

Colnertia

Colnertia : Coinertia analysis.....	2
Colnertia : Coinertia test - Fixed D.....	11
Colnertia : Coinertia test - Fixed Tab 1.....	17
Colnertia : Coinertia test - Fixed Tab 2.....	20
Colnertia : Fully Matched Tables.....	22
Colnertia : Matching two statistical triplets.....	26

Colnertia : Coinertia analysis



Méthode d'analyse de données couplant deux tableaux dite analyse de co-inertie.



L'objectif est de décrire la structure commune de deux tableaux portant sur les mêmes unités statistiques. La méthode est la version générale de l'analyse inter-batterie de Tucker ¹, l'analyse canonique sur variables qualitatives de Cazes ², l'analyse des correspondances sur tableaux de profils écologiques de Romane ³. Les justificatifs théoriques et bibliographiques sont dans ⁴. L'introduction en hydrobiologie est faite dans ⁵, en phyto-écologie dans ⁶, en nématologie dans ⁷, en chiomiométrie dans ⁸, en biologie moléculaire dans ⁹. On trouve des exemples d'utilisation dans ¹⁰ et la généralisation à K tableaux dans ¹¹ (KTA-MFA : Multiple CO-inertia Analysis).



L'option utilise une seule fenêtre de dialogue :

Coinertia analysis	
Matrix input file	---

Nom du fichier d'entrée de type ---.<>ta créé par CoInertia : Matching two statistical triplets.



Utiliser la carte Doubs ¹² de la pile ADE-4•Data. Ajouter les fichiers créés par la carte Doubs+1. Exécuter l'ACP normée du tableaux de variables de milieu (PCA : Correlation matrix PCA) :

Correlation matrix PCA	
Matrix input file	DouMil 30 11

Exécuter l'ACP centrée du tableau faunistique (PCA : Covariance matrix PCA) :

Covariance matrix PCA	
Matrix input file	DouPoi 30 27

Associer les deux analyses (CoInertia : Matching two statistical triplets) :

Matching two statistical triplets	
First input file	DouMil.cnta 30 11
Second input file	DouPoi.cpta 30 27
Output file name	W

```
First input statistical triplet: table DouMil.cnta
Number of rows: 30, columns: 11
Second input statistical triplet: table DouPoi.cpta
Number of rows: 30, columns: 27
```

```
File W.<>ta contains the cross table YtDnX
It has 27 rows and 11 columns
```

```
File W.<>pl contains the weight of each row
It has 27 rows and 1 column
```

```
File W.<>pc contains the weight of each column
It has 11 rows and 1 column
```

Bien noter, à cet endroit, l'ordre des tableaux dans le couple (l'option met en premier le tableau de rang le plus petit) :

```
File W.<>ma contains the name of the input statistical triplets
1 ---> DouMil.cnta
2 ---> DouPoi.cpta
```

```
total inertia: 134.703
```

Exécuter la présente option :



La première partie du programme est constituée de l'analyse du tableau croisé. Le premier triplet s'écrit $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$. \mathbf{X} a n lignes et p colonnes. Le second triplet s'écrit $(\mathbf{Y}, \mathbf{D}_q, \mathbf{D}_n)$. \mathbf{Y} a n lignes et q colonnes. Le triplet croisé s'écrit $(\mathbf{Y}^t \mathbf{D}_n \mathbf{X}, \mathbf{D}_p, \mathbf{D}_q)$. Ses lignes sont les colonnes de \mathbf{Y} et ses colonnes sont les colonnes de \mathbf{X} . La matrice traitée contient ici les covariances entre l'abondance d'un taxon et une variable de milieu normalisée. On obtient une analyse d'inertie standard d'une matrice croisée :

DiagoRC: General program for two diagonal inner product analysis

Input file: W.<ta

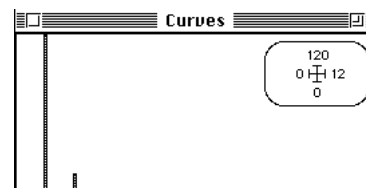
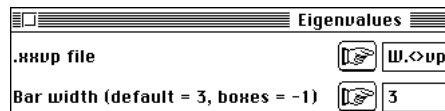
--- Number of rows: 27, columns: 11

Total inertia: 134.703

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.1902E+02	+0.8836	+0.8836	02	+1.3871E+01	+0.1030	+0.9865
03	+7.5658E-01	+0.0056	+0.9922	04	+5.2783E-01	+0.0039	+0.9961
05	+2.7089E-01	+0.0020	+0.9981	06	+1.6462E-01	+0.0012	+0.9993
07	+6.6011E-02	+0.0005	+0.9998	08	+1.8209E-02	+0.0001	+0.9999
09	+4.6041E-03	+0.0000	+1.0000	10	+2.7934E-03	+0.0000	+1.0000
11	+0.0000E+00	+0.0000	+1.0000				

File W.<vp contains the eigenvalues and relative inertia for each axis

--- It has 11 rows and 2 columns



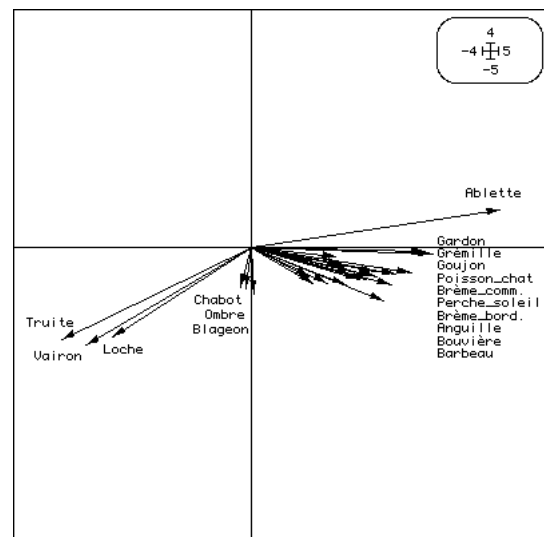
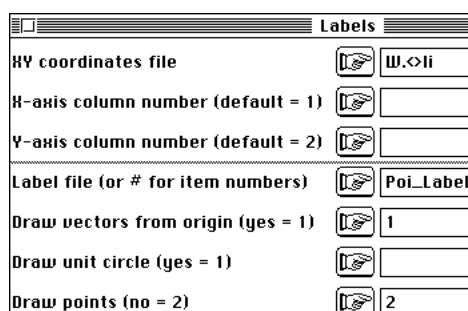
Les analyses de co-inertie fournissent souvent des graphes de valeurs propres très explicites car celles-ci sont des produits d'indices de structures de chaque tableau par un indice de co-structure. La dimension 2 s'impose ici.

File W.<co contains the column scores

--- It has 11 rows and 2 columns

File :W.<co

Col.	Mini	Maxi
1	-4.491e+00	5.376e+00
2	-1.621e+00	1.994e+00



File W.<>li contains the row scores
 --- It has 27 rows and 2 columns
 File :W.<>li

Col.	Mini	Maxi
1	-3.202e+00	4.180e+00
2	-1.647e+00	6.392e-01

Labels

XY coordinates file

X-axis column number (default = 1)

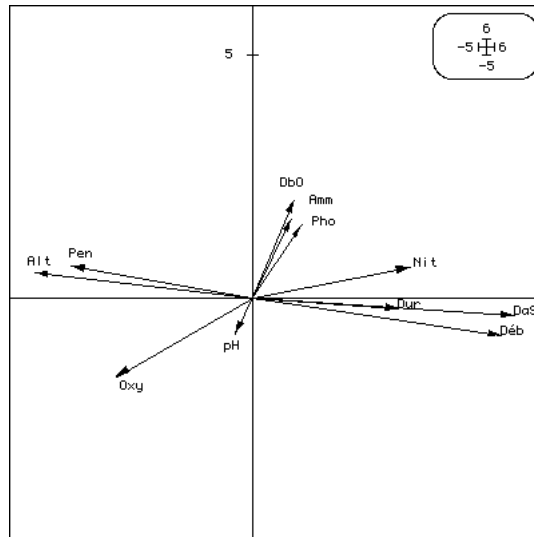
Y-axis column number (default = 2)

Label file (or # for item numbers)

Draw vectors from origin (yes = 1)

Draw unit circle (yes = 1)

Draw points (no = 2)



On obtient une carte des espèces et une carte des variables liées entre elles. On reconnaît que ces cartes donnent des informations sur la structure de chaque tableau et le lien qui les unit. Le tableau croisé a une signification qui dépend entièrement des analyses de base (ici deux ACP). Ici, ce sont des covariances. On peut éditer le tableau croisé (Tables : Values) et le légendier (Graph1D : Labels) :

Values

Input table file 27 11

X-axis position file 11 2

Column number (default = 1)

