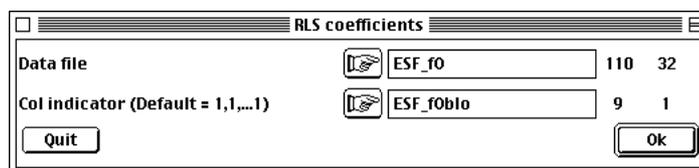
Utiliser la carte Coléoptères. Le tableau ES de variables floues est lu par FuzzyVar: Read Fuzzy File :



Centrer simplement le tableau par FuzzyVar: Fuzzy Centring :



Chaque variable floue définit un tableau centré par colonne.



```
Input file: ESF_f0
-> Rows: 110, columns: 32
-> 9 blocs: 4/5/4/2/4/3/5/3/2/
```

Les paramètres d'entrée sont rappelés. La matrice des coefficients RLS est calculée, éditée et stockée dans un fichier :

```
RLS coefficients
Lingoes & Schonemann 1974
```

Lazraq, Cleroux & Kiers 1992
 Kiers, Cleroux & Ten Berge 1994

```
----- Correlation matrix -----
[ 1] 1000
[ 2] 361 1000
[ 3] 432 812 1000
[ 4] 377 696 669 1000
[ 5] 325 743 732 691 1000
[ 6] 345 238 309 192 128 1000
[ 7] 348 378 421 329 371 817 1000
[ 8] 319 668 647 708 663 153 325 1000
[ 9] 301 762 709 812 650 172 334 798 1000
-----
```

ESF_fo_LS1 is a binary file with 9 rows and 9 columns
 Content: Coefficients RLS

La matrice qui précède contient des coefficients compris entre 0 et 1. Si K est le nombre de tableau, on obtient une matrice de distances :

$$\mathbf{R} = [\rho_{ij}]_{i \in \{1, \dots, K\}, j \in \{1, \dots, K\}} \quad \mathbf{D} = [d_{ij}]_{i \in \{1, \dots, K\}, j \in \{1, \dots, K\}}$$

avec $d_{ij} = \sqrt{2(1 - \rho_{ij})}$. Cette matrice est conservée dans un fichier avec le suffixe .dist qui permet d'utiliser une classification de tableaux. Cette matrice de distance est souvent euclidienne mais peut ne pas l'être (voir ci-dessous).

ESF_fo_LS.dist is a binary file with 9 rows and 9 columns
 Content: Among array distances = sqrt(2(1-r))

Quand on l'applique à deux tableaux contenant une seule variable, le coefficient RLS donne :

$$RLS(\mathbf{x}, \mathbf{y}) = |r(\mathbf{x}, \mathbf{y})|$$

C'est pourquoi on conserve aussi la matrice des RLS dans un fichier. Bien noter que pour les coefficient RV on a :

$$RV(\mathbf{x}, \mathbf{y}) = r^2(\mathbf{x}, \mathbf{y})$$

On a donc mis dans les fichiers _CC1, _LS1 et _RV1 des choses comparables à des corrélations (entre tableaux) et dans _CC2, _LS2 et _RV2 des choses comparables à des carrés de corrélations (entre tableaux).

Quand la matrice est positive, l'option en fait l'analyse en coordonnées principales :

RLS matrix is positive - Trace = 9.000e+00

```
PCO - RLS matrix diagonalization - Trace = 9.000e+00
Num Eigenval. | Num. Eigenval. | Num. Eigenval. | Num. Eigenval. |
001 5.136e+00 | 002 1.592e+00 | 003 7.317e-01 | 004 4.758e-01 |
005 3.284e-01 | 006 2.899e-01 | 007 1.862e-01 | 008 1.488e-01 |
009 1.106e-01 |
```

ESF_fo_RLS.divp is a binary file with 9 rows and 1 columns
 Content: Eigenvalues

ESF_fo_RLS.dico is a binary file with 9 rows and 6 columns
 Content: Array coordinates

```
File :ESF_fo_LS.dico
| Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | 4.036e-01 | 8.775e-01 |
| 2 | -8.596e-01 | 2.489e-01 |
| 3 | -2.366e-01 | 7.909e-01 |
| 4 | -3.071e-01 | 3.225e-01 |
| 5 | -2.275e-01 | 3.812e-01 |
| 6 | -3.557e-01 | 3.893e-01 |
|-----|-----|-----|
```

Le programme édite enfin les carrés de RLS :

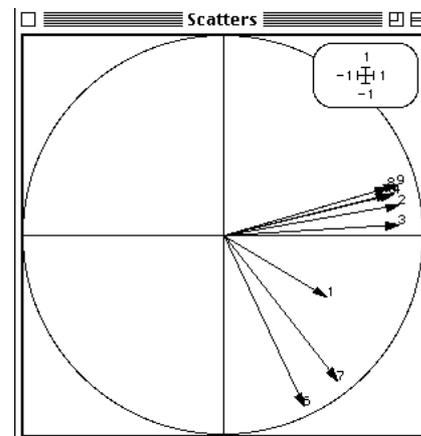
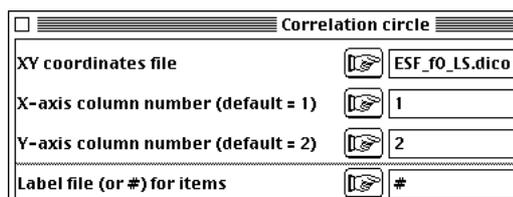
RLS squared

```
----- Correlation matrix -----
[ 1] 1000
[ 2] 130 1000
[ 3] 186 659 1000
[ 4] 142 484 447 1000
[ 5] 105 553 537 478 1000
[ 6] 119 57 96 37 16 1000
[ 7] 121 143 177 108 138 667 1000
[ 8] 101 446 418 502 439 24 106 1000
[ 9] 90 581 503 659 423 30 111 636 1000
```

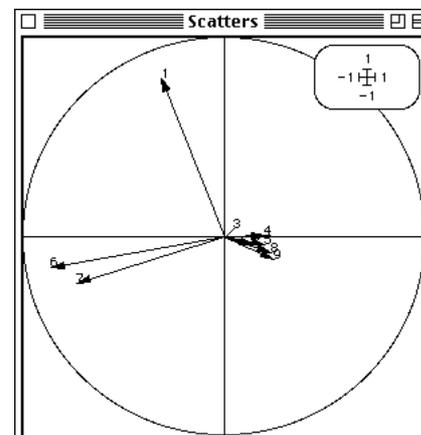
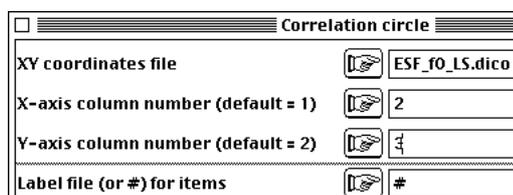
ESF_fO_LS2 is a binary file with 9 rows and 9 columns

Content: RLS squared

On a une propriété très particulière dans cette analyse. Comme la matrice RLS ne contient que des nombres positifs, la position de tous les tableaux se trouve d'un même côté du premier axe (théorème de Frobenius). Le premier axe est donc un axe de compromis du type ACP non centrée :



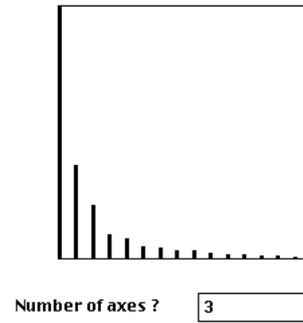
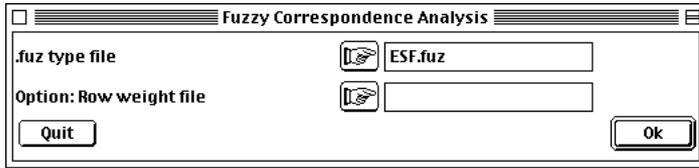
Par contre, à partir de l'axe 2, l'orthogonalité des axes introduit l'effet inverse :



Pour interpréter, en tenant compte des contraintes mathématiques, il faut donc voir le cône des vecteurs en combinant la lecture des deux figures.

