

Co-structure de deux analyses en composantes principales

Résumé

La fiche aborde un des aspects du couplage de tableaux, celui du test de l'existence d'une co-structure entre deux tableaux individus-variables dont on a étudié la structure par une analyse en composantes principales. ADE propose un test de permutation sur une analyse de co-inertie basée sur l'analyse inter-batteries de Tucker (1958, An inter-battery method of factor analysis. *Psychometrika* : 23, 2, 111-136) connue aussi comme le premier pas des *Partial Least Square regression* (Höskuldsson, 1988, PLS regression methods. *Journal of Chemometrics* : 2, 211-228).

Plan

1 - Type de données	2
2 - Structure du tableau de végétation	3
3 - Structure du tableau de milieu	5
4 - Co-inertie des deux nuages	9
5 - Test de permutation	14
6 - Axes d'inertie et axes de co-inertie	15
Références	20

S. Dolédec & D. Chessel

1 - Type de données

Le type de données auquel s'adresse la méthode est parfaitement décrit dans l'article de Jongman *et al.* (1987)¹ : "An important question is *What situations are typical of those where PLS methods can be expected to be good for modelling purposes?*. They are the ones where there are many variables but not necessarily many samples or observations. This a common situation in many laboratories. Typically it may take some time to get a new sample, but each sample may give a large amount of information (variables)".

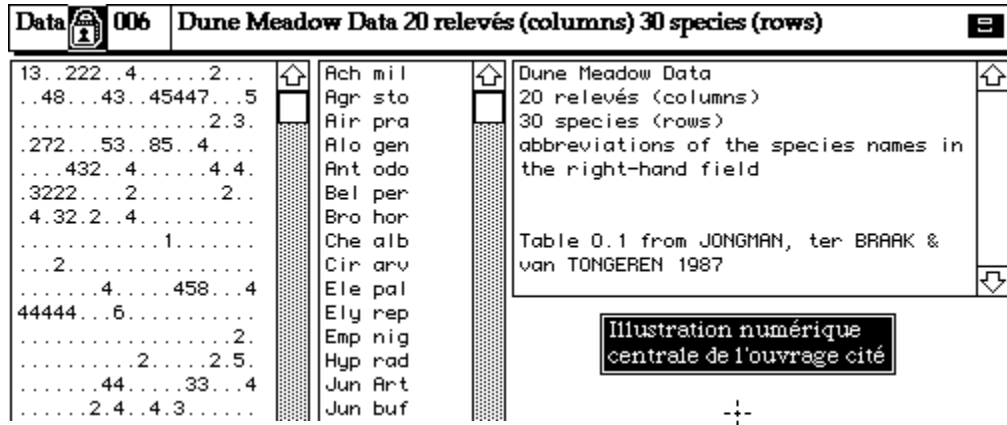


Figure 1 Carte Dune+1 de la pile ADE•Data

On traite dans cette fiche l'exemple central de l'ouvrage cité. Vingt relevés de végétation dunaire (lignes) ont permis de récolter 30 espèces (colonnes). Le code des espèces issu de l'ouvrage est résumé par les abréviations qui suivent pour être accessible pour les illustrations :

- | | | | | |
|------------|-------------|-------------|-------------|-------------|
| 1- Ach mil | 8- Che alb | 15- Jun buf | 22- Ran fla | 29- Bra rut |
| 2- Agr sto | 9- Cir arv | 16- Leo aut | 23- Rum ace | 30- Cal cus |
| 3- Air pra | 10- Ele pal | 17- Lol per | 24- Sag pro | |
| 4- Alo gen | 11- Ely rep | 18- Pla lan | 25- Sal rep | |
| 5- Ant odo | 12- Emp nig | 19- Poa pra | 26- Tri pra | |
| 6- Bel per | 13- Hyp rad | 20- Poa tri | 27- Tri rep | |
| 7- Bro hor | 14- Jun Art | 21- Pot pal | 28- Vic lat | |

	Veg-t																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	7	0	4	2	0	0	0	0	0	0	0	0	0	0
2	3	0	0	2	0	3	4	0	0	0	4	0	0	0	0	5	5	0	4	7	0	0	0	0	0	0	5	0	0	
3	0	4	0	7	0	2	0	0	0	0	4	0	0	0	0	2	6	0	5	6	0	0	0	0	0	2	0	2	0	
4	0	8	0	2	0	2	3	0	2	0	4	0	0	0	0	2	5	0	4	5	0	0	0	5	0	0	1	0	2	
5	2	0	0	0	4	2	2	0	0	0	4	0	0	0	0	3	2	5	2	6	0	5	0	0	2	2	0	2	0	
6	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	3	6	5	3	4	0	0	6	0	0	5	5	0	6	
7	2	0	0	0	2	0	2	0	0	0	0	0	0	0	0	2	3	6	5	4	5	0	0	3	0	0	2	2	0	
8	0	4	0	5	0	0	0	0	0	4	0	0	0	4	0	3	4	0	4	4	0	2	0	2	0	0	2	0	2	
9	0	3	0	3	0	0	0	0	0	0	6	0	0	4	4	2	2	0	4	5	0	0	2	2	0	0	3	0	2	
10	4	0	0	0	4	2	4	0	0	0	0	0	0	0	0	3	6	3	4	4	0	0	0	0	0	6	1	2	0	
11	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	5	7	3	4	0	0	0	0	2	0	0	3	2	4	
12	0	4	0	8	0	0	0	0	0	0	0	0	0	4	2	0	0	4	0	0	4	0	0	2	4	0	0	3	0	
13	0	5	0	5	0	0	0	1	0	0	0	0	0	0	0	2	0	0	2	9	0	2	0	2	0	0	2	0	0	
14	0	4	0	0	0	0	0	0	4	0	0	0	0	3	2	0	0	0	0	2	2	0	0	0	0	6	0	0	4	
15	0	4	0	0	0	0	0	0	5	0	0	0	3	0	2	0	0	0	0	2	2	0	0	0	0	1	0	4	0	
16	0	7	0	4	0	0	0	0	8	0	0	0	3	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	4	
17	2	0	2	0	4	0	0	0	0	0	0	0	2	0	0	2	0	2	1	0	0	0	0	0	0	0	0	0	0	
18	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	5	2	3	3	0	0	0	0	0	3	0	2	1	6	
19	0	0	3	0	4	0	0	0	0	0	2	5	0	0	6	0	0	0	0	0	0	0	0	3	3	0	2	0	3	
20	0	5	0	0	0	0	0	0	4	0	0	4	0	0	4	2	0	0	0	0	0	4	0	0	5	0	0	0	4	

Aller à la carte «Dune+1» de la pile ADE•Data (Fig. 1). Faire un fichier TXT avec le champ de valeurs numériques. Transformer ce fichier en binaire sous le nom VegTR (30-20) par l'option **Char->Binary** du module **TextToBin**. Transposer ce fichier sous le nom Veg (20-30) à l'aide de l'option **Transpose** du module **FilesUtil**. Utiliser la

commande **Edit with** du menu **Data Folder** qui crée un fichier Veg- t. Choisir Excel™ comme logiciel d'accès pour obtenir le tableau de végétation ci-dessus.

A ces 20 relevés correspondent des mesures de variables environnementales donnant un fichier Mi1.txt (20-8) obtenu directement dans la carte «Dune+1» de la pile **ADE•Data**. La procédure habituelle fournit le fichier Mi1 (20-8) que l'on peut lister :

