

CCA

Last update: 23/08/96

CCA : CCA.....	2
CCA : Initialize explanatory variables.....	7
CCA : Projected inertia decomposition.....	8

CCA: CCA



Méthode d'ordination dite Analyse Canonique des Correspondances (CCA: Ter Braak, 1986, 1987; AFCVI: Chessel, Lebreton & Yoccoz, 1987)



L'option est un module complet de CANOCO (Ter Braak, 1988), à l'exception des options de detrending qui sont hors du champ du modèle euclidien (voit le volume 5 de la documentation thématique).



L'option utilise une seule fenêtre de dialogue :

CCA

Explanatory variables: .@ob file

Dependant variables: .**ta

Output file name

Quit Ok

Nom du fichier .@ob contenant une base orthonormée du sous-espace de projection. L'option CCA : Initialize explanatory variables doit avoir été utilisée auparavant pour obtenir ce fichier à partir de l'ACP du tableau des variables de milieu.

Nom du fichier ---.##ta associé à l'analyse d'un tableau. Le module COA doit avoir été utilisé pour obtenir ce fichier.

Nom générique des fichiers de sortie (création).



Le fichier .@ob est toujours associé à un fichier .@pl qui contient la pondération pour laquelle la base utilisée est orthonormée. Le fichier ---.##ta est toujours associé à un fichier ---.##pl qui contient la pondération des lignes dans l'analyse préliminaire utilisée. Une seule contrainte est imposée : ces deux pondérations doivent être identiques. L'ACP initiale doit donc être réalisée en utilisant comme pondération des lignes la pondération de l'AFC (fichier ---.fcpl).



Reprendre l'exemple de la carte Banyuls de la pile ADE-4•Data. Après l'exécution de l'option CCA : Initialize explanatory variables, l'analyse canonique des correspondances du couple Mil/Veg est simplement obtenue par :

CCA

Explanatory variables: .@ob file Mil.@ob 12 4

Dependant variables: .**ta Veg.fcta 12 22

Output file name A 22

Quit Ok

Instrumental variables

----- input -----

Orthonormal basis: Mil.@ob

It has 12 rows and 4 columns

Dependant variable file: Veg.fcta

It has 12 rows and 22 columns

----- output -----

Projected variable file: A.ivta

It has 12 rows and 22 columns

Inertia: 8.5923e-01

File A.ivpc contains the column weights

It has 22 rows and 1 column

File A.ivpl contains the row weights

It has 12 rows and 1 column

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+4.1430E-01	+0.4822	+0.4822	02	+2.0203E-01	+0.2351	+0.7173
03	+1.3766E-01	+0.1602	+0.8775	04	+1.0524E-01	+0.1225	+1.0000

File A.ivvp contains the eigenvalues and relative inertia for each axis
It has 12 rows and 2 columns

On peut dépouiller en utilisant une partie des fichiers créés :

```
-----
files A.ivfa
      A.ivl1
      A.ivco
allow a convenient interpretation
-----
```

On cherche des poids des variables de milieu pour fabriquer des combinaisons linéaires des variables explicatives (Mil) de variance unité. Ces poids sont dans ---.ivfa :

File A.ivfa contains weights
coefficients of linear combination of explanatory variables with unit norm
It has 4 rows and 2 columns

	1	2
1	-0.5310	1.2896
2	0.3828	0.9871
3	0.1746	-0.0234
4	-0.0126	-0.5041

File :A.ivfa

Col.	Mini	Maxi
1	-5.310e-01	3.828e-01
2	-5.041e-01	1.290e+00

La première combinaison linéaire des variables de Mil (première variable canonique) est un code numérique des relevés qu'on trouve dans ---.ivl1 :

File A.ivl1 contains canonical row scores
linear combination of explanatory variables with unit norm
It has 12 rows and 2 columns

File :A.ivl1

Col.	Mini	Maxi
1	-2.961e+00	1.476e+00
2	-2.035e+00	2.276e+00

Chaque espèce se positionne par averaging sur ce gradient. Les positions moyennes par espèces sont dans ---.ivco :

File A.ivco contains column scores (dependant variables) with lambda norm
It has 22 rows and 2 columns