Il y a trois types de structure, la première associée aux variables 2 (Sediment), 3 (Velocity), 4 (Pollution), 5 (Saprobity), 8 (Temperature) et 9 (Temp_Amplitude), la seconde définie par les variables 6 (Diet) et 7 (Feeding) et la troisième associée à la seule variable 1 (Vegetation).

Faire l'AFC floue pour savoir comment s'est exprimé ce fait très clair :



CorRatioFCA: Correlation ratios after a FCA
 Title of the analysis: ESF.fl
 Number of rows: 110, columns: 9

```

Variable : 1
> Categ= 1 Weight= 0.468 -0.248 -0.108 -0.132
> Categ= 2 Weight= 0.261 0.265 0.184 0.208
> Categ= 3 Weight= 0.152 -0.045 -0.061 0.074
> Categ= 4 Weight= 0.119 0.450 0.097 -0.032
-----> r= 0.072 0.016 0.020
Variable : 2
> Categ= 1 Weight= 0.069 -1.853 -0.300 -0.935
...
> Categ= 5 Weight= 0.587 0.676 -0.124 -0.297
-----> r= 0.814 0.051 0.385
Variable : 3
-----> r= 0.666 0.008 0.165
Variable : 4
-----> r= 0.789 0.027 0.012
Variable : 5
-----> r= 0.569 0.041 0.147
Variable : 6
-----> r= 0.170 0.804 0.022
Variable : 7
-----> r= 0.464 0.644 0.193
Variable : 8
> Categ= 1 Weight= 0.332 -0.340 0.097 -0.021
> Categ= 2 Weight= 0.502 0.019 0.023 0.024
> Categ= 3 Weight= 0.165 0.626 -0.264 -0.031
-----> r= 0.104 0.015 0.001
Variable : 9
> Categ= 1 Weight= 0.318 -1.320 0.334 0.061
> Categ= 2 Weight= 0.682 0.616 -0.156 -0.028
-----> r= 0.813 0.052 0.002
  
```

Le premier axe récupère l'association 2-3-4-5-9 (8 est un peu oubliée), le second associe 6 et 7, le troisième est peu interprétable. La variable 1 est ignorée. Ceci est fort logique. Comme cela se passe dans une analyse canonique généralisée ou une ACP normée, toute variable non liée aux autres ne fait pas partie des résultats.

C'est un fait très important. La variable isolée peut être une série aléatoire (ajoutée par facétie) ou un fait biologique majeur, l'analyse d'un tableau ne fait pas la différence. Tout ceux qui veulent choisir automatiquement le nombre d'axes font cela sans le savoir.

Ici, on voit clairement que le profil écologique des espèces a trois composantes (i) la position dans le gradient amont-aval enregistrée plusieurs fois (ii) le régime alimentaire encodée deux fois et (iii) le lien avec la végétation. Ces trois composantes sont largement indépendantes.



L'approche d'une variable biologique comme tableau est donc utile.



La matrice des coefficients RLS n'est pas nécessairement positive. Cette question a été posée à Henk A.L. Kiers (h.a.l.kiers@ppsw.rug.nl), Heymans Institute (PA), University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.

Sa réponse est très claire : *“Je crois que la matrice des coefficients RLS n'est pas forcément positive définie. Ce n'est pas facile de trouver une contre-exemple, surtout pas pour les cas des trois matrices de données, mais si on a quatre matrices (ou plus) j'en ai trouvé quelques contre-exemples, et facilement.*

Une exemple est :

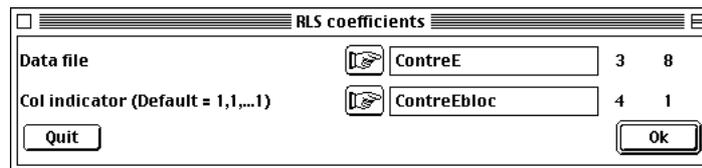
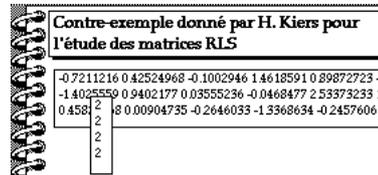
```
X1 =
-0.72112162110218    0.42524968296264
-1.40255586063406    0.94021769658085
 0.45827467861155    0.00904734521449

X2 =
-0.10029456518483    1.46185910109154
 0.03555235704854   -0.04684765238389
-0.26460333857548   -1.33686342786366

X3 =
 0.89872722739664   -0.53803581216303
 2.53373232882644    1.63716591549666
-0.24576058808020    0.13513860817495

X4 =
-0.12180822358350   -1.35337027524149
-0.04104612225877    1.28333265183618
-0.03411501939330   -0.17006391366464
```

On trouvera ce contre-exemple sur la carte RLS_CE :



```
Input file: ContreE
-> Rows: 3, columns: 8
-> 4 blocs: 2/2/2/2/
```

```
RLS coefficients
Lingoes & Schonemann 1974
Lazraq, Cleroux & Kiers 1992
Kiers, Cleroux & Ten Berge 1994
```

```
----- Correlation matrix -----
[ 1] 1000
[ 2] 449 1000
[ 3] 917 303 1000
[ 4] 305 487 607 1000
```

```
ContreE_LS1 is a binary file with 4 rows and 4 columns
Content: Coefficients RLS
```

```
ContreE_LS.dist is a binary file with 4 rows and 4 columns
Content: Among array distances = sqrt(2(1-r))
```

```
RLS matrix is NOT positive
Num Eigenval. | Num. Eigenval. | Num. Eigenval. | Num. Eigenval. |
001 2.565e+00 | 002 8.505e-01 | 003 6.038e-01 | 004 -1.909e-02 |
```

New name for the distance matrix : ContreE_LS_dist

La matrice de distance est dans ce cas conservée avec le suffixe `_dist` pour l'utilisation de Distances: Additive constante.

On peut essayer après centrage des tableaux :

Centring

Input file 3 8

Option: file for row weight

Option for X matrix (no default)

Output file

RLS coefficients

Data file 3 8

Col indicator (Default = 1,1,...1) 4 1

Le résultat est du même type :

Input file: ContreE0
 -> Rows: 3, columns: 8
 -> 4 blocs: 2/2/2/2/

RLS coefficients

Lingoes & Schonemann 1974
 Lazraq, Cleroux & Kiers 1992
 Kiers, Cleroux & Ten Berge 1994

----- Correlation matrix -----
 [1] 1000
 [2] 579 1000
 [3] 878 442 1000
 [4] 462 493 798 999

 ContreE0_LS1 is a binary file with 4 rows and 4 columns
 Content: Coefficients RLS

ContreE0_LS.dist is a binary file with 4 rows and 4 columns
 Content: Among array distances = sqrt(2(1-r))

RLS matrix is NOT positive
 Num Eigenval. | Num. Eigenval. | Num. Eigenval. | Num. Eigenval. |
 001 2.848e+00 | 002 6.302e-01 | 003 5.311e-01 | 004 -9.681e-03 |

New name for the distance matrix : ContreE0_LS_dist

La démonstration de H.A.L. Kiers est sans appel et limite considérablement la valeur théorique de la méthode. Nous le remercions de son intervention. L'option Canonical: RV coefficients est plus générale.