Input file: Mi1
Row: 20 Col: 8

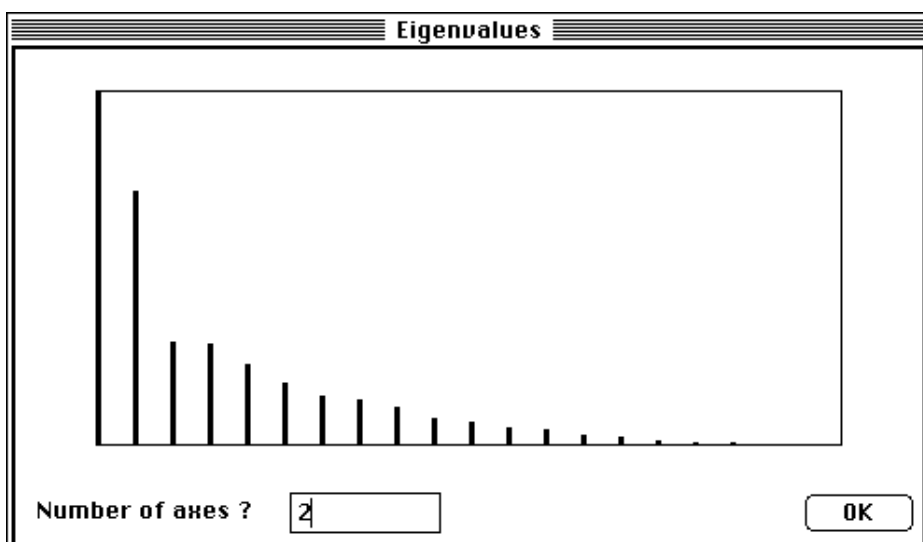
1	2.8	1.0	2.0	4.0	1.0	0.0	0.0	0.0
2	3.5	1.0	2.0	2.0	0.0	1.0	0.0	0.0
3	4.3	2.0	2.0	4.0	1.0	0.0	0.0	0.0
4	4.2	2.0	2.0	4.0	1.0	0.0	0.0	0.0
5	6.3	1.0	1.0	2.0	0.0	0.0	1.0	0.0
6	4.3	1.0	2.0	2.0	0.0	0.0	1.0	0.0
7	2.8	1.0	3.0	3.0	0.0	0.0	1.0	0.0
8	4.2	5.0	3.0	3.0	0.0	0.0	1.0	0.0
9	3.7	4.0	1.0	1.0	0.0	0.0	1.0	0.0
10	3.3	2.0	1.0	1.0	0.0	1.0	0.0	0.0
11	3.5	1.0	3.0	1.0	0.0	1.0	0.0	0.0
12	5.8	4.0	2.0	2.0	1.0	0.0	0.0	0.0
13	6.0	5.0	2.0	3.0	1.0	0.0	0.0	0.0
14	9.3	5.0	3.0	0.0	0.0	0.0	0.0	1.0
15	11.5	5.0	2.0	0.0	0.0	0.0	0.0	1.0
16	5.7	5.0	3.0	3.0	1.0	0.0	0.0	0.0
17	4.0	2.0	1.0	0.0	0.0	0.0	0.0	1.0
18	4.6	1.0	1.0	0.0	0.0	0.0	0.0	1.0
19	3.7	5.0	1.0	0.0	0.0	0.0	0.0	1.0
20	3.5	5.0	1.0	0.0	0.0	0.0	0.0	1.0

La variable 1 est l'**épaisseur du sol** (cm), la variable 2 mesure l'**humidité du sol** (5 classes ordonnées), la variable 3 donne l'**intensité d'utilisation de la végétation** (3 modalités ordonnées : 1-fauche, 2-mixte, 3-pâturage), la variable 4 donne la **quantité d'engrais** apportée (4 classes ordonnées) et les variables 5 à 8 codent en disjonctif complet le **mode d'aménagement** (5 : agriculture traditionnelle, 6 : agriculture biologique, 7 : hobby farming, 8 : conservation de la nature).

La situation expérimentale est typique. Si on se pose la question "Quelle combinaison d'espèces rend-elle compte de telle ou telle variable de milieu ?" la réponse statistique est que 30 espèces pour 20 relevés rendent parfaitement compte d'une série arbitraire de valeurs. L'analyse canonique des deux tableaux n'a pas de sens pour la même raison. La question duale "Quelle combinaison de variables rend-elle compte de l'abondance des espèces ?" ne bute pas immédiatement sur le même obstacle mais il suffirait de rendre qualitatives toutes les variables pour qu'il en soit ainsi. L'analyse canonique des correspondances (Ter Braak, 1986)² prend en compte cet aspect des choses mais lorsque le nombre de variables explicatives est assez important on se rapproche d'une simple analyse des correspondances du tableau de végétation. Dans cette fiche nous commencerons par l'analyse de la structure de chacun des tableaux puis nous étudierons la présence éventuelle d'une co-structure.

2 - Structure du tableau de végétation

Effectuer l'ACP centrée du tableau Veg en utilisant l'option **Covariance matrix PCA** du menu **PCA**. La répartition de l'inertie montre l'existence de deux facteurs utiles :



La carte factorielle des relevés (au demeurant voisine de celle de l'AFC du même tableau, cf. op. cit. p. 102) suggère l'existence d'une partition. Utiliser l'option **Labels** du module **Scatters** pour obtenir la Fig. 2.

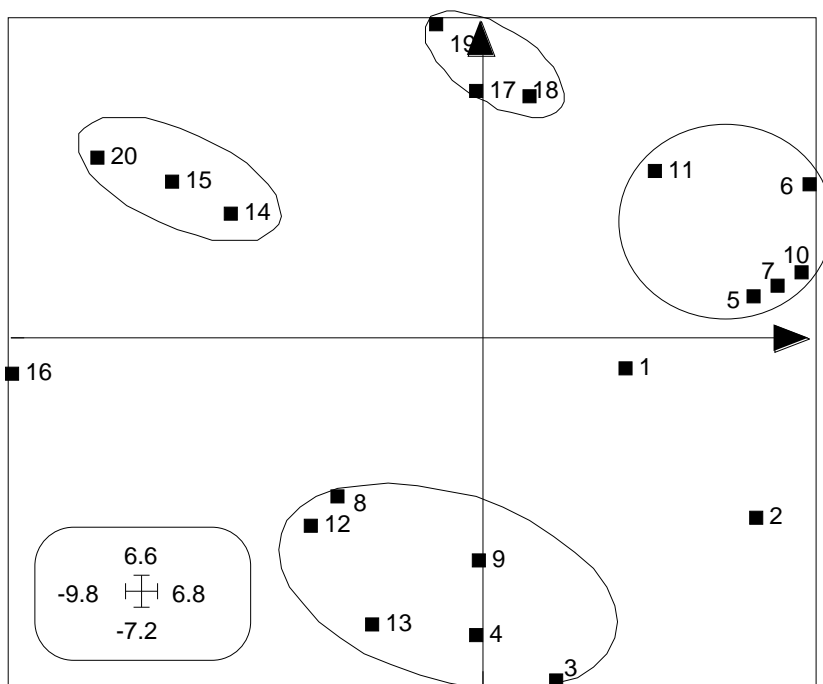


Figure 2 ACP du tableau de végétation.

Sur cette illustration, les ellipses ont été ajoutées sous ClarisDraw™.

La répartition de l'abondance des espèces (Fig. 3) indique plutôt un changement continu de composition floristique. Utiliser l'option **Values** du module **Scatters** pour obtenir cette illustration.