Y-axis position file 27 2

Labels

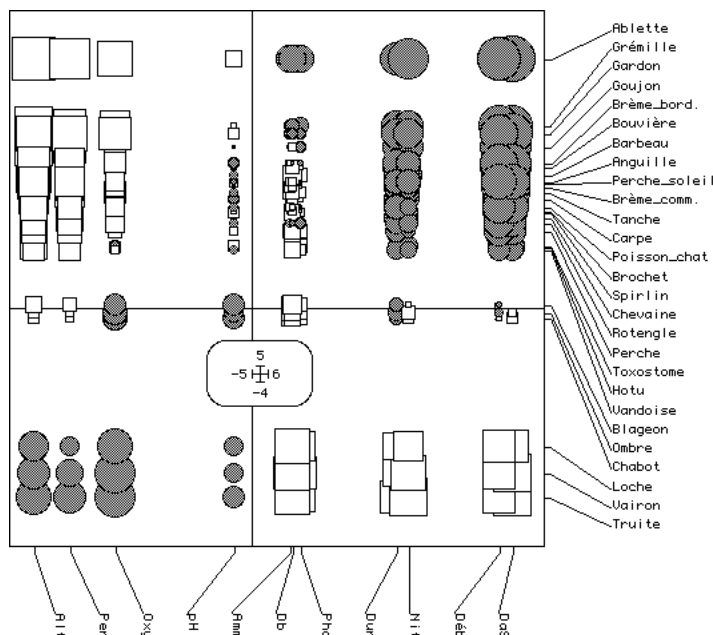
Data file (no default)

Rows label file (default = #)

Labels

Data file (no default)

Rows label file (default = #)



Tous les utilitaires complémentaires d'analyse d'inertie de DDUtil sont ici disponibles. Le listing donne ensuite les aides à l'interprétation spécifique. Il convient de les utiliser de manière variée en fonction des analyses de bases.

```
-----
Co-inertia analysis between two statistical triplets
  1 ---> DouMil.cnta (rows: 30, col: 11, axes: 3, inertia: 11.000000)
  2 ---> DouPoi.cpta (rows: 30, col: 27, axes: 3, inertia: 66.077797)
Co-inertia: 134.7, RV coefficient: 0.45056
```

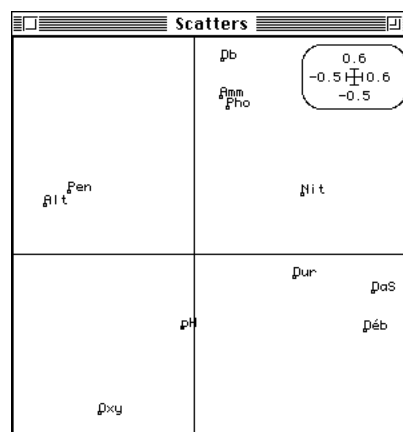
Le coefficient RV se comprend dans la théorie des analyses K-tableaux. En première lecture, il peut être ignoré (voir la fiche STATIS).

W.<>w1 is a binary file with 11 rows and 2 columns
It contains the canonical weights of variables of table 1

```
File :W.<>w1
|Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | -4.117e-01 | 4.928e-01 |
| 2 | -4.352e-01 | 5.353e-01 |
|-----|-----|-----|
```

Les coefficients de W.<>w1 sont des poids canoniques en cas d'ACP initiale. Ce sont des coefficients de somme de carrés unité définissant les combinaisons linéaires optimales des variables du tableau 1 (variable de milieu). Ce sont les coordonnées de variance 1 donnée par DDUtil : Add normed scores. Ce sont aussi les composantes des axes de co-inertie du tableau 1 :

Labels	
XY coordinates file	<input type="button" value="W.<>w1"/>
X-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="button" value="VarMil_label"/>



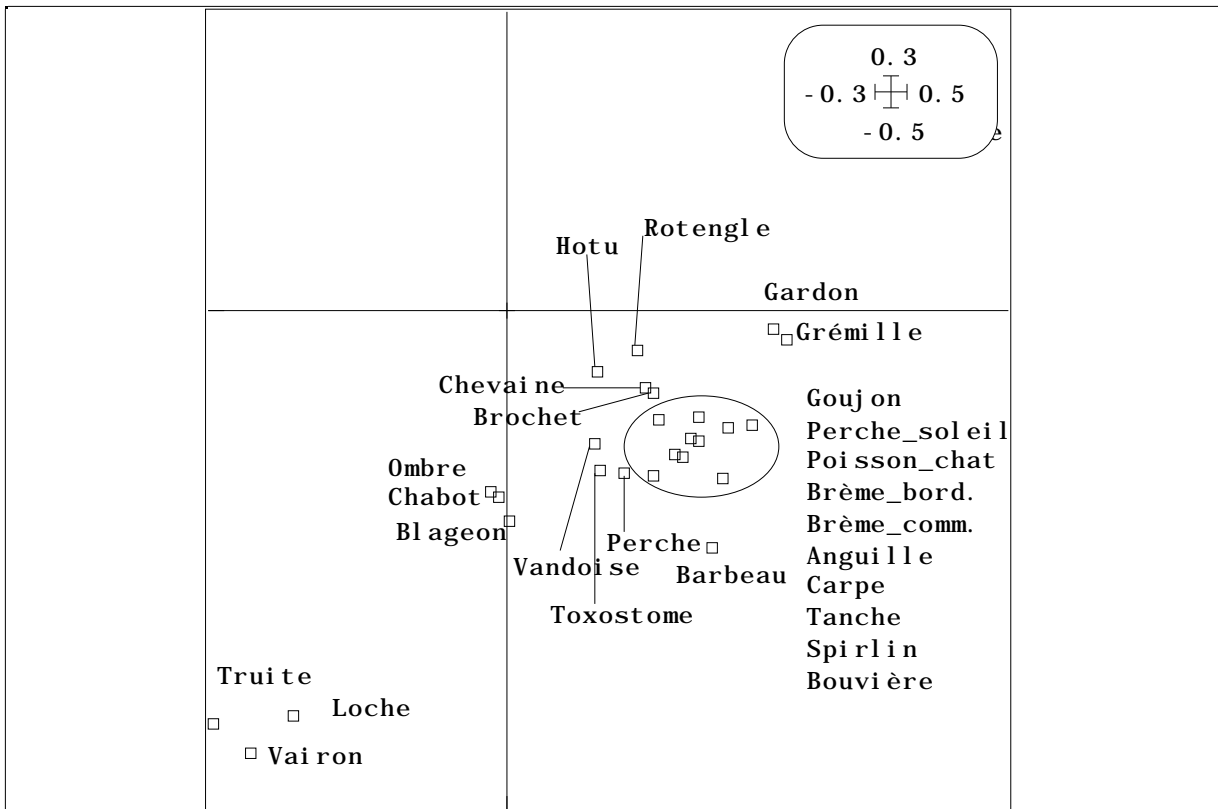
L'axe 1 est le gradient amont-aval et l'axe 2 est le gradient de pollution.

W.<>w2 is a binary file with 27 rows and 2 columns
It contains the canonical weights of the variables of table 2

```
File :W.<>w2
|Col. | Mini | Maxi |
|-----|-----|-----|
| 1 | -2.935e-01 | 3.832e-01 |
| 2 | -4.421e-01 | 1.716e-01 |
|-----|-----|-----|
```

Les coefficients de W.<>w2 sont des poids canoniques en cas d'ACP initiale. Ce sont des coefficients de somme de carrés unité définissant les combinaisons linéaires optimales des variables du tableau 2 (abondances spécifiques). Ce sont les coordonnées de variance 1 donnée par DDUtil : Add normed scores. Ce sont aussi les composantes des axes de co-inertie du tableau 2 :

Labels	
XY coordinates file	<input type="button" value="W.<>w2"/> 27 2
X-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Label file (or # for item numbers)	<input type="button" value="Poi_Label"/>

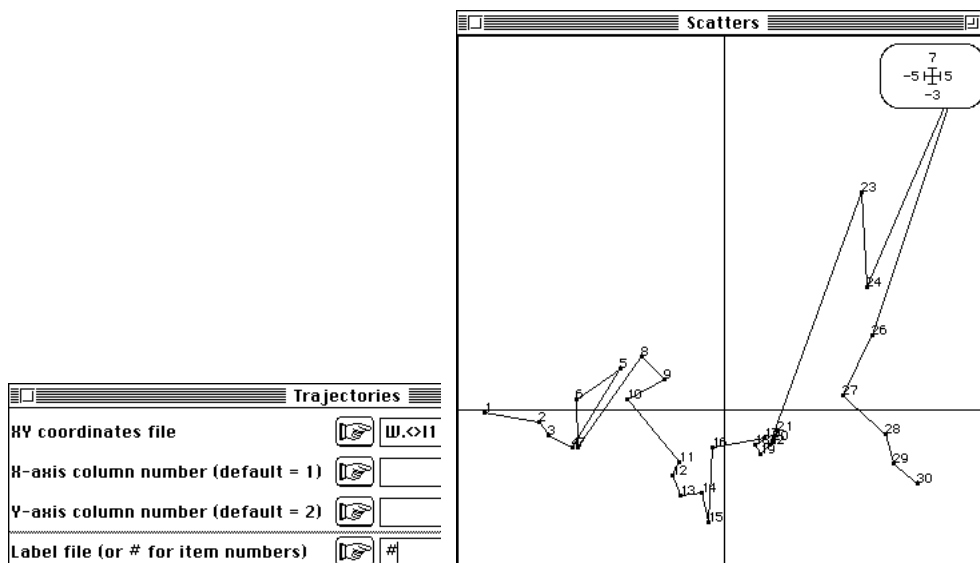


On reconnaît l'ordination standard (Truite/Ombre/Barbeau/Brème) et la bascule opérée par le rôle de la pollution d'autant plus contraignante qu'elle intervient plus haut ou d'autant plus forte qu'elle intervient plus bas.