- 1 Lazraq, A. & Cléroux, R. (1988) Etude comparative de différentes mesures de liaison entre deux vecteurs aléatoires et tests d'indépendance. *Statistique et Analyse des données* : 13, 15-38.
- 2 Lingoes, J.C. & Schönemann, P.H. (1974) Alternative measures of fit for the Schönemann-Carrol matrix fitting algorithm. *Psychometrika* : 39, 423-427.
- 3 Lazraq, A., Cléroux, R. & Kiers, H.A.L. (1992) Mesures de liaison vectorielle et généralisation de l'analyse canonique. *Revue de Statistique Appliquée* : 39, 23-35.
- 4 Kiers, H.A.L., Cléroux, R. & Ten Berge, M.F. (1994) Generalized analysis based on optimizing matrix correlations and a relation with IDIOSCAL. *Computational Statistics and Data Analysis* : 18, 331-340.

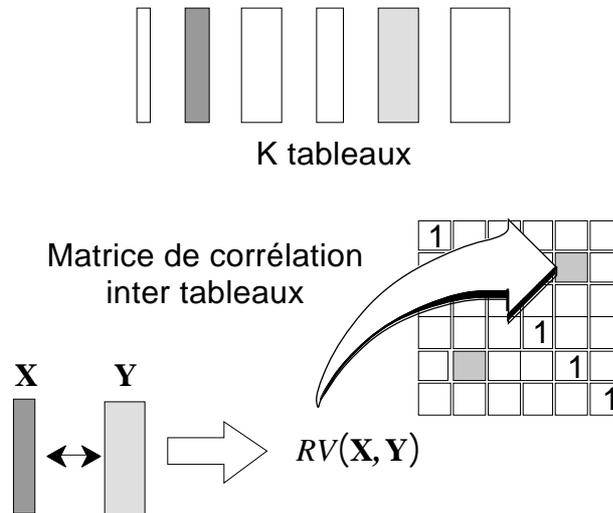
Canonical : RV coefficients



Typologie de tableaux sur coefficients RV.



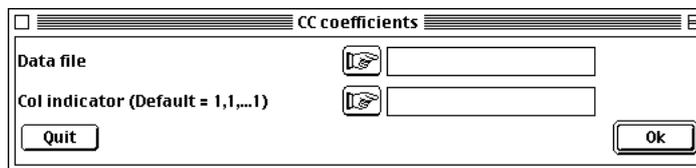
Dans un K -tableaux, on cherche à définir ressemblance et dissemblances entre tableaux. Ici la ressemblance est mesurée par le coefficient RV de la méthode ACT-STATIS.



La différence est mesurée par la distance déduite de cette mesure de corrélation. cette option est un programme ACT-STATIS allégé et centré sur l'interstructure.



L'option utilise une seule fenêtre de dialogue :



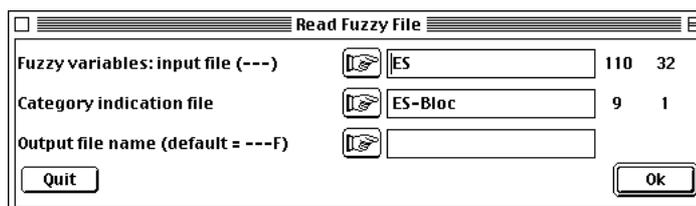
Nom du fichier binaire d'entrée (n lignes-individus et p colonnes-variables). L'option ne fait aucun centrage préalable et s'utilise normalement sur un tableau centré ou normalisé *a priori*.

Nom du fichier binaire indicateur de blocs de variables. Ce fichier doit être compatible avec le précédent. Il a une seule colonne et autant de lignes qu'il y a de tableaux (K). A la ligne k , on y trouve le nombre de variables (p_k) du tableau k . On doit avoir :

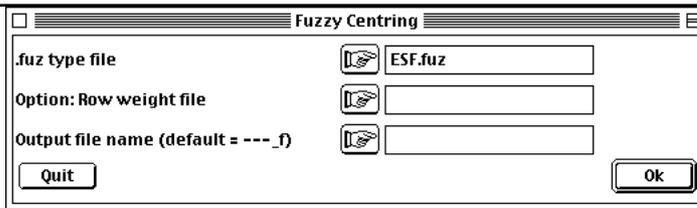
$$\sum_{k=1}^K p_k = p$$



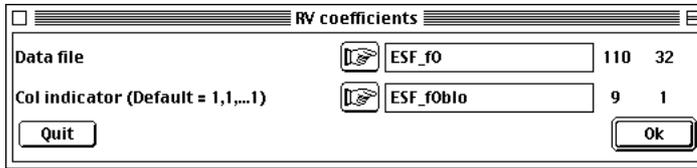
Utiliser la carte Coléoptères. Le tableau ES de variables floues est lu par FuzzyVar: Read Fuzzy File :



Centrer simplement le tableau par FuzzyVar: Fuzzy Centring :



Chaque variable floue définit un tableau centré par colonne.



Input file: ESF_fo
 -> Rows: 110, columns: 32
 -> 9 blocks: 4/5/4/2/4/3/5/3/2/
 RV coefficients Escoufier 1973

```
----- Correlation matrix -----
[ 1] 1000
[ 2] 159 1000
[ 3] 241 755 1000
[ 4] 199 604 514 1000
[ 5] 128 653 583 604 1000
[ 6] 153 60 108 38 20 1000
[ 7] 200 170 226 152 154 894 1000
[ 8] 108 612 498 612 555 17 128 1000
[ 9] 127 725 579 659 534 31 157 775 1000
```

ESF_fo_RV2 is a binary file with 9 rows and 9 columns
 Content: Coefficients RV

La matrice des coefficients RV¹ est stockée dans un fichier _RV2 parce que, sur un couple de tableau contenant une seule variable chacun on a :

$$RV(\mathbf{x}, \mathbf{y}) = r^2(\mathbf{x}, \mathbf{y})$$

Les coefficients RV se lisent comme des carrés de corrélation. On peut comparer des carrés de *RLS* et des *RV* ou des *RLS* et des \sqrt{RV} . L'analyse exécute l'interstructure de STATIS sur les opérateurs normés (voir par exemple ²). La matrice des coefficients *RV* est diagonalisée. Comme matrice de produits scalaires elle est symétrique et positive :

```
Interstructure - RV matrix diagonalization - Trace = 9.000e+00
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +4.2199E+00  +0.4689  +0.4689  | 02  +1.8833E+00  +0.2093  +0.6781  |
03  +9.2813E-01  +0.1031  +0.7813  | 04  +6.1917E-01  +0.0688  +0.8501  |
05  +4.9192E-01  +0.0547  +0.9047  | 06  +3.5628E-01  +0.0396  +0.9443  |
07  +2.3797E-01  +0.0264  +0.9707  | 08  +1.7120E-01  +0.0190  +0.9898  |
09  +9.2168E-02  +0.0102  +1.0000
```

Cette diagonalisation a deux significations qu'il vaut mieux bien comprendre pour s'en servir.



D'une part, les composantes du **premier vecteur propre** sont toutes de même signe (théorème de Frobenius). Elles sont systématiquement calculées avec le signe + et définissent des poids a_k , dits poids des tableaux dans le compromis. Si \mathbf{u} est le premier vecteur propre normé de la matrice des *RV* et λ_1 la première valeur propre :

$$\|\mathbf{u}\|^2 = \sum_{k=1}^K u_k^2 = 1$$

On a retenu les poids $a_k = u_k$.