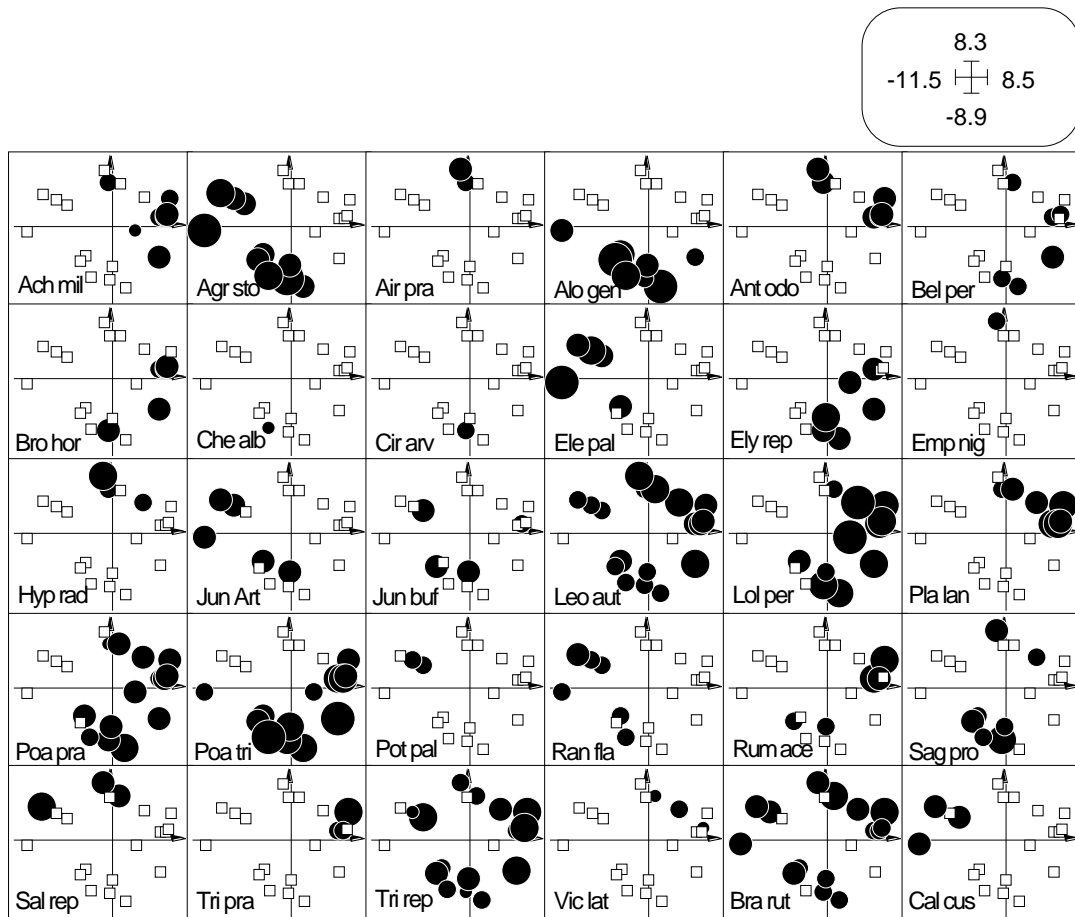


Figure 3 Distribution d'abondance des 30 espèces floristiques sur la carte factorielle de l'ACP. L'identification des espèces par une étiquette est faite sous ClarisDraw™. Les carrés blancs indiquent l'absence d'un taxon.

Cette analyse simple souligne l'existence d'une typologie de relevés à partir de leur contenu floristique.

3 - Structure du tableau de milieu

Faire l'ACP centrée du tableau Mi1 à l'aide de l'option **Covariance matrix PCA** du module **PCA**. On note dans le listing la répartition de la variance entre les variables de la manière suivante :

```

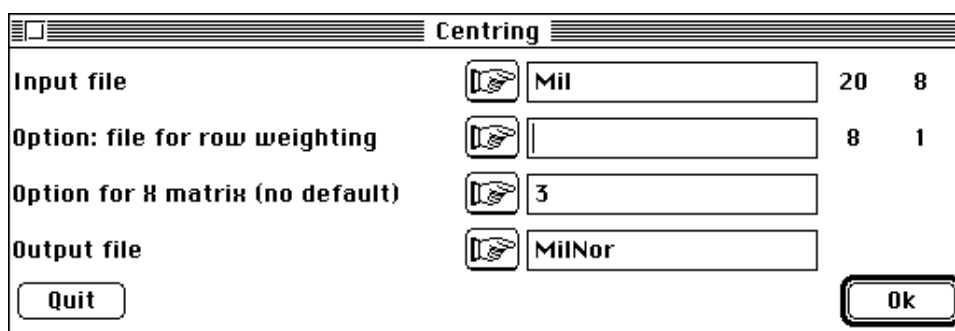
---- Info: means and variances
File Mi1.cpma contains the descriptive of the analysis
It contains successively:
Number of rows: 20
Number of columns: 8
means and variances:
Col.: 1 Mean: 4.85 Variance: 4.5115
Col.: 2 Mean: 2.9 Variance: 3.09
Col.: 3 Mean: 1.9 Variance: 0.59
Col.: 4 Mean: 1.75 Variance: 2.0875
Col.: 5 Mean: 0.3 Variance: 0.21
Col.: 6 Mean: 0.15 Variance: 0.1275
Col.: 7 Mean: 0.25 Variance: 0.1875
Col.: 8 Mean: 0.3 Variance: 0.21

```

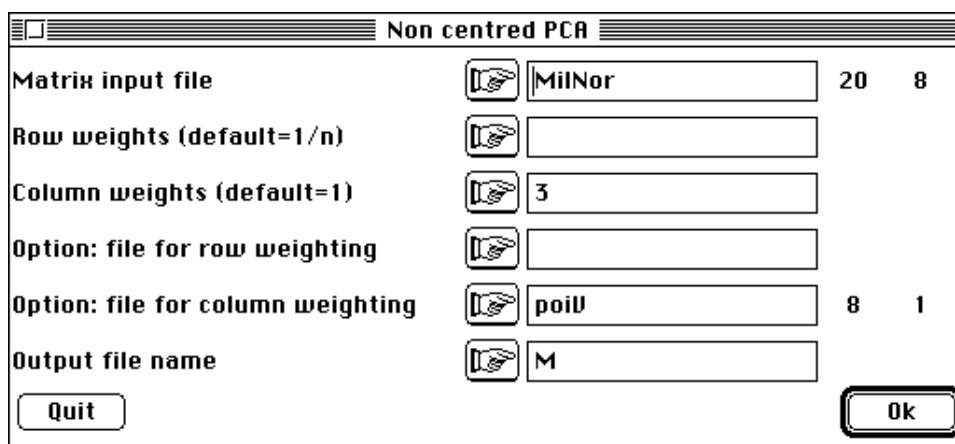
Cette analyse affecte aux quatre modalités (5,6,7,8), vues comme 4 variables, une inertie de l'ordre de grandeur de l'inertie de la variable 3. Les variables 1, 2 et 4 sont privilégiées. Les trois premiers facteurs de cette analyse seront ainsi liées aux variables 1, 2 et 4 qui ont une importance numérique très (trop) forte. Si on utilisait une ACP normée, ce serait le contraire : le type d'agriculture compterait quatre fois plus que les autres variables. Une pondération a priori peut résoudre ce problème. Créer un fichier texte PoiV/txt contenant les valeurs suivantes :

```
-----
Option: Excel - Binary
Input text file: poiV/txt
-----
Row 1      Words 1      : 1
Row 2      Words 1      : 1
Row 3      Words 1      : 1
Row 4      Words 1      : 1
Row 5      Words 1      : .25
Row 6      Words 1      : .25
Row 7      Words 1      : .25
Row 8      Words 1      : .25
8 numbers found in file poiV/txt
Row number: 8 Column number: 1
-----
Output binary file poiV
-----
```

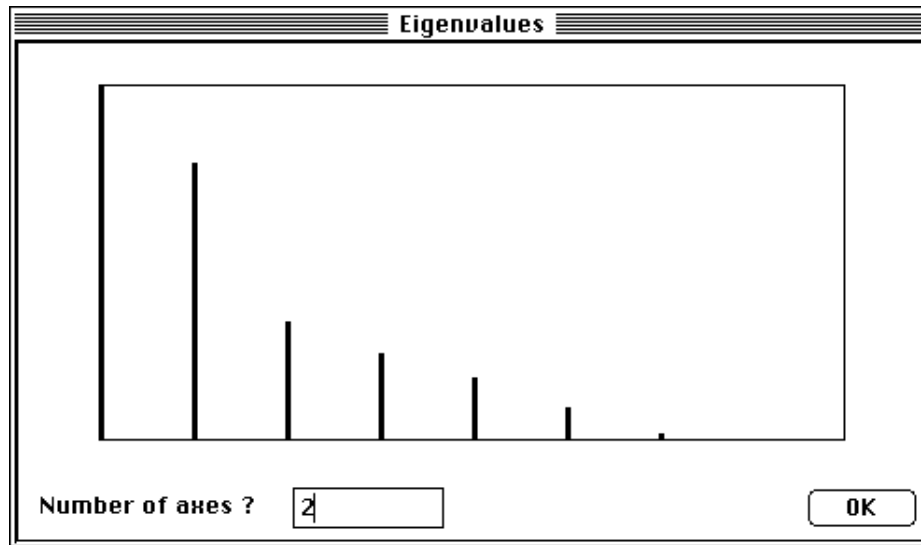
On équilibre ainsi l'importance des cinq variables expérimentales. Transformer ce fichier en binaire PoiV (8-1). Normaliser le tableau Mi1 (20-8) en un tableau Mi1Nor (20-8) à l'aide de l'option **Centring** du module **Bin->Bin** :



Faire l'ACP non centrée de Mi1Nor en tenant compte de la pondération externe des colonnes (PoiV). Pour cela, utiliser l'option **Non centred PCA** du module **PCA** :



On note que ce programme d'ACP générale permet de faire des ACP centrée, normée ou à pondération quelconque ce qui souligne la flexibilité du logiciel ADE. Une valeur propre nulle liée aux variables 5 à 8 qui sont des modalités disjonctives et une inertie totale de 1 sont à noter dans cette analyse. Le graphe des valeurs propres est le suivant :