W.<>11 is a binary file with 30 rows and 2 columns
It contains the coordinates of the rows (table 1)
File :W.<>11

Col.	Mini	Maxi
1	-4.479e+00	4.429e+00
2	-2.114e+00	6.342e+00

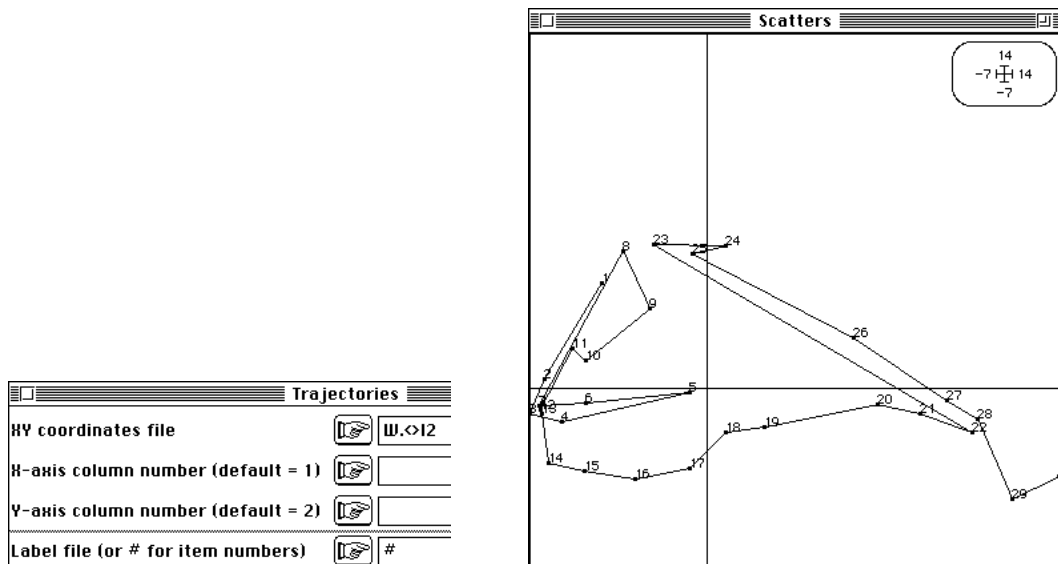
Les poids canoniques donnent des combinaisons de variables qui sont dans le fichier W.<>11. Les valeurs sont des coordonnées des projections des relevés sur les axes de co-inertie (espace des variables). On obtient une véritable carte factorielle des relevés (milieu) qu'on peut voir par Scatters : Trajectories :



W.<>l2 is a binary file with 30 rows and 2 columns
 It contains the coordinates of the rows (table 2)
 File :W.<>l2

Col.	Mini	Maxi
1	-6.948e+00	1.386e+01
2	-4.348e+00	5.674e+00

Les poids canoniques donnent des combinaisons de variables qui sont dans le fichier W.<>l2. Les valeurs sont des coordonnées des projections des relevés sur les axes de co-inertie (espace des espèces). On obtient une véritable carte factorielle des relevés (faune) qu'on peut voir par Scatters : Trajectories :



W.<>m1 is a binary file with 30 rows and 2 columns
 It contains the normalized coordinates of the rows (table 1)
 In earlier version of ADE this file name uses suffix bl1
 File :W.<>m1

Col.	Mini	Maxi
1	-1.925e+00	1.904e+00
2	-1.254e+00	3.763e+00

W.<>m2 is a binary file with 30 rows and 2 columns
 It contains the normalized coordinates of the rows (table 2)
 In earlier version of ADE this file uses suffix bl2
 File :W.<>m2

Col.	Mini	Maxi
1	-1.082e+00	2.158e+00
2	-1.518e+00	1.981e+00

Les deux systèmes d'axes de co-inertie donnent donc deux systèmes de coordonnées. L'analyse maximise la covariance entre les deux système, donc le produit de la corrélation par le produit des écarts-types de ces coordonnées. La représentation avec les coordonnées l1 et l2 exprime la part inertie dans la co-inertie. Si on les ramène aux variances unités (coordonnées m1 et m2), les deux systèmes deviennent superposables et on exprime la part corrélation dans la co-inertie. Le graphe canonique est donc obtenu avec Scatters : Match two scatters :

Match two scatters

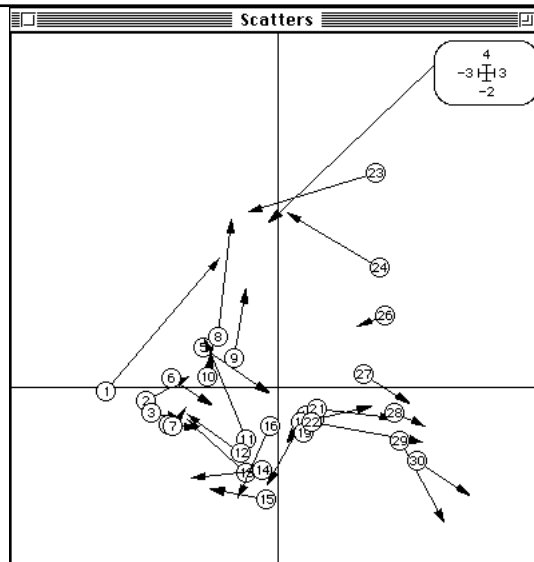
HY coordinates file

X-axis column number (default = 1)

Y-axis column number (default = 2)

Second HY coordinates file

Label file (or #) for rows



Chaque flèche représente chaque station par les variables de milieu (origine) et par la faune qu'on y trouve (extrémité). La pollution replie la structure longitudinale en réduisant l'abondance et la richesse quelque soit le cortège normal de la zone polluée.

W.<>a1 is a binary file with 3 rows and 2 columns
 It contains the coordinates of the projections of inertia axes onto the co-inertia axes (table 1)
 In earlier version of ADE this file uses the extension cp1
 File :W.<>a1

Col.	Mini	Maxi
1	-8.930e-01	3.884e-01
2	-9.079e-01	1.555e-02

Les axes d'inertie sont projetées sur les axes de co-inertie : cela permet de se rendre compte rapidement des relations existant entre structure des tableaux et co-structure. L'opération a lieu dans le premier espace :

Labels

HY coordinates file

X-axis column number (default = 1)

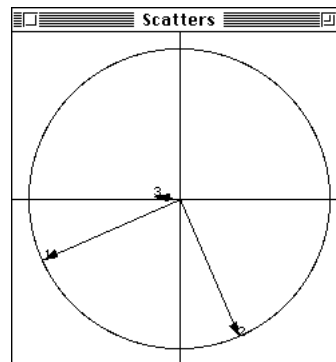
Y-axis column number (default = 2)

Label file (or # for item numbers)

Draw vectors from origin (yes = 1)

Draw unit circle (yes = 1)

Draw points (no = 2)

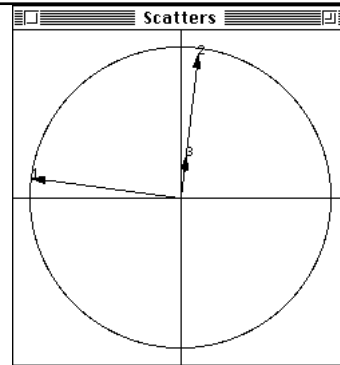


Le plan 1-2 de co-inertie est le plan 1-2 d'inertie avec une rotation. L'opération a lieu dans le second espace :

W.<>a2 is a binary file with 3 rows and 2 columns
 It contains the coordinates of the projections of inertia axes onto the co-inertia axes (table 2)
 In earlier version of ADE this file uses the extension cp2
 File :W.<>a2

Col.	Mini	Maxi
1	-9.803e-01	1.158e-01
2	1.261e-01	9.413e-01

Labels	
X-Y coordinates file	W, <a2
X-axis column number (default = 1)	
Y-axis column number (default = 2)	
Label file (or # for item numbers)	#
Draw vectors from origin (yes = 1)	1
Draw unit circle (yes = 1)	1
Draw points (no = 2)	2



Le listing final résume l'information acquise. Les colonnes `INER1` et `INER2` rappellent les inerties maximales qu'on peut projeter sur des axes dans chaque espace (obtenues par les analyses séparées). Les colonnes `Varian1` et `Varian2` donnent les inerties qu'on projette sur les axes de co-inertie. On voit ici que les plans 1-2 de co-inertie sont très sensiblement de même qualité que les plans 1-2 de chaque analyse. Mais en plus les deux projections sont reliées par les corrélations (`Correla.`) entre systèmes de coordonnées de même rang, ici 0.89 et 0.83, respectivement pour l'axe 1 et l'axe 2 de co-inertie. Ces valeurs élevées sont obtenues partiellement par optimisation car on optimise la covariance c'est-à-dire le produit de la corrélation et des radicaux des variances projetées. On a donc obtenu de bonnes valeurs sur les trois éléments optimisés soient l'inertie projetée dans chaque espace et la corrélation des systèmes de coordonnées.

Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	10.91	5.412	41.25	0.7302	6.322	42.75
2	3.724	2.839	8.201	0.7718	2.232	8.158



D'autres exemples sont donnés dans cette fiche.



- 1 Tucker, L.R. . (1958) An inter-battery method of factor analysis. *Psychometrika* : 23, 2, 111-136.
- 2 Cazes, P. (1980) L'analyse de certains tableaux rectangulaires décomposés en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. I. Définitions et applications à l'analyse canonique des variables qualitatives. II. Questionnaires : variantes des codages et nouveaux calculs de contributions. *Les Cahiers de l'Analyse des Données* : 5, 145-161 & 387-406.
- 3 Romane, F. (1972) Utilisation de l'analyse multivariée en Phytoécologie. *Investigación pesquera* : 36, 131-139.
Mercier, P, Chessel, D. & Dolédec, S. (1992) Complete correspondence analysis of an ecological profile data table: a central ordination method. *Acta Oecologica* : 13, 25-44.
- 4 Chessel, D. & Mercier, P. (1993) Couplage de triplets statistiques et liaisons espèces-environnement. In : *Biométrie et Environnement*. Lebreton, J.D. & Asselain, B. (Eds.) Masson, Paris. 15-44.
- 5 Dolédec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.
- 6 Bornette, G., Amoros, A. & Chessel, D. (1994) Rejuvenation in former braided channels of the Rhône River: successional patterns and allogenic processes. *Journal of Vegetation Science* : 5, 237-246.
- 7 Cadet, P., Thioulouse, J. & Albrecht, A. (1994) Relationships between ferrisol properties and the structure of plant parasitic nematode communities on sugarcane in Martinique (French West Indies). *Acta Oecologica* : 15, 767-780.

- 8** Devillers, J. & Chessel, D. (1995) Comparison of in vivo and in vitro toxicity tests from co-inertia analysis. In : *Computer-Aided Molecular Design. Applications in Agrochemicals, Materials and Pharmaceuticals*. Reynolds, C.H., Holloway, M.K. & Cox, H.K. (Eds.) ACS Symposium Series 589. American Chemical Society, Washington. 250-266.
- 9** Thioulouse, J. & Lobry, J.R. (1995) Co-inertia analysis of amino-acid physico-chemical properties and protein composition with the ADE package. *Computer Applications in the Biosciences* : 11, 321-329.
- 10** Castella, E. & Speight, M.C.D. (1996) Knowledge representation using fuzzy coded variables: an example based on the use of Syrphidae (Insecta, Diptera) in the assessment of riverine wetlands. *Ecological Modelling* : 85, 13-25.

Lamouroux, N., Pellegrin, F., Nandris, D. & Kohler, F. (1995) The Coffea arabica fungal pathosystem in New Caledonia: interactions at two different spatial scales. *Journal of Phytopathology* : 143, 403-413.

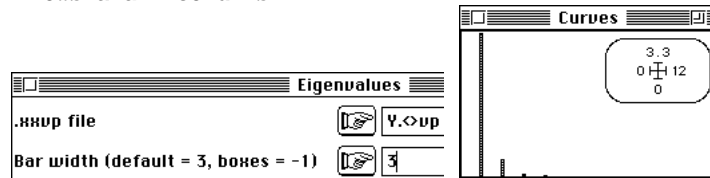
Verneaux, J., Schmitt, A. & Verneaux, V. (1995) Classification biologique des lacs jurassiens à l'aide d'une nouvelle méthode d'analyse des peuplements benthiques III Relations entre données biologiques et variables du milieu. *Annales de Limnologie* : 31, 277-286.
- 11** Chessel, D. & Hanafi, M. (1996) Analyses de la co-inertie de K nuages de points. *Revue de Statistique Appliquée* : 44, 35-60.
- 12** Verneaux, J. (1973) *Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie*. Thèse d'état, Besançon. 1-257.


```

03 +7.3517E-02 +0.0193 +0.9842 |04 +3.7693E-02 +0.0099 +0.9941 |
...
09 +2.0074E-04 +0.0001 +1.0000 |10 +5.0208E-05 +0.0000 +1.0000 |
11 +0.0000E+00 +0.0000 +1.0000

```

File Y.<>vp contains the eigenvalues and relative inertia for each axis
--- It has 11 rows and 2 columns



Co-inertia analysis between two statistical triplets

```

1 ---> DouMil.cnta (rows: 30, col: 11, axes: 3, inertia: 11.000000)
2 ---> Fau.fcta (rows: 30, col: 27, axes: 4, inertia: 1.776020)

```

Co-inertia: 3.7996, RV coefficient: 0.69486

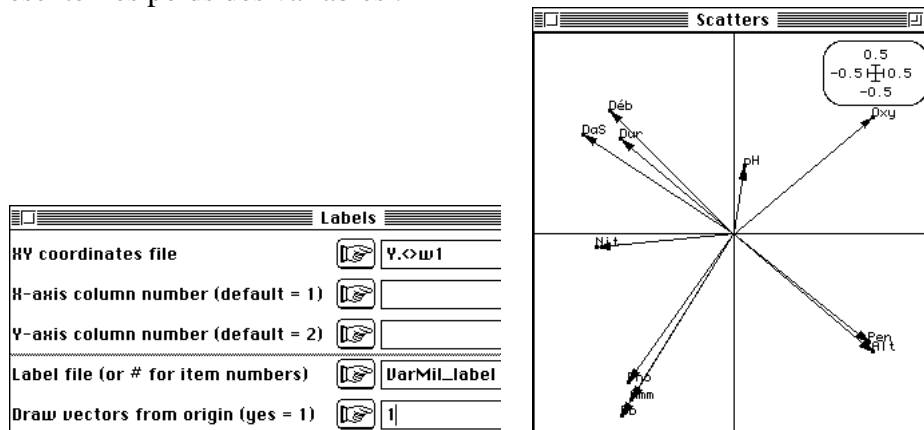
Y.<>w1 is a binary file with 11 rows and 2 columns

It contains the canonical weights of the variables of table 1

File :Y.<>w1

Col.	Mini	Maxi
1	-3.753e-01	3.497e-01
2	-4.550e-01	3.054e-01

Représenter les poids des variables :



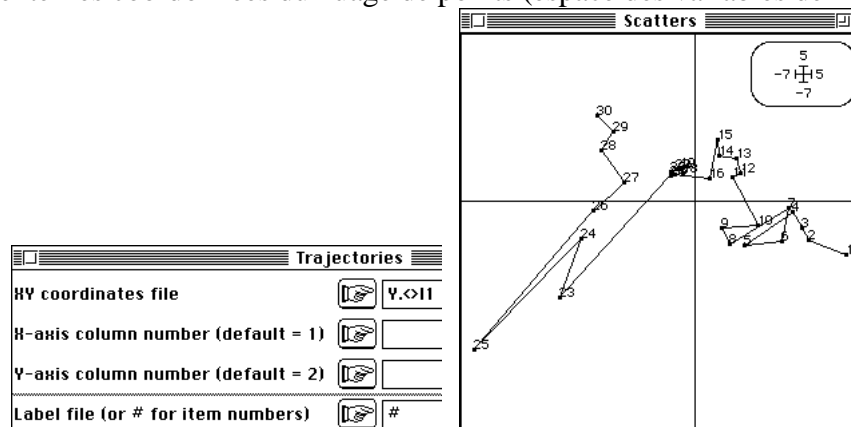
Y.<>l1 is a binary file with 30 rows and 2 columns

It contains the coordinates of the rows (table 1)