Sum of $a_k = 2.757e+00$

On peut prendre :

$$a_k = u_k \sqrt{\lambda_1} \quad \text{ou} \quad b_k = a_k \Big/ \prod_{k=1}^K a_k$$

ou un autre ensemble de paramètres proportionnels aux u_k , mais cela ne change rien à la suite et n'a qu'une importance secondaire. Les RV sont des produits scalaires entre opérateurs d'inertie :

$$RV(\mathbf{X}, \mathbf{Y}) = \frac{\text{Trace} \frac{1}{n} \mathbf{X}\mathbf{X}^t \frac{1}{n} \mathbf{Y}\mathbf{Y}^t}{\sqrt{\text{Trace} \frac{1}{n} \mathbf{X}\mathbf{X}^t \frac{1}{n} \mathbf{X}\mathbf{X}^t \quad \text{Trace} \frac{1}{n} \mathbf{Y}\mathbf{Y}^t \frac{1}{n} \mathbf{Y}\mathbf{Y}^t}}$$

Ici, le tableau k définit l'opérateur :

$$\mathbf{W}_k \mathbf{D} = \frac{1}{n} \mathbf{X}_k \mathbf{X}_k^t$$

dont la norme :

$$\|\mathbf{W}_k \mathbf{D}\| = \sqrt{\text{Trace} \frac{1}{n} \mathbf{X}_k \mathbf{X}_k^t \frac{1}{n} \mathbf{X}_k \mathbf{X}_k^t} = \sqrt{\sum_{j=1}^{r_k} \lambda_j^2}$$

Vérifier qu'en sélectionnant les quatre premières colonnes de ESF_fO dans un fichier 110-4 et en faisant l'ACP centrée (inutilement) de ce fichier on obtient 3 valeurs propres (rang 3) qui valent 0.067362, 0.026064 et 0.00847. En faire la somme des carrés et en prendre la racine. On trouve 0.07272 éditée ci-dessous.

Le premier vecteur propre donne les poids a_k qui permettent de combiner les opérateurs dans un compromis :

$$\mathbf{W}\mathbf{D} = \sum_{k=1}^K a_k \mathbf{W}_k \mathbf{D}$$

qui se caractérise par sa norme optimale :

$$\|\mathbf{W}\mathbf{D}\| = \sqrt{\text{Trace}(\mathbf{W}\mathbf{D}\mathbf{W}\mathbf{D})} = \sqrt{\sum_{j=1}^s \lambda_j^2(\mathbf{W}\mathbf{D})}$$

Cet optimum est la racine de la première valeur propre de la matrice des RV :

Consensus (Normed W_k) : Norm HS = 2.054e+00

La programme édite alors une vue synthétique de la position des tableaux par rapport au compromis, à savoir le nombre de colonnes de chaque tableau :

Cols = column number of each table

Le poids de chacun des opérateurs dans le calcul du compromis :

Weights = Weights of operators in the consensus

La norme de chacun des opérateurs :

HS norm = Norm (Hilbert-Schmidt) of operators

Le coefficient RV de chaque opérateur avec le compromis :

RVcons = RV (Wk, Sum akWk) 1/1000

la racine carré du RV qui se lit comme un coefficient de corrélation :

RV1/2 = sqrt(RV (Wk, Sum akWk)) 1/1000

la distance de chaque opérateur avec le compromis :

Dist2 = squared distance = 2(1-RVcons)

Number	Cols	Weights	HS norm	RVcons	RV1/2	Dist2
1	4	1.323e-01	7.272e-02	272	521	1.456e+00
2	5	4.251e-01	2.762e-01	873	934	2.536e-01
3	4	3.889e-01	2.242e-01	799	894	4.021e-01
4	2	3.887e-01	3.681e-01	798	894	4.032e-01
5	4	3.789e-01	1.788e-01	778	882	4.435e-01
6	3	8.282e-02	3.972e-01	170	412	1.660e+00
7	5	1.533e-01	2.707e-01	315	561	1.370e+00
8	3	3.910e-01	3.405e-02	803	896	3.938e-01
9	2	4.156e-01	4.339e-01	854	924	2.924e-01



D'autre part, les composantes de **tous les vecteurs propres** analyse une matrice de distances. La matrice des *RV* contient des coefficients compris entre 0 et 1 qui sont des produits scalaires. Si *K* est le nombre de tableau, on obtient une matrice de distances :

$$\mathbf{R} = [\rho_{ij}]_{i \in K, j \in K} \quad \mathbf{D} = [d_{ij}]_{i \in K, j \in K}$$

avec $d_{ij} = \sqrt{2(1 - \rho_{ij})}$. Cette matrice est conservée dans un fichier avec le suffixe .dist qui permet d'utiliser une classification de tableaux. Cette matrice de distance est euclidienne puisque que celle des *RV* est positive (théorème 6 dans ³), ou plus simplement parce qu'une matrice de distance est euclidienne s'il existe une représentation euclidienne qui rend compte de cette distance. Or cette représentation euclidienne est construite en amont du problème par l'introduction des opérateurs et de leur produit scalaire d'Hilbert-Schmidt.

ESF_fo_RV.dist is a binary file with 9 rows and 9 columns
Content: Among array distances = sqrt(2(1-RV))
Euclidean distance

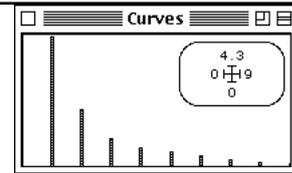
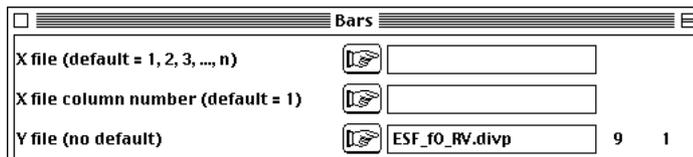
PCO analysis on array distances
ESF_fo_RV.divp is a binary file with 9 rows and 1 columns
Content: Eigenvalues

ESF_fo_RV.dico is a binary file with 9 rows and 6 columns
Content: Array coordinates

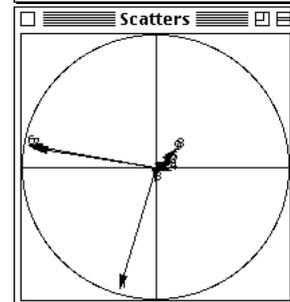
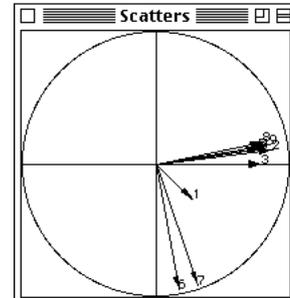
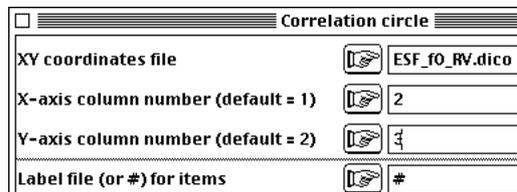
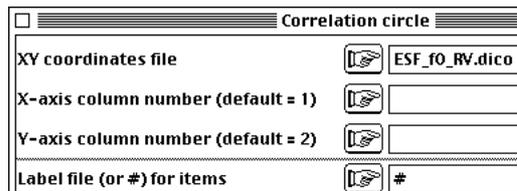
File :ESF_fo_RV.dico

Col.	Mini	Maxi
1	1.701e-01	8.732e-01
2	-9.465e-01	1.632e-01
3	-9.102e-01	1.685e-01
4	-4.428e-01	3.875e-01
5	-2.414e-01	4.627e-01
6	-4.164e-01	2.951e-01

Le programme garde au plus 6 coordonnées de la représentation euclidienne.



L'utilisation de cette représentation euclidienne suit les mêmes lois que celle de Canonical: RLS coefficients (voir p. 26) :

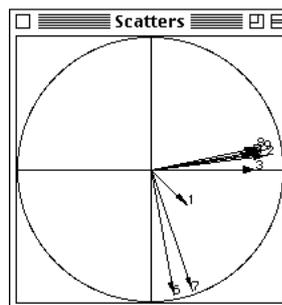


Bien voir le cône des variables : le résultat est celui obtenu p. 26 en plus net encore, vraisemblablement du fait qu'avec *RLS* on travaille avec des corrélations et qu'avec *RV* on travaille avec des carrés de corrélations.

On retiendra que STATIS permet la mise en évidence de plusieurs compromis dans un multi-tableau. Ici il serait logique d'étudier séparément 2 compromis et un tableau indépendant.