Le listing obtenu est :

```

General Principal Component Analysis (Lebart & al. 1982)
Input file: MilNor
---- Row weights:
File M.ncpl contains the row weights
It has 20 rows and 1 column
Each row has      0.05 weight (Sum = 1)
---- Column weights:
File M.ncpc contains the column weights
It has 20 rows and 1 column
Column weights are selected from file poiV (Sum = 1)
---- Table:
File M.ncta contains the table duplicated from MilNor
It has 20 rows and 8 columns
-----
DiagoRC: General program for two diagonal inner product analysis
Input file: M.ncta
--- Number of rows: 20, columns: 8
-----
Total inertia:          1
-----
Num. Eigenval.   R. Iner.   R. Sum   | Num. Eigenval.   R. Iner.   R. Sum   |
01  +3.7746E-01  +0.3775  +0.3775  | 02  +2.9560E-01  +0.2956  +0.6731  |
03  +1.2705E-01  +0.1271  +0.8001  | 04  +9.2053E-02  +0.0921  +0.8922  |
05  +6.5739E-02  +0.0657  +0.9579  | 06  +3.5313E-02  +0.0353  +0.9932  |
07  +6.7831E-03  +0.0068  +1.0000  | 08  +0.0000E+00  +0.0000  +1.0000  |
-----
File M.ncvp contains the eigenvalues and relative inertia for each
axis
--- It has 8 rows and 2 columns

File M.ncco contains the column scores
--- It has 8 rows and 2 columns

File M.ncli contains the row scores
--- It has 20 rows and 2 columns

```

On conserve encore deux axes. La carte des variables (option **Labels** du module **Scatters**) associe les fortes valeurs de l'épaisseur (1) et de l'humidité du sol (2) et oppose usage d'engrais (4), utilisation forte de la végétation (3) et agriculture traditionnelle (5) d'une part à aménagement pour la conservation de la nature (8) de l'autre (Fig. 4).

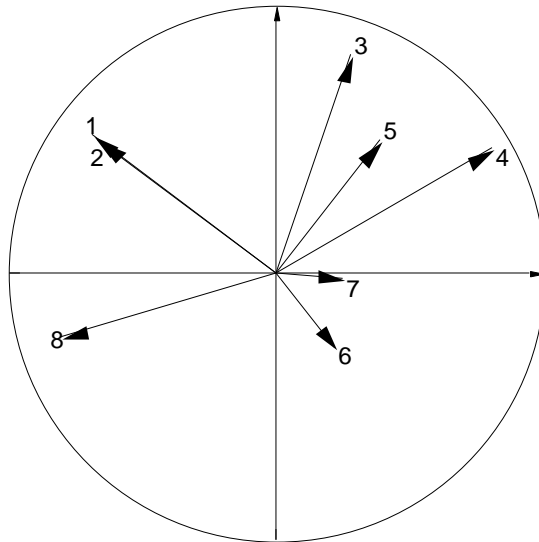


Figure 4 Cercle des corrélations des variables environnementales avec les facteurs de l'ACP centrée.

La carte des relevés (Fig. 5) est également obtenue par l'option **Labels** du module **Scatters**.

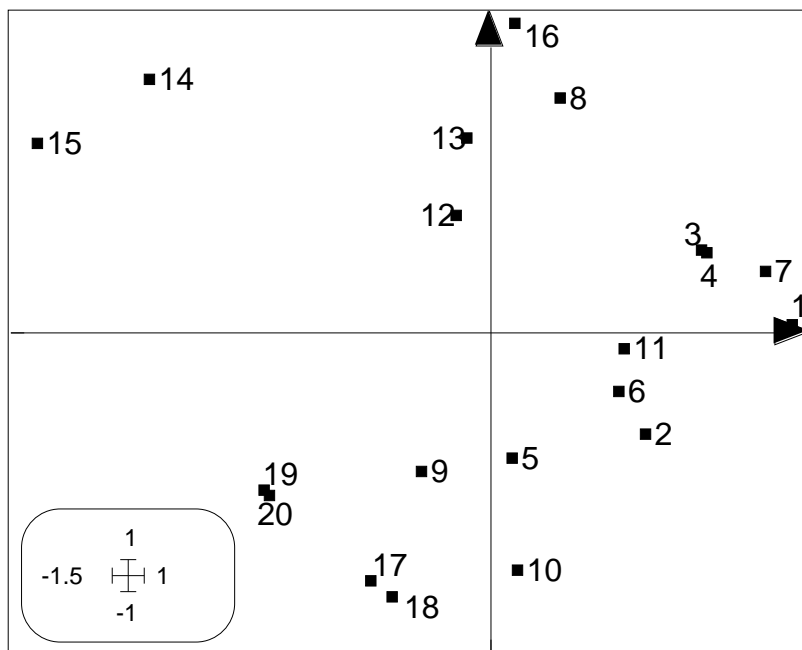


Figure 5 Carte des relevés de l'ACP du tableau milieu.

Cette carte des relevés de milieu rappelle certains points de la carte des relevés de végétation (proximité des points 14, 15 et des points 17, 18, 19 par exemple) tout en offrant des différences sensibles (le point 20 est avec le second groupe (17, 18, 19) dans l'analyse du milieu et avec le premier groupe (14, 15) dans l'analyse de la végétation par exemple).

L'objet de l'analyse de co-inertie va être de construire dans la logique de chaque analyse des cartes factorielles des relevés ayant une signification tout en optimisant la lecture simultanée.

4 - Co-inertie des deux nuages

4.1 - Le modèle mathématique

Le modèle mathématique de l'analyse de co-inertie peut être explicité sous l'angle du schéma de dualité synthétisé par Cailliez & Pagès (1976)³ et introduit en écologie statistique par Escoufier (1987)⁴. On trouvera une revue détaillée du modèle général de l'analyse de co-inertie dans Chessel & Mercier (1993)⁵. Les éléments suivants donnent les principaux repères de ce modèle général. Ils sont extraits d'un article de Dolédec & Chessel (1994)⁶. Si l'on utilise les notations du schéma de dualité, on peut justifier le terme de co-inertie de la manière suivante :

$(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ et $(\mathbf{Y}, \mathbf{D}_q, \mathbf{D}_n)$ sont deux triplets statistiques. \mathbf{X} est le tableau des données environnementales (après centrage initial). \mathbf{D}_p est la matrice des poids associés aux colonnes de \mathbf{X} . \mathbf{D}_n est la matrice des poids associés aux lignes de \mathbf{X} . \mathbf{Y} est le tableau floristique (après centrage initial). \mathbf{D}_q est la matrice des poids associés aux colonnes de \mathbf{Y} . \mathbf{D}_n est la matrice des poids associés aux lignes de \mathbf{Y} (cf. Dolédec & Chessel, 1991)⁷.

Le triplet $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ définit une analyse d'inertie des n points de \mathbb{R}^p et des p points de \mathbb{R}^n . Après diagonalisation (recherche de valeurs singulières), r axes principaux sont conservés et les matrices $\mathbf{R}_r, \mathbf{C}_r, \mathbf{N}_r$ sont générées. \mathbf{R}_r contient les coordonnées des n lignes pour les r axes. \mathbf{C}_r contient les coordonnées des p colonnes pour les r axes. \mathbf{N}_r contient les valeurs propres ($\lambda_1 \dots \lambda_r$). Le triplet $(\mathbf{Y}, \mathbf{D}_q, \mathbf{D}_n)$ définit une analyse d'inertie des n points de \mathbb{R}^q et des q points de \mathbb{R}^n . Après diagonalisation (recherche de valeurs singulières), s axes principaux sont conservés et les matrices $\mathbf{R}_s, \mathbf{C}_s, \mathbf{M}_s$ sont générées. \mathbf{R}_s contient les coordonnées des n lignes pour les s axes. \mathbf{C}_s contient les coordonnées des q colonnes pour les s axes. \mathbf{M}_s contient les valeurs propres ($\mu_1 \dots \mu_s$).