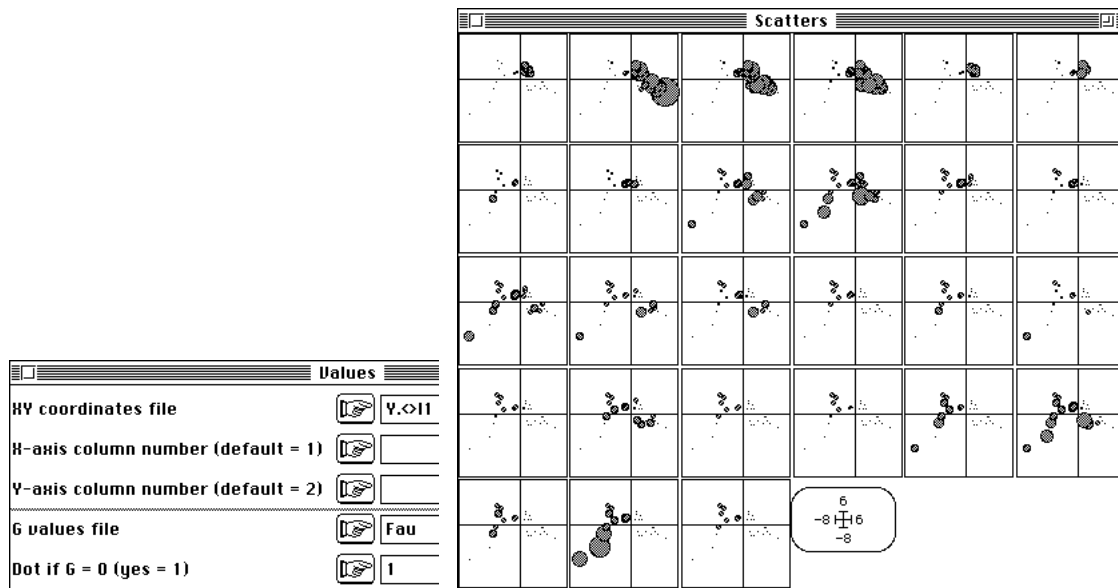
File :Y.<>l1

Col.	Mini	Maxi
1	-6.570e+00	4.525e+00
2	-4.451e+00	2.526e+00

Représenter les coordonnées du nuage de points (espace des variables de milieu) :



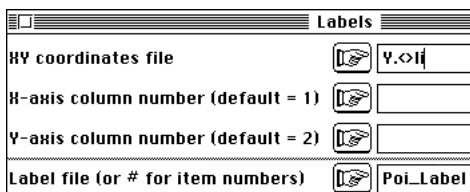
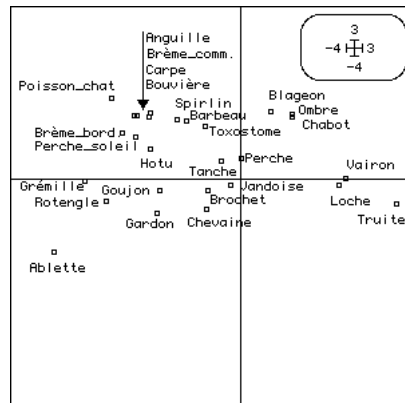
Chaque espèce est une distribution de fréquence sur cet ensemble de points :



La position moyenne de chaque espèce est dans :

File Y.<>11 contains the row scores
 --- It has 27 rows and 2 columns
 File :Y.<>11

Col.	Mini	Maxi
1	-3.238e+00	2.738e+00
2	-1.275e+00	1.391e+00



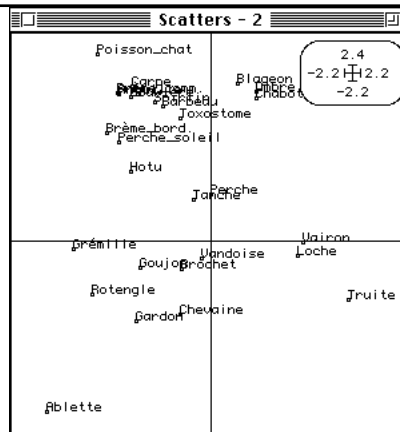
L'analyse de co-inertie fonctionne ici comme une analyse canonique des correspondances ¹ avec une autre contrainte : poids des variables de milieu, combinaison linéaire de ces variables, position moyenne par espèces. La variance de ces positions moyennes (séparation de niches) est ici optimale sous contrainte que la somme des carrés des poids des variables est unitaire.

On a le résultat dual : les espèces sont positionnées par un code de variance unité :

Y.<>w2 is a binary file with 27 rows and 2 columns
 It contains the canonical weights of the variables of table 2
 File :Y.<>w2

Col.	Mini	Maxi
1	-1.797e+00	1.519e+00
2	-1.967e+00	2.148e+00

Labels	
HV coordinates file	<input type="button" value="Y.<w2"/>
H-axis column number (default = 1)	<input type="button" value=""/>
Y-axis column number (default = 2)	<input type="button" value=""/>
Label file (or # for item numbers)	<input type="button" value="Poi_Label"/>

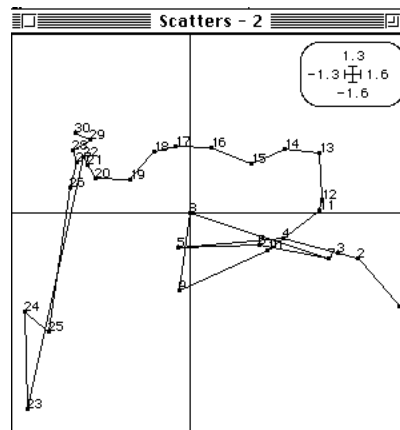


Ceci permet de positionner les relevés à la moyenne des espèces qui y sont :

Y.<>l2 is a binary file with 30 rows and 2 columns
 It contains the coordinates of the rows (table 2)
 File :Y.<>l2

Col.	Mini	Maxi
1	-1.203e+00	1.520e+00
2	-1.424e+00	5.814e-01

Trajectories	
HV coordinates file	<input type="button" value="Y.<l2"/>
H-axis column number (default = 1)	<input type="button" value=""/>
Y-axis column number (default = 2)	<input type="button" value=""/>
Label file (or # for item numbers)	<input type="button" value="#"/>

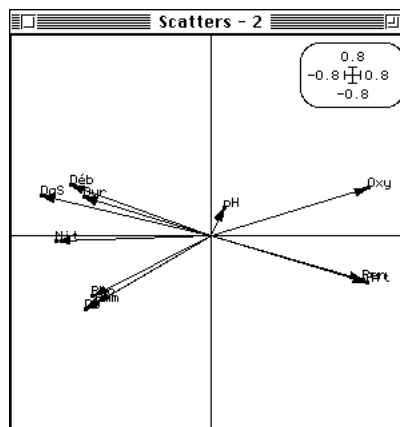


Ces coordonnées maximise les sommes des covariances au carré avec les variables de milieu. Ces covariances sont dans :

File Y.<>co contains the column scores
 --- It has 11 rows and 2 columns
 File :Y.<>co

Col.	Mini	Maxi
1	-6.761e-01	6.301e-01
2	-2.948e-01	1.979e-01

Labels	
HV coordinates file	<input type="button" value="Y.<co"/>
H-axis column number (default = 1)	<input type="button" value=""/>
Y-axis column number (default = 2)	<input type="button" value=""/>
Label file (or # for item numbers)	<input type="button" value="VarMil_Label"/>
Draw vectors from origin (yes = 1)	<input type="button" value="1"/>



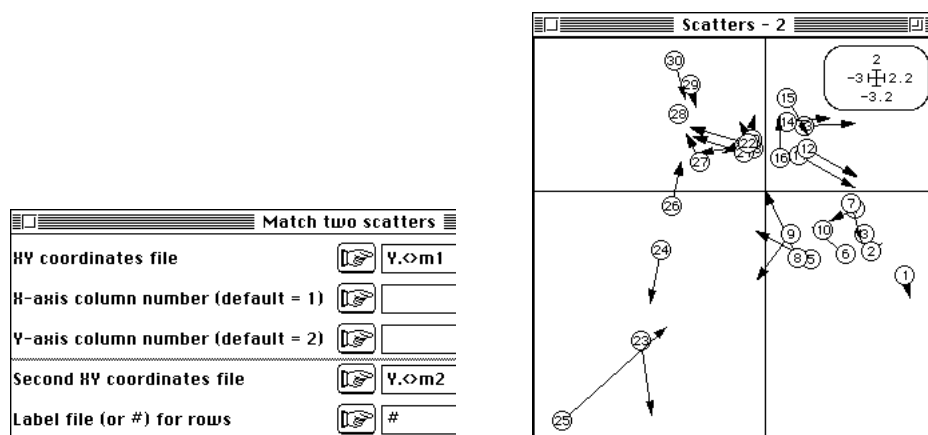
Y.<>m1 is a binary file with 30 rows and 2 columns
 It contains the normalized coordinates of the rows (table 1)
 In earlier version of ADE this file name uses suffix b11
 File :Y.<>m1

Col.	Mini	Maxi
1	-2.627e+00	1.810e+00
2	-2.986e+00	1.694e+00

Y.<>m2 is a binary file with 30 rows and 2 columns
 It contains the normalized coordinates of the rows (table 2)
 In earlier version of ADE this file uses suffix b12
 File :Y.<>m2

Col.	Mini	Maxi
1	-1.492e+00	1.886e+00
2	-2.907e+00	1.187e+00

On conserve également la notion de double représentation :



Remarquer ici la parenté des raisonnements en analyse de co-inertie ACP-AFC et analyse canonique des correspondances. Pour une dizaine de variables de milieu sur 30 stations, la première l'emporte sur la seconde qui implique des régressions linéaires sur beaucoup de variables (10) avec peu de points (30).



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

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

Colnertia : Coinertia test - Fixed Tab 1



Test de permutation sur la co-structure de deux tableaux.