Bien noter que sur la figure :



le vecteur unitaire horizontal qui permet (en partie) de positionner les tableaux est le compromis. Les résultats du tableaux de synthèse ne font qu'illustrer les propriétés de la figure : le tableau 6 est le plus éloigné du compromis par exemple. Ce compromis étant un opérateur, il définit lui-même une structure. Le listing utilise le terme de consensus retenu par Kiers & Coll. *op. cit.* Le passage de l'objet consensus vu comme un vecteur à la nature de cet objet vu comme typologie de point est une difficulté conceptuelle non négligeable. Le compromis est analysé :

Consensus analysis
 ESF_fo_RV.vp is a binary file with 110 rows and 1 columns
 Content: Eigenvalues

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.0039E+00	+0.6214	+0.6214	02	+2.9614E-01	+0.0918	+0.7133

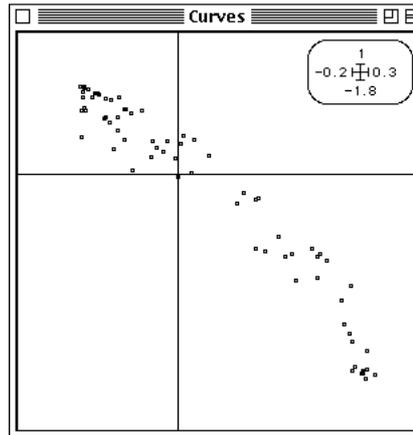
```

03 +2.4846E-01 +0.0771 +0.7903 | 04 +1.4657E-01 +0.0455 +0.8358 |
...
21 +1.0865E-03 +0.0003 +0.9999 | 22 +2.8267E-04 +0.0001 +1.0000 |
23 +2.0513E-05 +0.0000 +1.0000 | 24 +0.0000E+00 +0.0000 +1.0000 |
...
39 +0.0000E+00 +0.0000 +1.0000 | 40 +0.0000E+00 +0.0000 +1.0000 |
ESF_fO_RV.li is a binary file with 110 rows and 3 columns
Content: Row coordinates

```

File :ESF_fO_RV.li

Col.	Mini	Maxi
1	-1.204e-01	2.452e-01
2	-1.199e-01	8.847e-02
3	-1.248e-01	1.291e-01



Lines

X file (default = 1, 2, 3, ..., n)

X file column number (default = 1)

Y file (no default)

Cumulated data (1=yes, 2=no)

Variable label file (or #)

Draw curves (1=yes, 2=no)

On retrouve une forte communauté entre le premier score de l'AFC floue et la première coordonnée (fondamentale) du compromis. L'AFC floue est efficace pour décrire le compromis : on rajoute ici le moyen rapide et précis d'étude de la pertinence de ce compromis et de l'éventuelle nécessité d'en faire plusieurs. La matrice des \sqrt{RV} est enfin conservée.

sqrt(RV)

```

----- Correlation matrix -----
[ 1] 1000
[ 2] 399 1000
[ 3] 491 869 1000
[ 4] 447 777 717 1000
[ 5] 358 808 763 777 1000
[ 6] 391 245 328 196 143 1000
[ 7] 447 412 476 390 393 945 1000
[ 8] 329 782 705 782 745 131 358 1000
[ 9] 356 852 761 812 731 175 396 880 1000

```

ESF_fO_RV1 is a binary file with 9 rows and 9 columns
Content: sqrt(Coefficients RV)



Le programme a été vérifié sur le jeu de données illustrant l'article de fond de Lavit & Coll. (1994)⁴. Les données sont disponibles sur la carte Statis94 :

Données de démonstration de ACT (STATIS method) Lavit & Coll 1994		Card
13 18 8 1 12 2 20 10 10 2	10 2 14 12 18 15 18 10 4 10 0 18 14 17 7 12 2	
12 10 4	Architecture	12 16 12 5 4 8 10 8 10 6 5 7 5 10 7 12 8 8 15
4 10 6	Theory	20 10 10 12 5 6 5 3 10 3 2 3 2 4 3 5 2 0 0 2 8 0 6
15 12	Languages	16 14 12 10 18 10 13 12 12 12 11 11 7 6 7 4
16 18	Economics	13 13 12 14 16 14 13 12 15 15 14 12 14 14 16 16
12 12	Accounting	+ 10 lignes et 55 colonnes
10 17	Management	seurs définis par les étiquettes
17 17	Inf Sys Design	es répartis en 8 blocs par l'indi
11 8 7	Statistics	loc est un étudiant qui note les
17 16	Op-Research	Compétence
	English	Explanation
		Pedagogy
		Cheerfulness
		Dynamic
		Accessible
		Interesting
		1_Comet
		1_Explan
		1_Pedago
		1_Cheerf
		1_Dynami
		1_Access
		1_Intere
		2_Comet
		2_Explan
		2_Pedago

Centrer simplement le tableau de données :

Centring

Input file 10 55

Option: file for row weight

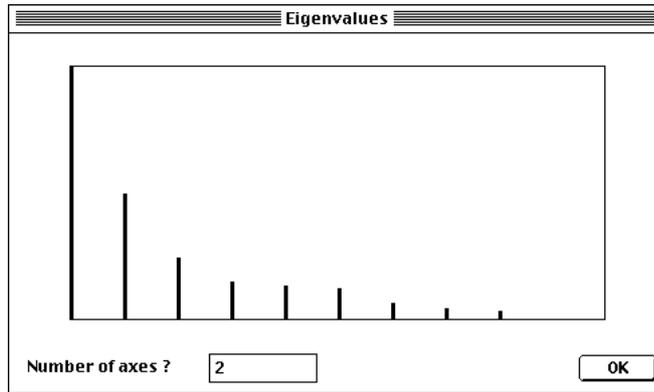
Option for X matrix (no default)

Output file

RV coefficients

Data file 10 55

Col indicator (Default = 1,1,...,1) 8 1



Input file: Tab0
 -> Rows: 10, columns: 55
 -> 8 blocks: 7/7/7/6/7/7/7/7/

RV coefficients Escoufier 1973

----- Correlation matrix -----

```
[ 1] 1000
[ 2] 173 1000
[ 3] 449 282 1000
[ 4] 426 304 421 1000
[ 5] 691 220 528 542 1000
[ 6] 481 161 539 631 760 1000
[ 7] 627 182 485 412 868 728 1000
[ 8] 417 108 306 226 290 191 398 1000
```

Le résultat est conforme au tableau 2 p. 111.

Interstructure - RV matrix diagonalization - Trace = 8.000e+00

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+4.1751E+00	+0.5219	+0.5219	02	+1.0054E+00	+0.1257	+0.6476
03	+9.2040E-01	+0.1150	+0.7626	04	+6.0202E-01	+0.0753	+0.8379
05	+5.6900E-01	+0.0711	+0.9090	06	+4.4442E-01	+0.0556	+0.9645
07	+1.8283E-01	+0.0229	+0.9874	08	+1.0080E-01	+0.0126	+1.0000

Sum of ak = 2.734e+00

Consensus (Normed Wk) : Norm HS = 2.043e+00

Number	Cols	Weights	HS norm	RVcons	RV1/2	Dist2
1	7	3.758e-01	1.190e+02	768	876	4.641e-01
2	7	1.672e-01	1.510e+02	342	585	1.317e+00
3	7	3.433e-01	8.678e+01	701	838	5.970e-01
4	6	3.405e-01	7.797e+01	696	834	6.087e-01
5	7	4.436e-01	1.005e+02	906	952	1.873e-01
6	7	4.086e-01	1.170e+02	835	914	3.302e-01
7	7	4.244e-01	1.369e+02	867	931	2.658e-01
8	7	2.306e-01	1.020e+02	471	686	1.057e+00

Les carrés des distances sont ceux de la page 109. L'article donne pour les poids :

0.18 0.08 0.17 0.17 0.22 0.20 0.21 0.11. Pour retrouver ces valeurs, il suffit de diviser les poids édités par $\sqrt{\lambda_1} = \sqrt{4.1751} = 2.043$

```
Interstructure - RV matrix diagonalization - Trace = 8.000e+00
Num. Eigenval.  R.Iner.  R.Sum  Num. Eigenval.  R.Iner.  R.Sum
01 +4.1751E+00 +0.5219 +0.5219 02 +1.0054E+00 +0.1257 +0.6476
03 +9.2040E-01 +0.1150 +0.7626 04 +6.0202E-01 +0.0753 +0.8379
05 +5.6900E-01 +0.0711 +0.9090 06 +4.4442E-01 +0.0556 +0.9645
07 +1.8283E-01 +0.0229 +0.9874 08 +1.0080E-01 +0.0126 +1.0000
```

Par exemple $0.3758/2.043 = 0.1839 = 0.18$ ou $0.2306/2.043 = 0.1129 = 0.11$.