Considérons deux vecteurs \mathbf{u} et \mathbf{v} . Le premier est normalisé par \mathbf{D}_p dans \mathbb{R}^p et le second est normalisé par \mathbf{D}_q dans \mathbb{R}^q . La projection de l'espace multidimensionnel associé au tableau \mathbf{X} sur le vecteur \mathbf{u} génère n coordonnées dans une matrice colonne :

$$= \mathbf{X} \mathbf{D}_p \mathbf{u} \quad (1).$$

La projection de l'espace multidimensionnel associé au tableau \mathbf{Y} sur le vecteur \mathbf{v} génère n coordonnées dans une matrice colonne :

$$= \mathbf{Y} \mathbf{D}_q \mathbf{v} \quad (2).$$

La co-inertie associée à \mathbf{u} et \mathbf{v} est égale à :

$$H(\mathbf{u}, \mathbf{v}) = {}^t \mathbf{D}_n \quad (\text{t for transposed}) \quad (3).$$

Si les deux tableaux \mathbf{X} et \mathbf{Y} sont préalablement centrés alors la co-inertie est la covariance entre les deux nouveaux jeux de coordonnées :

$$\text{Cov} \left(\begin{matrix} \text{ } \\ \text{ } \end{matrix} \right) = (\text{Iner}_1(\mathbf{u}))^{\frac{1}{2}} (\text{Iner}_2(\mathbf{v}))^{\frac{1}{2}} \text{Corr} \left(\begin{matrix} \text{ } \\ \text{ } \end{matrix} \right) \quad (4),$$

avec $\text{Iner}_1(\mathbf{u})$ inertie projetée sur \mathbf{u} ou encore variance des nouvelles coordonnées sur \mathbf{u} , $\text{Iner}_2(\mathbf{v})$ inertie projetée sur \mathbf{v} ou encore variance des nouvelles coordonnées sur \mathbf{v} , et

$Corr(\cdot, \cdot)$ corrélation entre les deux nouveaux jeux de coordonnées. On notera que c'est le carré de $Corr(\cdot, \cdot)$ qui est maximisé en analyse canonique des corrélations. Par contre un axe de co-inertie associé à deux vecteurs \mathbf{u} et \mathbf{v} maximise la covariance $Cov(\cdot, \cdot)$.

Les axes de co-inertie sont obtenus en diagonalisant \mathbf{W} :

$$\mathbf{W} = \mathbf{D}_p^{-\frac{1}{2}} \mathbf{X}^t \mathbf{D}_n \mathbf{Y} \mathbf{D}_q \mathbf{Y}^t \mathbf{D}_n \mathbf{X} \mathbf{D}_p^{-\frac{1}{2}} \quad (5),$$

Considérons \mathbf{U}_z la matrice contenant les z premiers vecteurs propres normalisés de \mathbf{W} et Λ_z la matrice contenant les z premières valeurs propres correspondantes ($\lambda_k, 1 \leq k \leq z$). Les z premiers axes de co-inertie (\mathbf{D}_p normés) dans \mathbb{R}^p et les z premiers axes de co-inertie (\mathbf{D}_q normés) dans \mathbb{R}^q résultent respectivement de :

$$\mathbf{A}_z = \mathbf{D}_p^{-\frac{1}{2}} \mathbf{U}_z \quad \text{et} \quad \mathbf{B}_z = \mathbf{Y}^t \mathbf{D}_n \mathbf{X} \mathbf{D}_p^{-\frac{1}{2}} \mathbf{U}_z \quad \Lambda_z^{-\frac{1}{2}} \quad \text{avec} \quad \mathbf{Y}^t \text{ matrice transposée de } \mathbf{Y}. \quad (6).$$

\mathbf{A}_z et \mathbf{B}_z sont les poids de co-inertie des variables optimaux respectivement dans les tableaux \mathbf{X} et \mathbf{Y} . Les coordonnées des lignes de \mathbf{X} et des lignes de \mathbf{Y} dans la co-inertie sont respectivement :

$$\mathbf{X}_z^* = \mathbf{X} \mathbf{D}_p \mathbf{A}_z \quad \text{et} \quad \mathbf{Y}_z^* = \mathbf{Y} \mathbf{D}_q \mathbf{B}_z \quad (7).$$

On peut utilement comparer la variabilité projetée des analyses d'inertie de chaque tableau et celles obtenues par l'analyse de co-inertie. Pour cela, on calcule les coordonnées des projections des axes d'inertie sur les axes de co-inertie. Appelons \mathbf{C}_r^* et \mathbf{C}_s^* les matrices de ces coordonnées :

$$\mathbf{C}_r^* = \mathbf{N}_r^{-\frac{1}{2}} \mathbf{C}_r^t \mathbf{D}_p \mathbf{A}_z \quad \text{et} \quad \mathbf{C}_s^* = \mathbf{M}_s^{-\frac{1}{2}} \mathbf{C}_s^t \mathbf{D}_q \mathbf{B}_z \quad (8).$$

Les éléments diagonaux des matrices $(\mathbf{X}_z^*)^t \mathbf{D}_n \mathbf{X}_z^*$ et $(\mathbf{Y}_z^*)^t \mathbf{D}_n \mathbf{Y}_z^*$ sont les pseudo-valeurs propres. Considérons λ_k^* les k pseudo-valeurs propres de \mathbf{X} et μ_k^* les k pseudo-valeurs propres de \mathbf{Y} . Ces valeurs sont utilisées pour situer les axes d'inertie des analyses séparés sur les axes de co-inertie ; ceci permet d'interpréter quelle part des structures de chacun des tableaux est prise en compte dans la co-structure. Enfin la quantité :

$$\frac{\lambda_k^{*2}}{\lambda_k^* \mu_k^*} \quad (9)$$

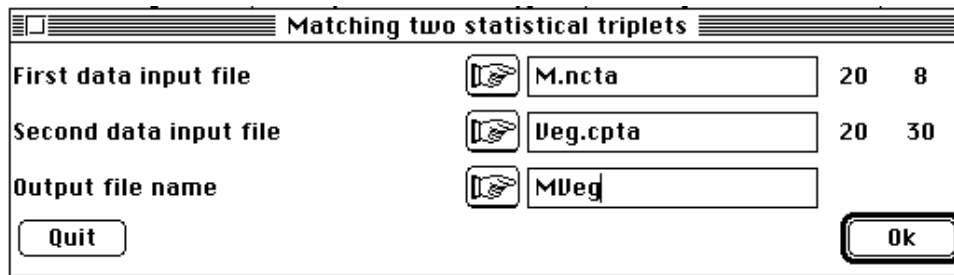
est l'expression de la corrélation entre les deux nouveaux jeux de coordonnées.

4.2 - Application

Utiliser le module **CoInertia** et la première option **Matching two statistical triplets** :



Le dialogue d'entrée demande de coupler deux analyses soit :



On veut ici coupler l'analyse du tableau M de type *nc* et l'analyse du tableau Veg de type *cp*. On déclare comme premier tableau celui qui a le moins de colonnes. M. ncpa est un fichier 20-8 et Veg. cpta est un fichier 20-30. Le couplage crée des fichiers de nom générique MVeg. <> qui symbolise une analyse symétrique des deux tableaux du couple :

```
First input statistical triplet: table M.ncta
Number of rows: 20, columns: 8
Second input statistical triplet: table Veg.cpta
Number of rows: 20, columns: 30
```

```
File MVeg.<>ta contains the crossed table YtDnX
It has 30 rows and 8 columns
```

```
File MVeg.<>pl contains the weight of each row
It has 30 rows and 1 column
```

```
File MVeg.<>pc contains the weight of each column
It has 8 rows and 1 column
```

```
File MVeg.<>ma contains the name of the input statistical
triplet
```

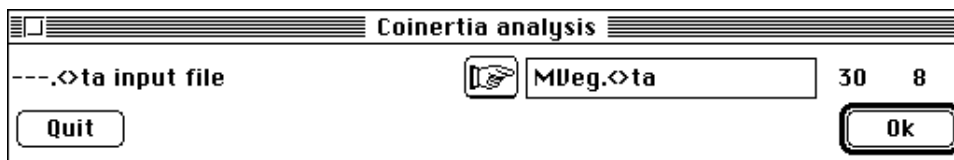
```
1 ---> M.ncta
2 ---> Veg.cpta
```

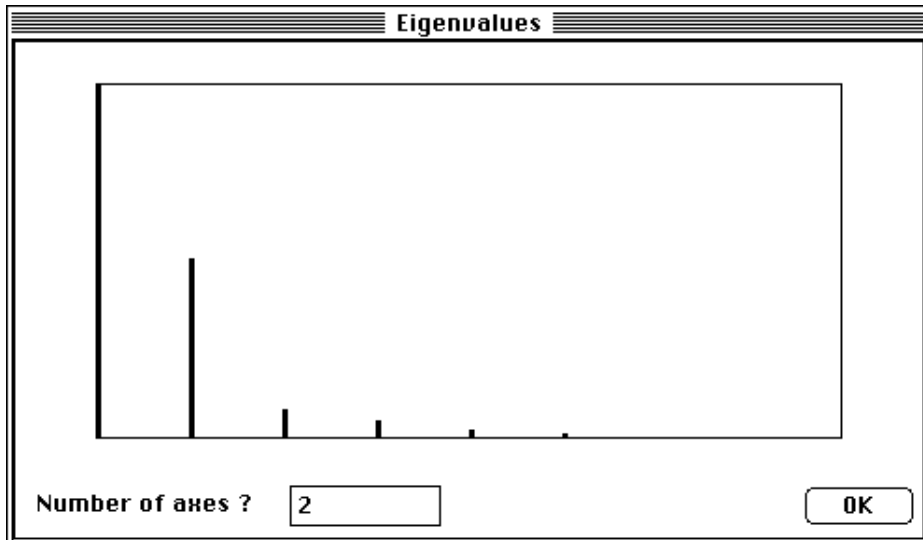
```
total inertia: 10.288411
```

Le tableau croisé MVeg. <>ta (30-8) est formé dans cet exemple des covariances entre l'abondance des 30 espèces et les valeurs des 8 variables environnementales. C'est une matrice de covariances inter-tableaux.