On compare la co-structure observée de deux tableaux à un grand nombre de permutations aléatoires dans l'appariement des deux tableaux. Ici on réalise un test de Monte-Carlo où seules les lignes du second tableau sont permutées. A chaque permutation, les poids des lignes du premier tableau restent fixés et le second tableau est recentré. Cette option s'utilise en particulier lorsque le tableau 1 est un tableau d'AFC et le tableau 2 est un tableau d'ACP par exemple.



L'option utilise une seule fenêtre de dialogue :

Nom du fichier d'entrée de type ---.<gt;ta créé par CoInertia : Matching two statistical triplets.

Nombre de permutations utilisées.



Reprendre l'exemple de l'option CoInertia : Coinertia analysis.



La seule difficulté de l'analyse de co-inertie est la nécessité de coupler deux tableaux ayant la même pondération des lignes. Si dans le même exemple on veut coupler l'ACP normée — qui semble s'imposer — et l'AFC du tableau faunistique, on ne pourra pas le faire sans précaution. Nous pouvons transporter dans l'ACP la pondération issue de l'AFC :

Faire l'AFC du tableau faunistique (COA : COrrespondence Analysis) :

Faire l'ACP du tableau de variables de milieu en utilisant la pondération précédente (PCA : Correlation matrix PCA) :

Coupler les deux tableaux (CoInertia : Matching two statistical triplets) :

```
First input statistical triplet: table DouMil.cnta
Number of rows: 30, columns: 11
Second input statistical triplet: table DouPoi.fcta
Number of rows: 30, columns: 27
```

```
File AA.<gt;ta contains the cross table YtDnX
It has 27 rows and 11 columns
```




```
File AA.<gt;pl contains the weight of each row
It has 27 rows and 1 column
```

File AA.<>pc contains the weight of each column
It has 11 rows and 1 column

File AA.<>ma contains the name of the input statistical triplet
1 ---> DouMil.cnta
2 ---> DouPoi.fcta

total inertia: 2.590287

Remarquer que l'ordre d'entrée des deux tableaux n'a pas d'importance mais que c'est le programme qui décide de l'ordre tableau 1 et tableau 2.

Matching two statistical triplets			
First data input file		DouMil.cnta	30 11
Second data input file		DouPoi.fcta	30 27
Output file name		BB	

First input statistical triplet: table DouMil.cnta
Number of rows: 30, columns: 11
Second input statistical triplet: table DouPoi.fcta
Number of rows: 30, columns: 27

File BB.<>ta contains the crossed table YtDnX
It has 27 rows and 11 columns

File BB.<>pl contains the weight of each row
It has 27 rows and 1 column

File BB.<>pc contains the weight of each column
It has 11 rows and 1 column

File BB.<>ma contains the name of the input statistical triplet
1 ---> DouMil.cnta
2 ---> DouPoi.fcta

total inertia: 2.590287

Les deux options **CoInertia : Coinertia test - Fixed Tab 1** et **CoInertia : Coinertia test - Fixed Tab 2** font exactement le même travail. Ici le tableau qui est doit être fixé et impose sa pondération à l'autre est celui d'AFC. Il est placé en position 2 car il y a plus d'espèces que de variables. Utiliser donc **CoInertia : Coinertia test - Fixed Tab 2**.



La version 3.7 proposait un test de permutations sur les valeurs propres qui a été utilisé dans ¹. Sur la première valeur propre, ce test est correct mais il est discutable sur les suivantes. Suite aux travaux récents ², nous avons préféré suivre la proposition plus simple de tester la co-inertie globale (somme des valeurs propres de l'analyse de co-inertie) et laisser le graphe des valeurs propres imposer — dans les cas les plus fréquents — le choix du nombre d'axes à conserver. En cas de signification faible et/ou de graphe de valeurs propres peu explicite, l'analyse est à rejeter.

Noter que si on note θ_r les valeurs propres de l'analyse de co-inertie, λ_k les valeurs propres de la première analyse et μ_j les valeurs propres de la seconde analyse, le coefficient RV ³ est :

$$RV = \frac{\sum_r \theta_r}{\sqrt{\sum_k \lambda_k^2} \sqrt{\sum_j \mu_j^2}}$$

Tester l'inertie totale est donc équivalent à tester la corrélation vectorielle entre les deux tableaux, le dénominateur restant constant par permutation. c'est un test global de corrélation entre structures.



- 1 Statzner, B., Resh, V. & Dolédec, S. (1994) Ecology of the Upper Rhône River: a test of habitat template theories. Special issue of *Freshwater Biology* (Editeurs invités). 31, 253-554.
- 2 Fraile, L., Escoufier, Y. & Raibaut, A. (1993) Analyse des correspondances de données planifiées : Etude de la chémotaxie de la larve infestante d'un parasite. *Biometrics* : 49, 1142-1153.
Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1996) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis*, 20, 643-656.
- 3 Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.

Colnertia : Coinertia test - Fixed Tab 2



Test de permutation sur la co-structure de deux tableaux.



On compare la co-structure observée de deux tableaux à un grand nombre de permutations aléatoires dans l'appariement des deux tableaux. Ici on réalise un test de Monte-Carlo où seules les lignes du premier tableau sont permutées. A chaque permutation, les poids des lignes du second tableau restent fixés et le premier tableau est recentré. Cette option s'utilise en particulier lorsque le tableau 2 est un tableau d'AFC et le tableau 1 est un tableau d'ACP par exemple.



L'option utilise une seule fenêtre de dialogue :

---	<ta input file		
Select a number of permutations			

Nom du fichier d'entrée de type ---.<ta créé par Colnertia : Matching two statistical triplets.

Nombre de permutations utilisées.



Utiliser la carte Doubs ¹ de la pile ADE-4•Data. Ajouter les fichiers créés par la carte Doubs+1.

Exécuter l'AFC du tableau faunistique (COA : COrrépondence Analysis) :

Data file		DouPoi	30	27
-----------	--	--------	----	----

Cette analyse donne un poids à chacune des stations. Ce poids est proportionnel à l'abondance des organismes qu'on y a trouvé. Il est nul pour les stations vides.

Exécuter l'ACP normée du tableaux de variables de milieu (PCA : Correlation matrix PCA) en utilisant la pondération de l'analyse précédente :

Matrix input file		DouMil	30	11
Row weights (default=1/n)		3		
Column weights (default=1)				
Option: file for row weighting		DouPoi.fcpl	30	1
Option: file for column weighting				
I = Save correlation matrix		1		

Associer les deux analyses (Colnertia : Matching two statistical triplets) :

First data input file		DouPoi.fcta	30	27
Second data input file		DouMil.cnta	30	11
Output file name		K		

First input statistical triplet: table DouMil.cnta
Number of rows: 30, columns: 11
Second input statistical triplet: table DouPoi.fcta
Number of rows: 30, columns: 27

File K.<>ta contains the crossed table YtDnX
It has 27 rows and 11 columns

File K.<>pl contains the weight of each row
It has 27 rows and 1 column

File K.<>pc contains the weight of each column
It has 11 rows and 1 column

```
File K.<ma contains the name of the input statistical triplet
1 ---> DouMil.cnta
2 ---> DouPoi.fcta
```

total inertia: 2.590287

Si on veut tester la co-inertie entre les deux tableaux, il faut respecter le fait que la pondération du tableau 2 sert au centrage et à la normalisation des variables du tableau 1. Le tableau 2 est fixé, ses poids des lignes sont fixés et on permute aléatoirement les lignes du tableau 1 qui est recentré à chaque simulation. La co-inertie totale observée dans les données est comparée à la co-inertie obtenue par permutations :



Le listing rappelle la condition de validité de ce test :

```
Warning: non uniform weight. The results from permutations
are valid only if the row weights come from the fixed table.
-----
```

```
The fixed table is table 2: DouPoi.fcta
number of random matching: 1000   Observed: 2.590287
Histogramm:  minimum = 0.124923, maximum = 2.590287
number of simulation X<Obs: 1000 (frequency: 1.000000)
number of simulation X>=Obs: 0 (frequency: 0.000000)
```

```
*****
*****
*****
*****
*****
****
***
**
*

•-->
```

Aucune des 1000 simulations donne une co-inertie supérieure à l'observation et le test est très significatif.



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

Colnertia : Fully Matched Tables



Méthode d'analyse de données à opérateur non positif.



Quand deux tableaux ont les mêmes lignes et les mêmes colonnes, les nuages de lignes sont dans un même espace et on peut chercher des axes de co-inertie communs aux deux tableaux. L'option met en œuvre cette méthode décrite dans ¹.



L'option utilise une seule fenêtre de dialogue :

Fully Matched Tables

First data input file

Second data input file

Output file name

Quit Ok

Nom du premier triplet statistique.

Nom du second triplet statistique.

Nom générique des fichiers de sortie.