Tab0_RV.dist is a binary file with 8 rows and 8 columns
Content: Among array distances = $\text{sqrt}(2(1-RV))$
Euclidean distance

On retrouve le contenu dans le tableau 3 ($1.2863 * 1.2863 = 1.6546 = 1.65$) :

	1	2	3	4	5	6	7	8
1	0.0000	1.2863	1.0501	1.0712	0.7858	1.0191	0.8633	1.0802
2	1.2863	0.0000	1.1986	1.1795	1.2487	1.2954	1.2788	1.3357
3	1.0501	1.1986	0.0000	1.0762	0.9713	0.9602	1.0147	1.1781
4	1.0712	1.1795	1.0762	0.0000	0.9570	0.8593	1.0840	1.2441
5	0.7858	1.2487	0.9713	0.9570	0.0000	0.6931	0.5134	1.1915
6	1.0191	1.2954	0.9602	0.8593	0.6931	0.0000	0.7379	1.2717
7	0.8633	1.2788	1.0147	1.0840	0.5134	0.7379	0.0000	1.0976
8	1.0802	1.3357	1.1781	1.2441	1.1915	1.2717	1.0976	0.0000

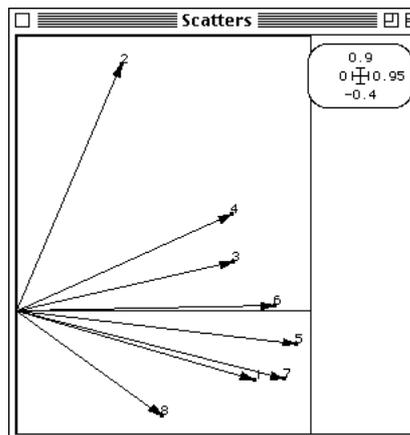
PCO analysis on array distances

Tab0_RV.divp is a binary file with 8 rows and 1 columns
Content: Eigenvalues

Tab0_RV.dico is a binary file with 8 rows and 6 columns
Content: Array coordinates

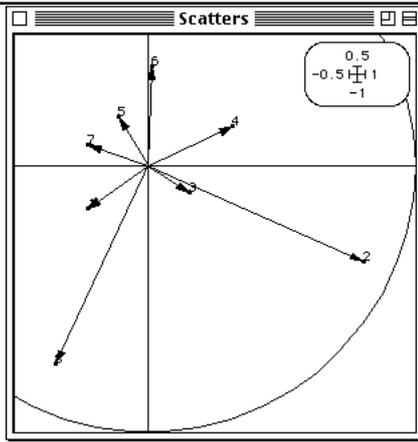
File :Tab0_RV.dico

Col.	Mini	Maxi
1	3.417e-01	9.063e-01
2	-3.407e-01	8.081e-01
...		
6	-2.446e-01	5.099e-01



On retrouve (à une symétrie près) la figure 2 p. 111, sur laquelle il y a une faute de frappe sur le report des valeurs propres (17.4 pour 4.17). Le tracé du plan 2-3 renforce l'interprétation (les juges 2 et 8 ne participent pas au compromis et de plus ne sont pas cohérents) :

XY coordinates file	<input type="text" value="Tab0_RV.dico"/>	8	6
X-axis column number (default = 1)	<input type="text" value="2"/>		
Y-axis column number (default = 2)	<input type="text" value="3"/>		
Label file (or #) for items	<input type="text" value="#"/>		



Consensus analysis

Tab0_RV.vp is a binary file with 10 rows and 1 columns

Content: Eigenvalues

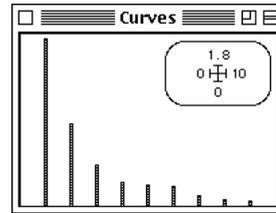
Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.7450E+00	+0.4360	+0.4360	02	+8.6896E-01	+0.2171	+0.6532
03	+4.3034E-01	+0.1075	+0.7607	04	+2.5948E-01	+0.0648	+0.8255
05	+2.3354E-01	+0.0584	+0.8839	06	+2.1281E-01	+0.0532	+0.9371
07	+1.1296E-01	+0.0282	+0.9653	08	+7.8653E-02	+0.0197	+0.9850
09	+6.0218E-02	+0.0150	+1.0000	10	+0.0000E+00	+0.0000	+1.0000

Bars

X file (default = 1, 2, 3, ..., n)

X file column number (default = 1)

Y file (no default)



Tab0_RV.li is a binary file with 10 rows and 2 columns

Content: Row coordinates

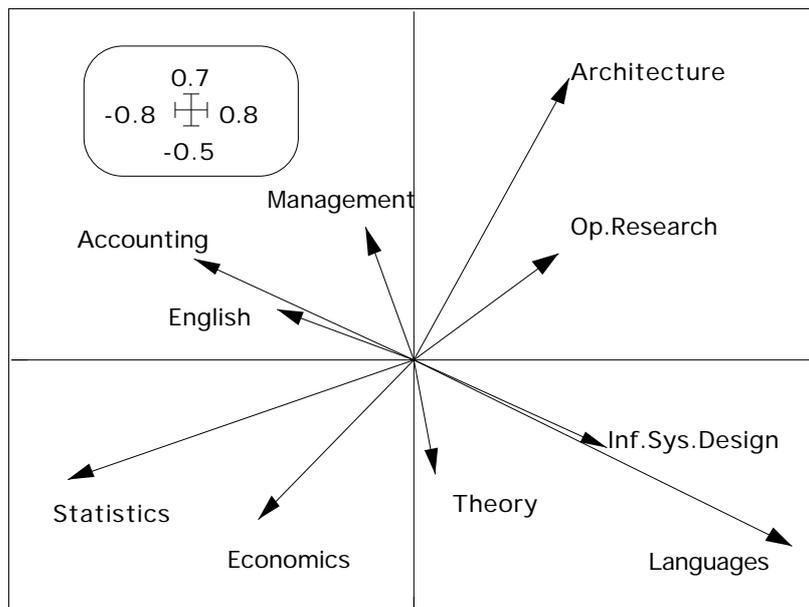
File :Tab0_RV.li

Col.	Mini	Maxi
1	-6.862e-01	7.569e-01
2	-3.728e-01	5.627e-01

Labels

1)

2)



On retrouve (à une double symétrie près) la figure 3 p. 112. Les paramètres affichés sont proportionnels au carrés des valeurs propres du présent programme. Pour les trajectoires, utiliser STATIS: Operator averaging.



- 1 Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.
- 2 Dazy, F. & Le Barzic, J.F. (1996) *L'analyse des données évolutives*. Technip, Paris. 227 p.
- 3 Gower, J.C. & Legendre, P. (1986) Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* : 3, 5-48.
- 4 Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis* : 18, 97-119.

Canonical : Test Sum_RV



Test de permutations.