Utiliser l'option **Coinertia analysis** pour réaliser le calcul central :

Sélectionner le tableau croisé dans la boîte de dialogue correspondante :





Conserver deux axes à l'examen du graphe des valeurs propres. Le listing qui suit indique la création d'un grand nombre de fichiers (en gras dans le listing). C'est la nature très générale de la procédure qui permet de coupler tous les types d'analyse qui en est la cause. On utilise dans chaque cas une partie d'entre eux en fonction des types d'analyses couplées et des objectifs.

```
DiagoRC: General program for two diagonal inner product analysis
Input file: MVeg.<>ta
--- Number of rows: 30, columns: 8
-----
Total inertia: 10.2884
-----
Num. Eigenval.  R. Iner.  R. Sum  |Num. Eigenval.  R. Iner.  R. Sum |
01  +6.1196E+00 +0.5948 +0.5948 |02  +3.1236E+00 +0.3036 +0.8984 |
03  +5.0782E-01 +0.0494 +0.9478 |04  +2.9621E-01 +0.0288 +0.9766 |
05  +1.3522E-01 +0.0131 +0.9897 |06  +8.9604E-02 +0.0087 +0.9984 |
07  +1.6316E-02 +0.0016 +1.0000 |08  +0.0000E+00 +0.0000 +1.0000 |
File MVeg.<>vp contains the eigenvalues and relative inertia for each
axis
--- It has 8 rows and 2 columns
File MVeg.<>co contains the column scores
--- It has 8 rows and 2 columns
File :MVeg.<>co
|Col. |  Mi ni  |  Maxi  |
|-----|-----|
|  1  |-2.054e+00| 4.044e+00|
|  2  |-2.297e+00| 2.880e+00|
|-----|-----|
File MVeg.<>li contains the row scores
--- It has 30 rows and 2 columns
File :MVeg.<>li
|Col. |  Mi ni  |  Maxi  |
|-----|-----|
|  1  |-1.342e+00| 8.762e-01|
|  2  |-4.306e-01| 8.711e-01|
|-----|-----|
-----
coinertia analysis between two statistical triplets
  1 ---> M.ncta (rows: 20, col: 8, axes: 2, inertia: 1.000000)
  2 ---> Veg.cpta (rows: 20, col: 30, axes: 2, inertia: 79.917503)
Coinertia: 10.288400, RV coefficient: 0.632414

MVeg.<>w1 is a binary file with 8 rows and 2 columns
```

It contains the canonical weights of the variables of table 1
File : MVeg. <>w1
----- Minimum/Maximum:
Col.: 1 Mini = -0.83035 Maxi = 1.6346
Col.: 2 Mini = -1.2998 Maxi = 1.6296

MVeg. <>co is a binary file with 8 rows and 2 columns
It contains the coordinates of the columns of table 1
In earlier version of ADE its name is MVeg. <>c1
File : MVeg. <>co
----- Minimum/Maximum:
Col.: 1 Mini = -2.0541 Maxi = 4.0438
Col.: 2 Mini = -2.2972 Maxi = 2.8801

MVeg. <>w2 is a binary file with 30 rows and 2 columns
It contains the canonical weights of the variables of table 2
File : MVeg. <>w2
----- Minimum/Maximum:
Col.: 1 Mini = -0.54239 Maxi = 0.35421
Col.: 2 Mini = -0.24363 Maxi = 0.49289

MVeg. <>li is a binary file with 30 rows and 2 columns
It contains the coordinates of the columns of table 2
In earlier version of ADE its name is MVeg. <>c2
File : MVeg. <>li
----- Minimum/Maximum:
Col.: 1 Mini = -1.3418 Maxi = 0.87625
Col.: 2 Mini = -0.43058 Maxi = 0.87111

MVeg. <>l1 is a binary file with 20 rows and 2 columns
It contains the coordinates of the rows (table 1)
File : MVeg. <>l1
----- Minimum/Maximum:
Col.: 1 Mini = -0.77157 Maxi = 1.3551
Col.: 2 Mini = -0.88888 Maxi = 0.87913

MVeg. <>l2 is a binary file with 20 rows and 2 columns
It contains the coordinates of the rows (table 2)
File : MVeg. <>l2
----- Minimum/Maximum:
Col.: 1 Mini = -5.8236 Maxi = 8.7283
Col.: 2 Mini = -6.9894 Maxi = 6.3238

MVeg. <>m1 is a binary file with 20 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 : MVeg. <>m1
----- Minimum/Maximum:
Col.: 1 Mini = -1.3105 Maxi = 2.3016
Col.: 2 Mini = -1.659 Maxi = 1.6408

MVeg. <>m2 is a binary file with 20 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 : MVeg. <>m2
----- Minimum/Maximum:
Col.: 1 Mini = -1.2311 Maxi = 1.8451
Col.: 2 Mini = -1.7505 Maxi = 1.5838

MVeg. <>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 : MVeg. <>a1

----- Minimum/Maximum:
Col.: 1 Mini = -0.87031 Maxi = 0.43047
Col.: 2 Mini = 0.4572 Maxi = 0.80403

MVeg. <>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 : MVeg. <>a2

----- Minimum/Maximum:
Col.: 1 Mini = -0.94877 Maxi = 0.23231
Col.: 2 Mini = -0.90653 Maxi = -0.25425

Les fichiers w1 et c1 concernent les variables (colonnes) du premier tableau. Les fichiers w2 et l1 concernent les variables (colonnes) du second tableau. Les fichiers l1 et m1 concernent les relevés (lignes) du premier tableau. Les fichiers l2 et m2 concernent les relevés (lignes) du second tableau. Les fichiers a1 et a2 concernent les axes principaux des analyses séparées.

Une attention particulière est à porter au tableau suivant intégré au listing ci-dessus :

Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	2.474	0.3466	22.38	0.8882	0.3775	23.58
2	1.767	0.2871	15.94	0.8261	0.2956	16.94

L'analyse de co-inertie ne cherche plus des axes d'inertie pour chaque nuage de 20 points mais recherche des axes de co-inertie qui maximise la covariance des coordonnées des projections.

Les colonnes INER1 et INER2 rappellent les inerties maximales qu'on peut projeter sur des axes dans chaque espace (cf. obtenues par les analyses séparées). Les colonnes Vari an1 et Vari an1 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 (Correl a.) 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. Un test de signification est nécessaire avant le dépouillement.

5 - Test de permutation

Trois options de test sont disponibles dans le module **CoInertia analysis** :

<input checked="" type="checkbox"/> Coinertia test - Fixed D
<input type="checkbox"/> Coinertia test - Fixed Tab 1
<input type="checkbox"/> Coinertia test - Fixed Tab 2

On reliera très simplement les analyses d'inertie séparées et l'analyse de co-inertie en utilisant le module **Scatters** et son option **Labels** respectivement sur les fichiers a1 et a2 (Fig. 6).