Utiliser la carte Buech 2 3 de la pile ADE-4•Data. Le premier tableau Mil1 contient 31 mesures sur 10 variables physiques et chimiques. Il est acquis en juin. Le second tableau Mil2 donne les valeurs des mêmes variables sur les mêmes individus. Il est acquis en septembre. Faire deux ACP normée et garder 3 facteurs arbitrairement :

Correlation matrix PCA

Matrix input file 31 10

Correlation matrix PCA

Matrix input file 31 10

Coupler les deux triplets :

Fully Matched Tables

First data input file 31 10

Second data input file 31 10

Output file name

La matrice diagonalisée est symétrique mais non positive. Elle a des valeurs propres négatives mais on ne s'intéresse qu'aux valeurs propres positives qui expriment comment une même combinaison de variables peut covarier entre les deux tableaux :



Number of axes ?

OK

On garde un facteur.

First input statistical triplet: table Mil1.cnta
Number of rows: 31, columns: 10
Second input statistical triplet: table Mil2.cnta
Number of rows: 31, columns: 10

File MilMil_S.croi contains the cross table YtDnX

It has 10 rows and 10 columns

On a d'abord l'équivalent de CoInertia : Matching two statistical triplets.

Symmetric operator diagonalization

```
-----  
Num Eigenval. | Num. Eigenval. | Num. Eigenval. | Num. Eigenval. |  
001 3.612e+00 | 002 8.758e-01 | 003 6.667e-01 | 004 2.673e-01 |  
005 7.196e-02 | 006 1.942e-02 | 007 -2.695e-02 | 008 -7.168e-02 |  
009 -1.297e-01 | 010 -3.807e-01 |
```

Les valeurs propres sont éditées et conservées.

File MilMil_S.valp contains the eigenvalues
--- It has 10 rows and 1 column



File MilMil_S.vecp contains the eigenvectors (Dp-normed)
--- It has 10 rows and 1 columns

File :MilMil_S.vecp

Col.	Mini	Maxi
1	7.884e-02	4.439e-01

Les vecteurs propres sont conservés : les composantes sont dans ce cas les coefficients des combinaisons linéaires des variables normées que sont les coordonnées des lignes. Il y a une seule série de coefficients utilisés pour les deux tableaux (0.0788 , 0.4439, 0.3258, 0.3561, 0.3881 , 0.3915, 0.2769, 0.2835, 0.2064, 0.2457).

File MilMil_S.cp1 contains the scores of the projections of inertia axes onto co-inertia axes (Tab 1)
--- It has 3 rows and 1 columns

File :MilMil_S.cp1

Col.	Mini	Maxi
1	1.477e-01	9.576e-01

File MilMil_S.cp2 contains the scores of the projections of inertia axes onto co-inertia axes (Tab 2)
--- It has 3 rows and 1 columns

File :MilMil_S.cp2

Col.	Mini	Maxi
1	-9.845e-01	1.177e-01

Les axes principaux des analyses séparées sont projetés sur les axes communs de co-inertie. Les fichiers cp1 et cp2 s'utilisent comme dans une analyse de co-inertie ordinaire :

MilMil_S.cp1		MilMil_S.cp2	
1	0.9576	1	-0.9845
2	0.2023	2	0.0106
3	0.1477	3	0.1177

L'axe de co-inertie commun conservé est voisin des axes 1 de chaque analyse (au sens près qui est aléatoire).

Les deux nuages des lignes de chaque tableau sont projetés sur les axes de co-inertie (fichiers li1 et li2) et les coordonnées sont normalisées (fichiers mi1 et mi2) :

File MilMil_S.li1 contains the coordinates of the rows of table 1
--- It has 31 rows and 1 columns
File :MilMil_S.li1

Col.	Mini	Maxi
1	-4.661e+00	4.379e+00

File MilMil_S.mi1 contains the scores (unit norm) of the rows of table 1
 --- It has 31 rows and 1 columns

File :MilMil_S.mi1

Col.	Mini	Maxi
1	-2.350e+00	2.207e+00

File MilMil_S.li2 contains the coordinates of the rows of table 2
 --- It has 31 rows and 1 columns

File :MilMil_S.li2

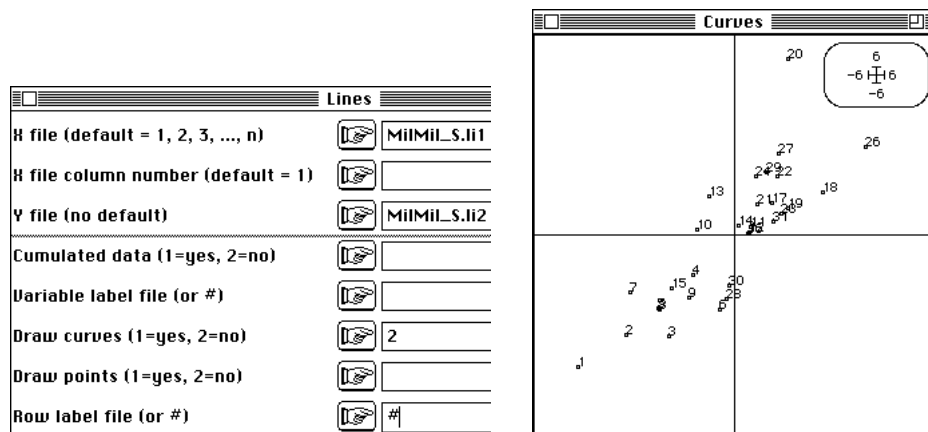
Col.	Mini	Maxi
1	-3.991e+00	5.250e+00

File MilMil_S.mi2 contains the scores (unit norm) of the rows of table 2
 --- It has 31 rows and 1 columns

File :MilMil_S.mi2

Col.	Mini	Maxi
1	-1.865e+00	2.454e+00

On voit la covariance entre les deux systèmes de coordonnées (critère optimisé) :



Les statistiques de synthèse ont la même signification que dans une analyse de co-inertie ordinaire :

Inertia statistics

Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	3.612	3.935	4.579	0.851	4.146	4.697

Dans 1, on trouve un exemple avec deux ACP centrées reproductible à partir de la carte Accidents 4.



Le programme vérifie que les deux tableaux ont le même nombre de lignes et de colonnes mais l'analyse n'a de sens que si les unités statistiques (lignes) et les variables (colonnes) ont la même signification dans les deux tableaux.



Dans ADE 3.7 l'option correspondait au module BiMatchSim.



¹ Torre, F. & Chessel, D. (1994) Co-structure de deux tableaux totalement appariés. *Revue de Statistique Appliquée* : 43, 109-121.

- 2 Vespini, F. (1985) *Contribution à l'étude hydrobiologique du Buech, rivière non aménagée de Haute-Provence*. Thèse de troisième cycle, Université de Provence. 1-148 + bibliographie + annexes.
- 3 Vespini, F., Légier, P. & Champeau, A. (1987) Ecologie d'une rivière non aménagée des Alpes du Sud : Le Buëch (France) I Evolution longitudinale des descripteurs physiques et chimiques. *Annales de Limnologie* : 23, 151-164.
- 4 Lafosse, R. (1985) *Analyses procustéennes de deux tableaux. Proposition d'une technique visant à la détection de points originaux. essai de présentation synthétique d'analyses de deux tableaux*. Thèse de 3^e cycle, Université de Toulouse III. 1-69.

Colnertia : Matching two statistical triplets



Utilitaire d'entrée dans le module CoInertia.



L'option teste la compatibilité de deux triplets statistiques et enregistre leur nom dans un fichier auxiliaire. Elle calcule le tableau croisé et met en place un triplet statistique dit de co-inertie (voir CoInertia : CoInertia analysis). L'exécution de cette option est obligatoire avant l'usage d'une autre option de ce module.



L'option utilise une seule fenêtre de dialogue :

Nom du premier triplet statistique.

Nom du second triplet statistique.

Nom générique des fichiers de sortie.



Utiliser la carte Dune ¹ de la pile ADE-4•Data. Éditer le tableau de variables de milieu :

ADE-4•Data,Dune,DMD.txt									
02	8	1	2	4	1	0	0	0	0
03	5	1	2	2	0	1	0	0	0
04	3	2	2	4	1	0	0	0	0
04	2	2	2	4	1	0	0	0	0
05	3	1	1	2	0	0	1	0	0
04	3	1	2	2	0	0	1	0	0
02	8	1	3	3	0	0	1	0	0
04	2	5	3	3	0	0	1	0	0
03	7	4	1	1	0	0	1	0	0
03	3	2	1	1	0	1	0	0	0
03	5	1	3	1	0	1	0	0	0
05	8	4	2	2	1	0	0	0	0
06	0	5	2	3	1	0	0	0	0
09	3	5	3	0	0	0	0	1	0
11	5	5	2	0	0	0	0	1	0
05	7	5	3	3	1	0	0	0	0
04	0	2	1	0	0	0	0	1	0
04	6	1	1	0	0	0	0	1	0
03	7	5	1	0	0	0	0	1	0
03	5	5	1	0	0	0	0	1	0

Tab2.txt	
	A
1	1
2	2
3	1
4	1
5	3
6	3
7	3
8	3
9	3
10	2
11	2
12	1
13	1
14	4
15	4
16	1
17	4
18	4
19	4
20	4

On y trouve 4 variables quantitatives (Thickness, Moisture, Use, Manuring) et une variable qualitative à 4 modalités (Farming management, Biological management, Hobby farming, Nature conservation). Sélectionner (FilesUtil : Row-Col Selection) les 4 premières colonnes et en faire l'ACP normée :

A l'aide d'un tableur, éditer la variable qualitative par numéro de modalités (ci-dessus) à droite de la fenêtre du fichier de départ). Passer en binaire ce fichier à 20 lignes et 1 colonne et le lire (CategVar : Read Categ File) :

Categorical variables: file Tab2
 Rows: 20, Variables: 1, Categories: 4, Missing data: 0

Description of categories:

Variable number 1 has 4 categories

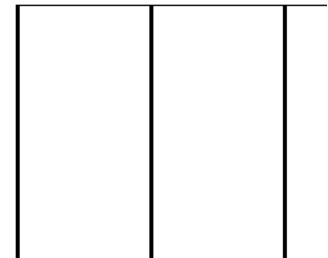
[1]Category:	1 Num:	6 Freq.:	0.3
[2]Category:	2 Num:	3 Freq.:	0.15
[3]Category:	3 Num:	5 Freq.:	0.25
[4]Category:	4 Num:	6 Freq.:	0.3

En faire l'ACM (MCA : Multiple Correspondence Analysis) :

Multiple Correspondence Analysis

.cat type file

Option: Row weighting file



Number of axes ?