Dans de rares cas, on peut se demander si les tableaux d'un K -tableaux sont complètement indépendants. Dans une telle circonstance, tous les coefficients RV entre tableaux sont faibles et il en est de même de leur somme. Le test porte sur l'hypothèse nulle : l'observation est tirée au hasard dans l'espace équiprobabilisé des $(n!)^K$ permutations indépendantes des lignes de chacun des tableaux. La variable aléatoire supportant le test est :

$$SRV = \sum_{j,k=1}^K RV(\mathbf{X}_j, \mathbf{X}_k)$$

L'alternative peut s'énoncer par : il existe au moins un couple de tableaux dont le RV est non nul.



L'option utilise une seule fenêtre de dialogue :

Nom du fichier binaire d'entrée (n lignes-individus et p colonnes-variables). L'option ne fait aucun centrage préalable et s'utilise normalement sur un tableau centré ou normalisé *a priori*.

Nom du fichier binaire indicateur de blocs de variables. Ce fichier doit être compatible avec le précédent. Il a une seule colonne et autant de lignes qu'il y a de tableaux (K). A la ligne k , on y trouve le nombre de variables (p_k) du tableau k . On doit avoir :

$$\sum_{k=1}^K p_k = p$$

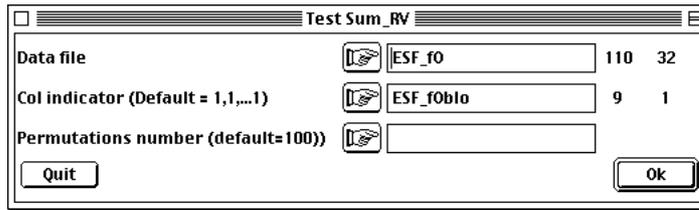
Nombre de permutations utilisées. Attention, le calcul est volumineux. A chaque simulation, chacun des tableaux subit une permutation aléatoire de ses lignes et la matrice des RV est recalculée.



Utiliser la carte Coléoptères. Le tableau ES de variables floues est lu par FuzzyVar: Read Fuzzy File :

Centrer simplement le tableau par FuzzyVar: Fuzzy Centring :

Chaque variable floue définit un tableau centré par colonne.

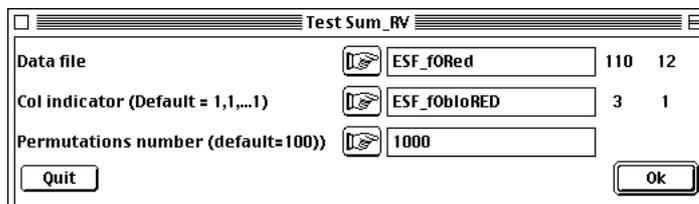
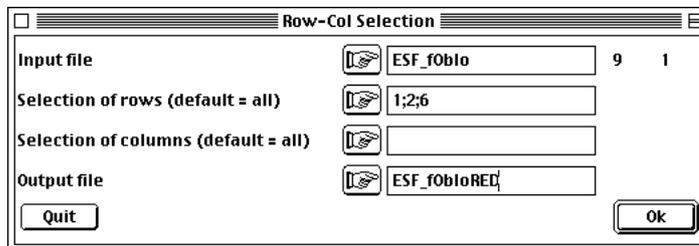
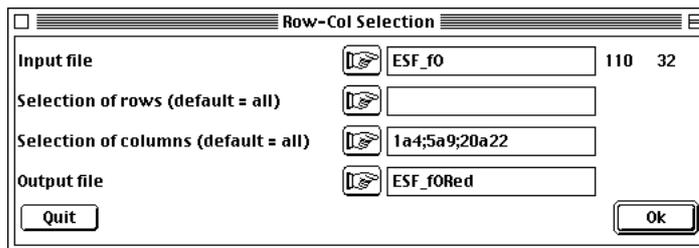


Input file: ESF_f0
-> Rows: 110, columns: 32
-> 9 blocs: 4/5/4/2/4/3/5/3/2/
Permutation number: 100

number of random permutations: 100 Observed: 1.273e+01
Histogram: minimum = 3.150e-01, maximum = 1.273e+01
number of simulations X<Obs: 100 (frequency: 1.000e+00)
number of simulations X>=Obs: 0 (frequency: 0.000e+00)

•->

L'hypothèse nulle est caricaturellement fausse. Extraire du tableau les variables 1, 2 et 6.



```

Input file: ESF_fORed
  -> Rows: 110, columns: 12
  -> 3 blocs: 4/5/3/
Permutation number: 1000

number of random permutations: 1000   Observed: 3.719e-01
Histogram: minimum = 7.511e-03, maximum = 3.719e-01
number of simulations X<Obs: 1000 (frequency: 1.000e+00)
number of simulations X>=Obs: 0 (frequency: 0.000e+00)

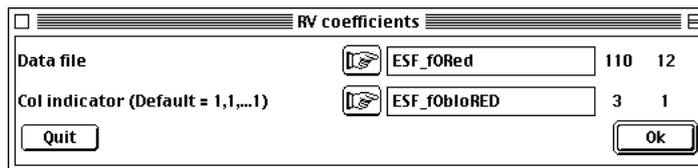
```

```

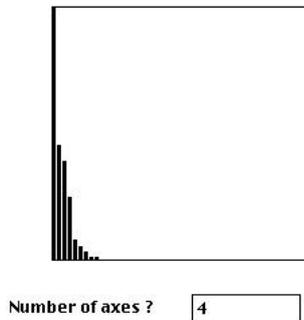
*****
*****
*****
*****
****
*

```

•-->



On ne s'attend pas à un compromis aussi complexe :



```

Input file: ESF_fORed
  -> Rows: 110, columns: 12
  -> 3 blocs: 4/5/3/
RV coefficients Escoufier 1973

```

```

----- Correlation matrix -----
[ 1] 1000
[ 2] 159 1000
[ 3] 153 60 1000

```

ESF_fORed_RV2 is a binary file with 3 rows and 3 columns
Content: Coefficients RV

Cette matrice de RV est bien extraite de la précédente (voir Canonical: RV coefficients).

```

Interstructure - RV matrix diagonalization - Trace = 3.000e+00
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +1.2526E+00  +0.4175  +0.4175  | 02  +9.4012E-01  +0.3134  +0.7309  |
03  +8.0726E-01  +0.2691  +1.0000

```

Sum of ak = 1.723e+00
 Consensus (Normed Wk) : Norm HS = 1.119e+00
 Cols = column number of each table
 Weights = Weights of operators in the consensus
 HS norm = Norm (Hilbert-Schmidt) of operators
 RVcons = RV (Wk, Sum akWk) 1/1000
 RV1/2 = sqrt(RV (Wk, Sum akWk)) 1/1000
 Dist2 = squared distance = 2(1-RVcons)

Number	Cols	Weights	NS norm	RVcons	RV1/2	Dist2
1	4	6.578e-01	7.272e-02	736	858	5.276e-01
2	5	5.397e-01	2.762e-01	604	777	7.919e-01
3	3	5.254e-01	3.972e-01	588	767	8.240e-01

On ne s'attendait pas à trouver un consensus en sélectionnant dans l'ensemble un tableau aussi différents que possible des deux autres.