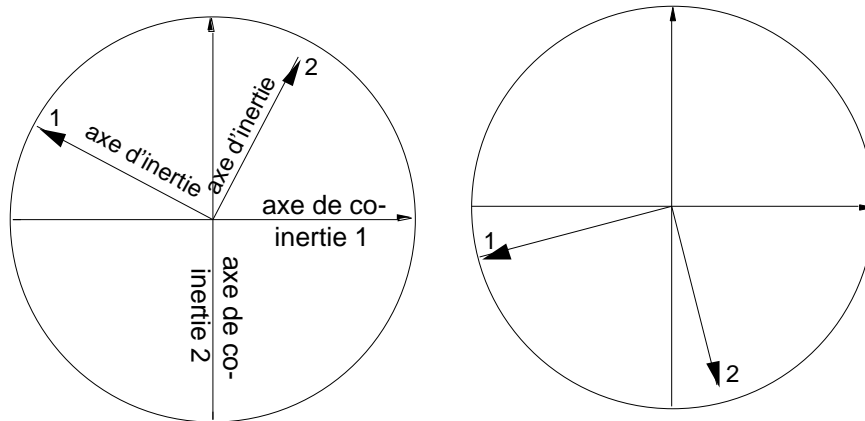


Figure 6 Projection des axes d'inertie des analyses séparées (milieu à gauche, végétation à droite) sur les axes de co-inertie.

Sur la Fig. 6, à gauche on a projeté les axes d'inertie 1 et 2 de l'ACP de Mi 1 Nor sur le plan des axes de co-inertie du même tableau. La carte des relevés positionnés par le milieu est sensiblement obtenue par symétrie et rotation de la carte d'ACP simple. A droite on a projeté les axes d'inertie 1 et 2 de l'ACP de Veg sur le plan des axes de co-inertie du même tableau. La carte des relevés positionnés par la végétation est sensiblement obtenue par symétrie et rotation de la carte d'ACP simple. Ce qui fait qu'on voit difficilement la ressemblance entre les deux systèmes de coordonnées.

On peut maintenant superposer les deux cartes de relevés en utilisant l'option **Match two scatters** du module **Scatters** :

Match two scatters			
HY coordinates file	<input type="button" value="👉"/>	MVeg.<0>m1	20 2
H-axis column number (default = 1)	<input type="button" value="👉"/>		
Y-axis column number (default = 2)	<input type="button" value="👉"/>		
Second HY coordinates file	<input type="button" value="👉"/>	MVeg.<0>m2	20 2
Label file (or #) for rows	<input type="button" value="👉"/>	##	
Label file (or #) for groups	<input type="button" value="👉"/>		
Quit		Copy graph	Save graph
		Print graph	Draw

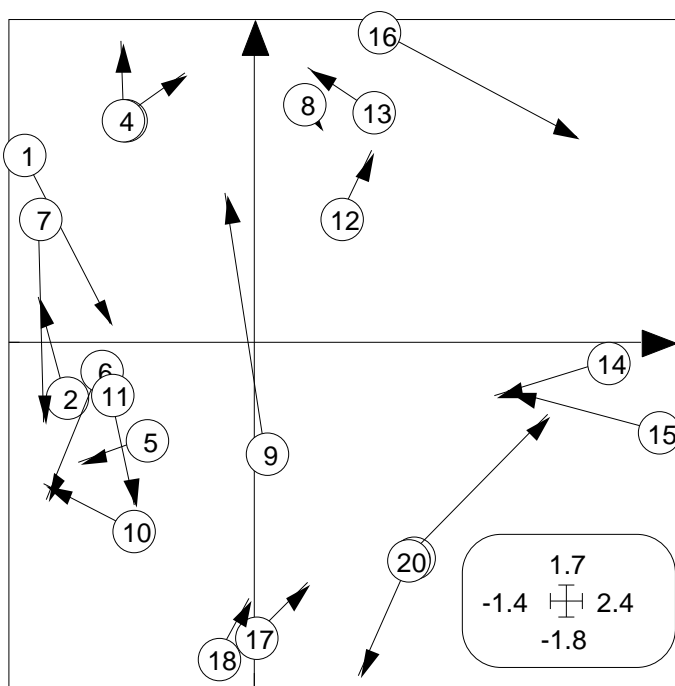


Figure 7 Projection simultanée en analyse de co-inertie. Le point portant une étiquette donne la position vue à travers le milieu (tableau 1). Le point à l'extrémité de la flèche donne la position vue par la végétation (tableau 2). Les variances sont unitaires sur chaque axe pour ramener les deux systèmes à la même échelle.

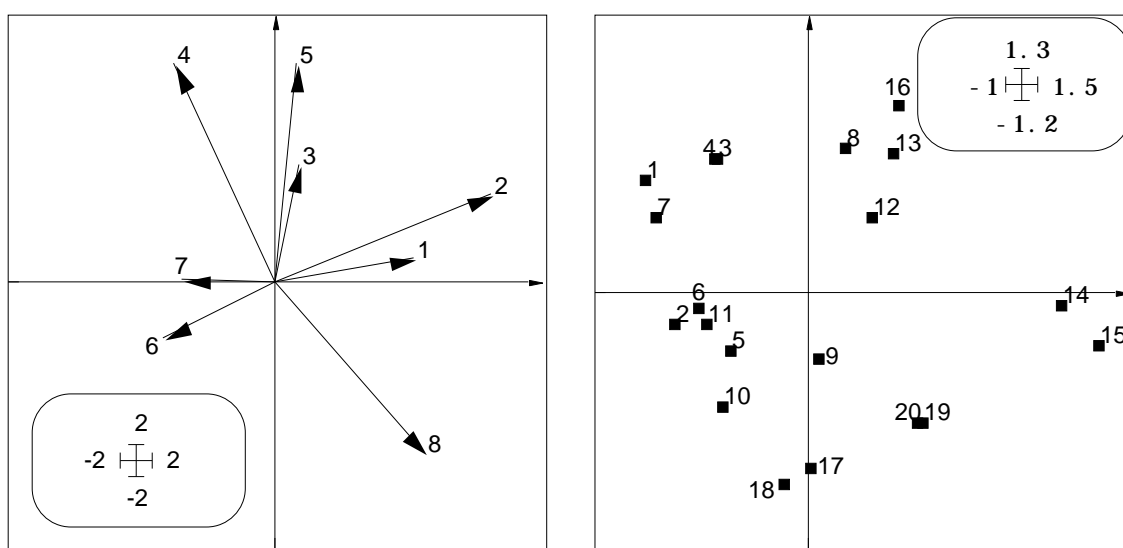


Figure 8 Projection du nuage de 20 points à 8 composantes dans l'analyse de co-inertie. En haut représentation de la base canonique, en bas le nuage des relevés associé. Les deux éléments forment un biplot classique d'ACP centrée. Mais l'inertie projetée est partiellement maximisée puisqu'on utilise d'autres éléments. On obtient sensiblement la carte de l'ACP initiale à une transformation près qui la connecte à la carte des relevés de végétation.

Les fichiers concernés contiennent les coordonnées normalisées des projections sur les axes de co-inertie, c'est-à-dire les coordonnées de variance unité. Ce changement d'échelle permet de superposer les deux cartes (Fig. 7) pour exprimer ce qui dans le caractère optimisé (covariance) concerne la corrélation. Les deux structures sont globalement les mêmes. On constate cependant pour certains relevés (9, 16 et 20)

quelques discordances entre les deux situations matérialisées par la longueur des flèches.

Chaque composante de la figure s'interprète de plus comme une ACP. On utilisera le fichier `Mveg. <>w1` pour placer les variables tableau 1 et `Mveg. <>11` pour placer les relevés (lignes) du tableau 1. On obtient un biplot (Gabriel, 1971⁹, 1981¹⁰) d'une ACP centrée traditionnelle (Fig. 8).