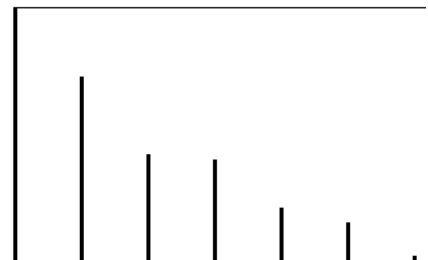
Cette opération n'a évidemment pas pour but d'analyser le tableau qui donne 3 valeurs propres égales à 1/3 et aucune indication sur la structure de la variable mais de regrouper la tableau 1 (ACP normée) et le tableau 2 (ACM) dans l'analyse de Hill et Smith² faite pour traiter les mélanges de variables qualitatives et quantitatives (MCA : Hill & Smith Analysis) :

Hill & Smith Analysis

Discrete characters (.cmta) 20 4

Continuous characters (.cnta) 20 4

Output file name



Number of axes ?

Transposer le tableau floristique (FilesUtil : Transpose) pour ramener les stations en lignes :

Transpose

Input file 30 20

Output file

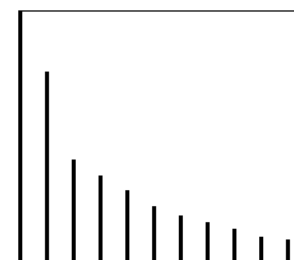
Exécuter l'AFC à pondération uniforme de ce tableau (COA : Row weighted COA) :

Row weighted COA

Data file 20 30

Row weight (default 1/n)

Generic output file



Number of axes ?

Pour savoir si les deux typologies de stations (la première vue par l'analyse des variables de milieu, l'autre vue par l'analyse de la végétation) sont cohérentes, associer les deux analyses :


```

03 +2.6206E-02 +0.0755 +0.9191 | 04 +1.9093E-02 +0.0550 +0.9741 |
05 +4.4420E-03 +0.0128 +0.9869 | 06 +3.7045E-03 +0.0107 +0.9976 |
07 +8.4780E-04 +0.0024 +1.0000 | 08 +0.0000E+00 +0.0000 +1.0000 |

```

File A.<>vp contains the eigenvalues and relative inertia for each axis
--- It has 8 rows and 2 columns

A.<>a1 is a binary file with 2 rows and 2 columns
It contains the coordinates of the projections of inertia axes onto the co-inertia axes (table 1)

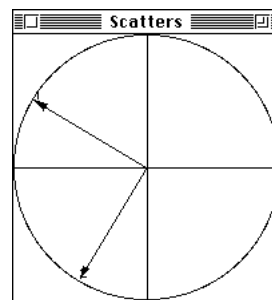
In earlier version of ADE this file uses suffix cp1

File :A.<>a1

Col.	Mini	Maxi
1	-8.485e-01	-4.999e-01
2	-8.399e-01	5.037e-01

Dessiner la projection sur le premier plan de co-inertie des axes d'inertie de l'analyse du tableau 1 :

Labels	
XY coordinates file	R.<>a1
X-axis column number (default = 1)	
Y-axis column number (default = 2)	
Label file (or # for item numbers)	#
Draw vectors from origin (yes = 1)	1
Draw unit circle (yes = 1)	1
Draw points (no = 2)	2



A.<>a2 is a binary file with 2 rows and 2 columns
It contains the coordinates of the projections of inertia axes onto the co-inertia axes (table 2)

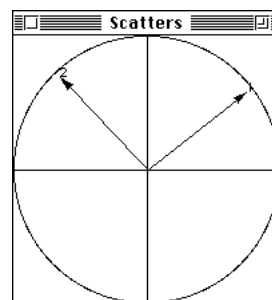
In earlier version of ADE this file uses suffix cp2

File :A.<>a2

Col.	Mini	Maxi
1	-6.407e-01	7.482e-01
2	5.738e-01	6.859e-01

Dessiner la projection sur le premier plan de co-inertie des axes d'inertie de l'analyse du tableau 2 :

Labels	
XY coordinates file	R.<>a2
X-axis column number (default = 1)	
Y-axis column number (default = 2)	
Label file (or # for item numbers)	#
Draw vectors from origin (yes = 1)	1
Draw unit circle (yes = 1)	1
Draw points (no = 2)	2



Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	0.4278	0.4329	0.4985	0.921	0.4753	0.5677
2	0.3315	0.3697	0.4255	0.8359	0.3475	0.4322

On vérifie très simplement que les deux plans d'inertie des deux analyses séparées sont les plans de co-inertie.



Fondamentalement cet exemple partiel attire l'attention sur le caractère très structuré de l'enchaînement des procédures dans ADE-4. Le module CoInertia avec ses options de préparation (CoInertia : Matching two statistical triplets), de test de signification (CoInertia : Coinertia test - Fixed D) et de calculs principaux (CoInertia : Coinertia analysis) autorise l'entrée dans un schéma très général de tous les types d'analyse classique. Potentiellement, on peut compter plus de 200 combinaisons possibles. Les plus communes sont :

- Variables de milieu qualitatives (ACM) / Tableau écologique traité en AFC ;
- Variables de milieu qualitatives (ACM) / Tableau écologique en présence-absence traité en AFC (procédure compatible avec la théorie des profils écologiques : EcolTools : Ecological Profiles) ;
- Variables de milieu quantitatives (ACP) / Tableau écologique traité en AFC (procédure compatible avec la théorie de la niche et voisine de l'analyse canonique des correspondances : Projectors : PCA on Instrumental Variables) ;
- Variables de milieu quantitatives (ACP) / Tableau écologique traité en ACP (procédure conseillée en cas de facteur écologique limitant et voisine de l'analyse en composantes principales sur variables instrumentales : Projectors : PCA on Instrumental Variables) ;

La version ACM-ACM est l'analyse discriminante sur variables qualitatives de Cazes ³.

La version ACP-ACP est l'analyse inter-batterie de Tucker ⁴.

On peut coupler deux analyses décentrées (PCA : Decentring X[i,j] - Model[i,j]) ⁵.

On peut coupler par la marge espèces de l'information biologique (MCA : Fuzzy Correspondence Analysis ⁶) et de l'information écologique (COA : Row weighted COA) ⁷.

Le module RLQ étend ces possibilités aux couplages doubles sur les marges espèces (variables biologiques) et relevés (variables mésologiques) ⁸.



- 1 Jongman, R.H., ter Braak, C.J.F. & van Tongeren, O.F.R. (1987) *Data analysis in community and landscape ecology*. Pudoc, Wageningen. 1-298.
- 2 Hill, M.O. & Smith, A.J.E. (1976) Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* : 25, 249-255.
- 3 Cazes, P. (1980) L'analyse de certains tableaux rectangulaires décomposé en blocs : généralisation des propriétés rencontrées dans l'étude des correspondances multiples. I. Définitions et applications à l'analyse canonique des variables qualitatives. II. Questionnaires : variantes des codages et nouveaux calculs de contributions. *Les Cahiers de l'Analyse des Données* : 5, 145-161 & 387-406.
- 4 Tucker, L.R. . (1958) An inter-battery method of factor analysis. *Psychometrika* : 23, 2, 111-136.
- 5 Castella, E., Bickerton, M., Armitage, P.D. & Petts, G.E. (1995) The effects of water abstractions on invertebrate communities in U.K. streams. *Hydrobiologia* : 308, 167-182.
- 6 Chevenet, F., Dolédec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309.
- 7 Statzner, B., Resh, V. & Dolédec, S. (1994) Ecology of the Upper Rhône River: a test of habitat templet theories. Special issue of *Freshwater Biology* (Éditeurs invités). 31, 253-554.
- 8 Dolédec, S., Chessel, D., Ter Braak, C.J.F. & Champely, S. (1996) Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* : in press.