ESF_fOred_RV.dist is a binary file with 3 rows and 3 columns
 Content: Among array distances = sqrt(2(1-RV))
 Euclidean distance

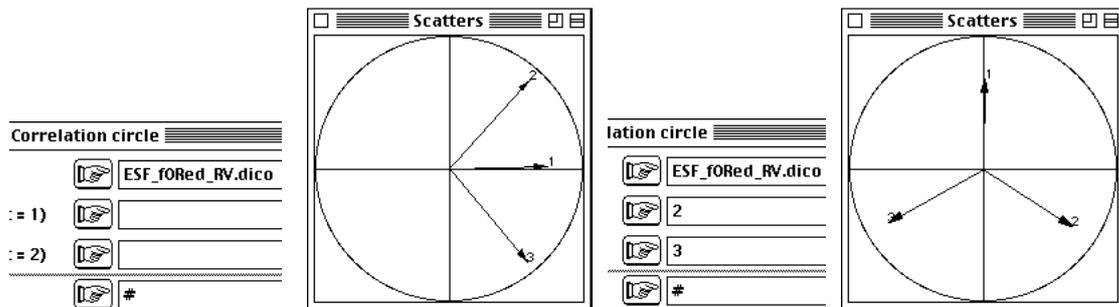
PCO analysis on array distances
 ESF_fOred_RV.divp is a binary file with 3 rows and 1 columns
 Content: Eigenvalues

ESF_fOred_RV.dico is a binary file with 3 rows and 3 columns
 Content: Array coordinates

File :ESF_fOred_RV.dico

Col.	Mini	Maxi
1	5.880e-01	7.362e-01
2	-7.029e-01	6.677e-01
3	-4.350e-01	6.766e-01

Le résultat est fort étonnant :



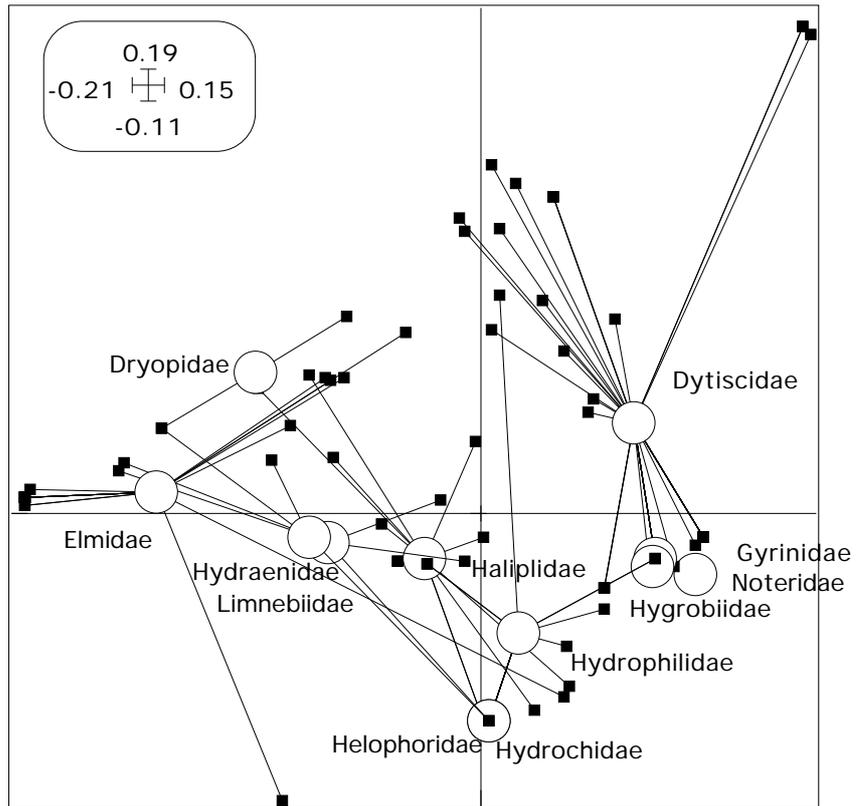
Les vecteurs représentant les opérateurs sont les arêtes d'une pyramide dont la base est un triangle équilatéral.

Consensus analysis
 ESF_fOred_RV.vp is a binary file with 110 rows and 1 columns
 Content: Eigenvalues

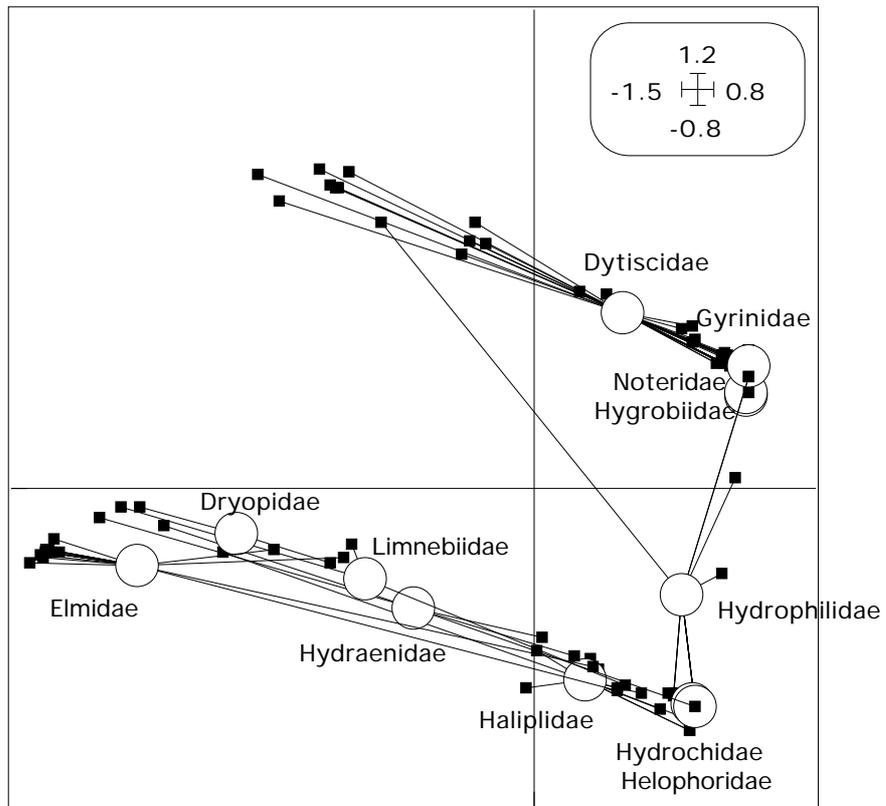
Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+9.3445E-01	+0.4362	+0.4362	02	+4.2720E-01	+0.1994	+0.6356
03	+3.6445E-01	+0.1701	+0.8058	04	+2.3346E-01	+0.1090	+0.9148
05	+7.5714E-02	+0.0353	+0.9501	06	+5.0769E-02	+0.0237	+0.9738
07	+3.1054E-02	+0.0145	+0.9883	08	+1.4716E-02	+0.0069	+0.9952
09	+1.0351E-02	+0.0048	+1.0000	10	+0.0000E+00	+0.0000	+1.0000
11	+0.0000E+00	+0.0000	+1.0000	12	+0.0000E+00	+0.0000	+1.0000
...							
37	+0.0000E+00	+0.0000	+1.0000	38	+0.0000E+00	+0.0000	+1.0000
39	+0.0000E+00	+0.0000	+1.0000	40	+0.0000E+00	+0.0000	+1.0000

ESF_fOred_RV.li is a binary file with 110 rows and 4 columns

Le fichier ESF_fORed_RV.li (plan 1-2) donne :



On le compare au plan 1-2 de l'AFC floue du tableau complet :



Noter le rotation associée à l'élimination de la redondance sur le gradient principal. La ressemblance plus forte entre espèces d'une même famille rend significative la somme des coefficients RV bien que l'essentiel des redites entre tableaux ait été enlevé.

```
sqrt(RV)
```

```
----- Correlation matrix -----
```

```
[ 1] 1000  
[ 2] 399 1000  
[ 3] 391 245 1000
```

```
-----  
ESF_fOred_RV1 is a binary file with 3 rows and 3 columns  
Content: sqrt(Coefficients RV)
```

Le résultat est remarquable. Bien que nous ayons enlevé 8 des 11 variables, la typologie des espèces est globalement conservée. De ceci on retiendra que la structure phylogénétique sous-jacente aux données biologiques implique automatiquement une part de corrélation entre structures et que l'analyse du compromis additionne des structures plus qu'elle n'en fait la moyenne.

La signification des méthodes multi-tableaux dans le champ de l'application est encore largement inconnue.