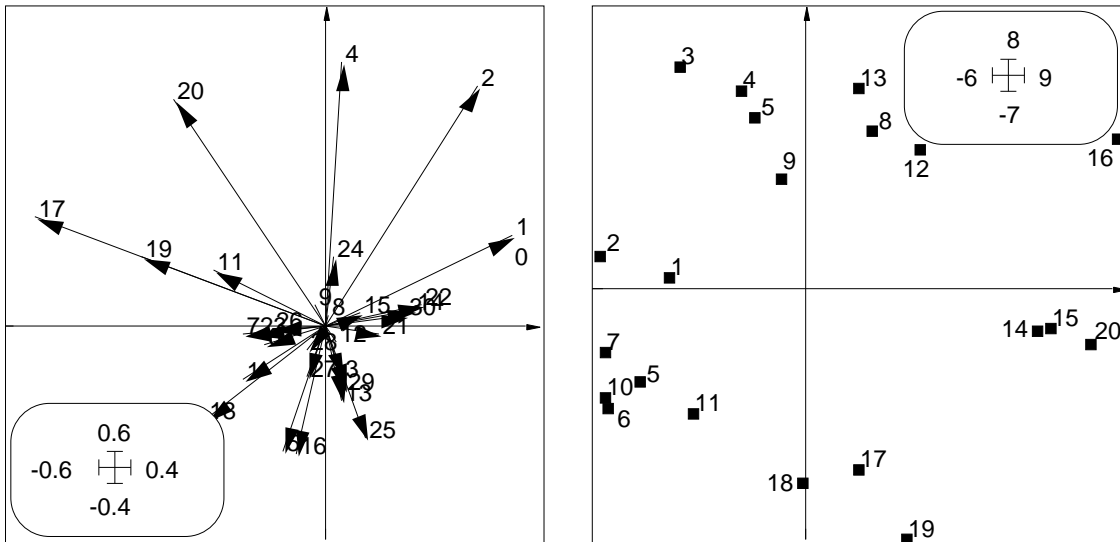
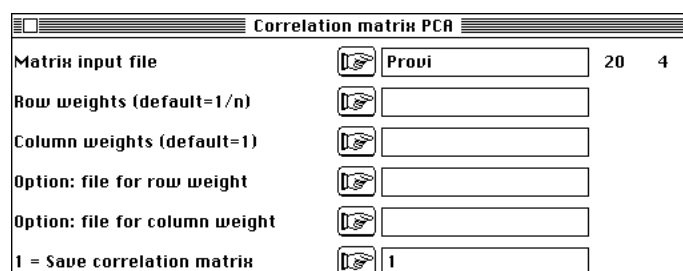


Figure 9 Projection du nuage de 20 points à 30 composantes dans l'analyse de co-inertie. En haut représentation de la base canonique, en bas le nuage des relevés associé. Les deux éléments forment un biplot classique d'ACP centrée. Mais l'inertie projetée est partiellement maximisée puisqu'on utilise d'autres éléments. On obtient sensiblement la carte de l'ACP initiale à une transformation près qui la connecte à la carte des relevés de milieu.

On exprime ainsi ce qui dans le caractère optimisé (covariance) concerne la variance dans le tableau 1 (milieu). On y retrouve l'interprétation des données de Ter Braak (op. cit. p. 140) : "We infer that the first axis is a moisture gradient and that the second axis is a manuring axis, separating the meadows managed as a nature reserve from the standarly farmed meadows". On utilisera le fichier `Mveg. <>w2` pour placer les espèces et `Mveg. <>1 2` pour placer les relevés (lignes) du tableau 2. On obtient un biplot d'une ACP centrée traditionnelle (Fig. 9). On exprime cette fois ce qui dans le caractère optimisé (covariance) concerne la variance dans le tableau 2 (végétation). On fait donc clairement deux ACP simultanées dans cette analyse de co-structure.

On notera que les coordonnées factorielles d'un même tableau ne sont pas nécessairement indépendantes. Pour vérifier cet élément et à titre de démonstration, accoler les tableaux `Mveg. <>1 1` et `Mveg. <>1 2` dans un fichier `Provi` à l'aide de l'option **PasteFiles-SameRow** du module **FilesUtil**. Puis, lancer le module **PCA** et cocher la case **Save correlation matrix** :



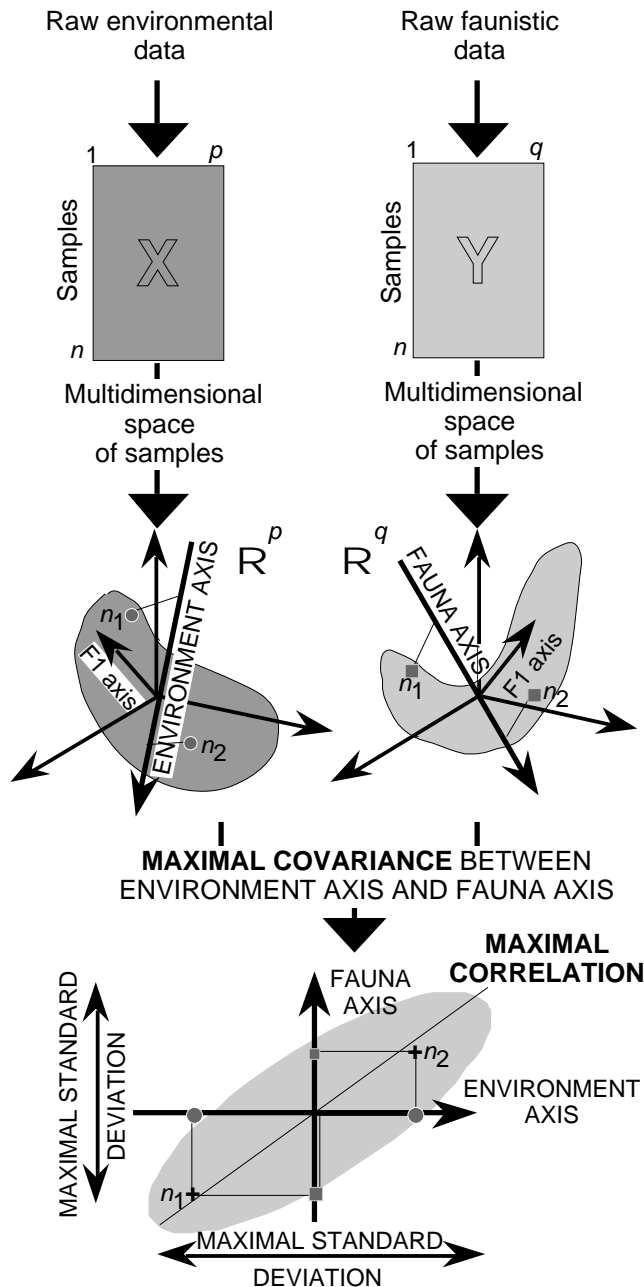


Figure 10 Schéma de principe des analyses de co-inertie.

La matrice des corrélations entre les quatre éléments de Provi peut être lue dans le listing :

```

----- Correlation matrix -----
[ 1] 1000
[ 2] -128 1000
[ 3] 888 0 1000
[ 4] -0 826 103 1000
-----

```

avec [1] = col 1 de MVeg.<>l1, [2] = col 2 de MVeg.<>l1, [3] = col 1 de MVeg.<>l2, [4] = col 2 de MVeg.<>l2.

On y observe une des propriétés fondamentales démontrées par Tucker¹¹ : les coordonnées 1 du tableau 1 sont non corrélées avec les coordonnées 2 du tableau 2 et

réciroquement. Par contre les coordonnées sur un même tableau ne sont pas sans corrélation comme d'ordinaire dans une ACP.

On retiendra de l'exemple le principe fondamental exprimé dans la Fig. 10 (Dolédec & Chessel, op. cit.). On trouvera une illustration récente du principe de la co-inertie se trouve dans l'article de Cadet & Coll.¹² et un exemple entre deux ACP à centrage multiplicatif dans ¹³.

Références

- 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 Ter Braak, C.J.F. (1986) Canonical correspondence analysis : a new eigenvector technique for multivariate direct gradient analysis. *Ecology* : 69, 69-77.
- ³ Cailliez, F. & Pages, J.P. (1976) Introduction à l'analyse des données. SMASH, 9 rue Duban, 75016 Paris. 1-616.
- ⁴ Escoufier, Y. (1987) The duality diagramm : a means of better practical applications. In : Development in numerical ecology. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute , Serie G .Springer Verlag, Berlin. 139-156.
- ⁵ 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.
- ⁶ Doledec, S. & Chessel, D. (1994) Co-inertia analysis: an alternative method for studying species-environment relationships. *Freshwater Biology* : 31, 277-294.
- ⁷ Doledec, S. & Chessel, D. (1991) Recent developments in linear ordination methods for environmental sciences. *Advances in Ecology, India* : 1, 133-155.
- ⁸ Kazi-Aoual, F., Hitier, S., Sabatier, R., Lebreton, J.D. (sous presse) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data analysis*.
- ⁹ Gabriel, K.R. (1971) The biplot graphical display of matrices with application to principal components analysis. *Biometrika* : 58, 453-467.
- ¹⁰ Gabriel, K.R. (1981) Biplot display of multivariate matrices for inspection of data and diagnosis. In : Interpreting multivariate data. Barnett, V. (Ed.) John Wiley and Sons, New York. pp. 147-174.
- ¹¹ Tucker, L.R. (1958) An inter-battery method of factor analysis. *Psychometrika* : 23, 2, 111-136.
- ¹² 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.
- ¹³ Devillers, J. & Chessel, D. (1995) Comparaison 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.