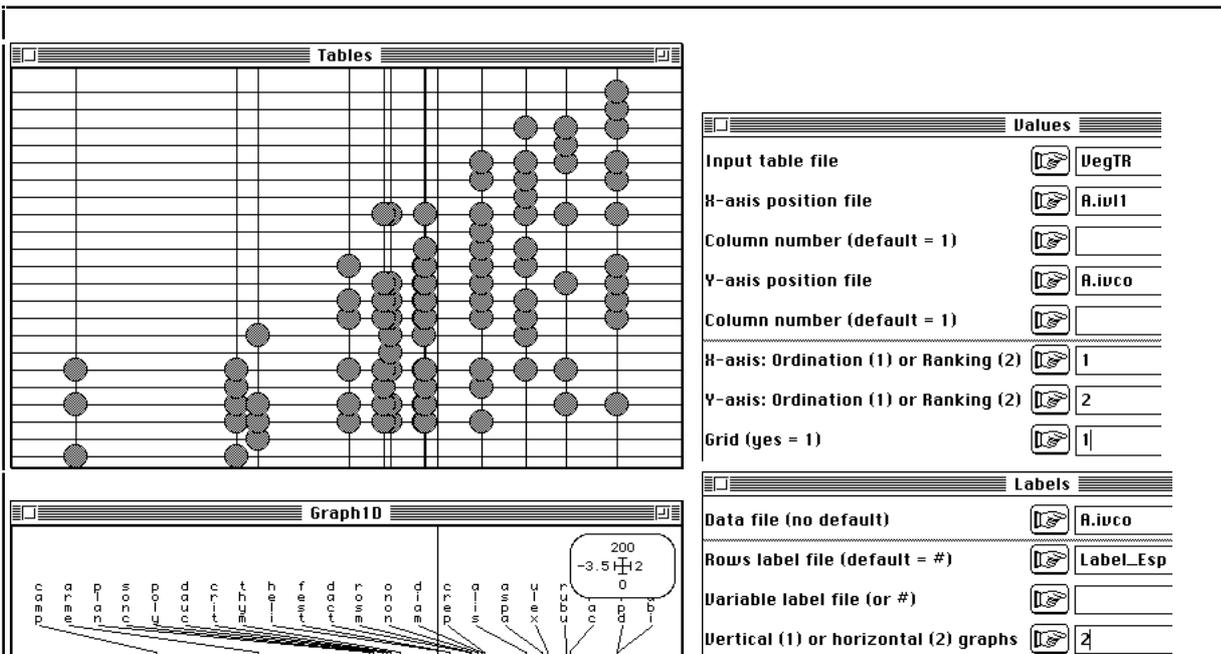
File :A.ivco

Col.	Mini	Maxi
1	-2.304e+00	1.476e+00
2	-2.035e+00	1.291e+00

Pour illustrer la démarche, transposer le tableau de végétation :

The image shows a software interface with three windows:

- Transpose**: A dialog box with 'Input file' set to 'veg' (dimensions 12 22) and 'Output file' set to 'vegTR'.
- Labels**: A dialog box with 'Data file (no default)' set to 'A.ivl1', 'Rows label file (default = #)' set to '#', 'Variable label file (or #)' set to an empty field, and 'Vertical (1) or horizontal (2) graphs' set to '2'.
- Graph1D**: A window showing a plot of 22 points (numbered 1-22) on a coordinate system. A small inset shows a 2x2 matrix: $\begin{pmatrix} 200 & -3.5 \\ -3.5 & 12 \\ 0 & 0 \end{pmatrix}$.



Les poids des variables de milieu ont été choisis pour maximiser la variance des moyennes par espèce. Ce maximum est la première valeur propre, soit 0.48 ou encore 48% de variance initiale. On repère ainsi le gradient de salinité structurant la communauté végétale. C'est le point de vue AFC sous contrainte où la coordonnée des relevées doit être une variable de milieu synthétique.

Il existe un raisonnement symétrique moins connu :

```
-----
files A.ivc1
      A.ivls
      A.ivli
allow a convenient interpretation
-----
```

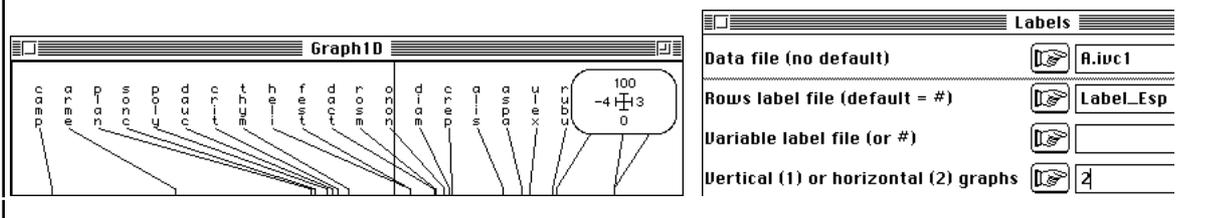
On part d'un score des espèces de variance unité pour la pondération des taxons dans le tableau Veg. Ce score est dans le fichier ---.ivc1 :

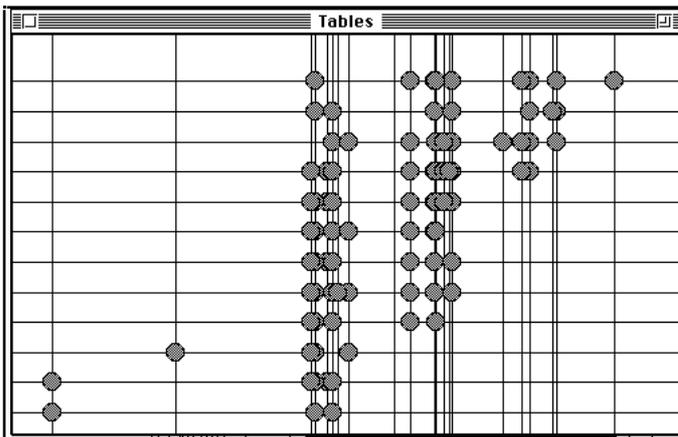
File A.ivc1 contains column scores (dependant variables) with unit norm
 It has 22 rows and 2 columns

File :A.ivc1

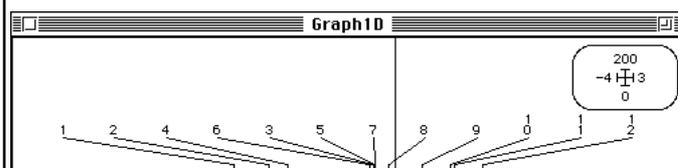
Col.	Mini	Maxi
1	-3.579e+00	2.294e+00
2	-4.526e+00	2.873e+00

Chaque relevé prend une position par averaging symétrique :





Values	
Input table file	<input type="text" value="Veg"/>
H-axis position file	<input type="text" value="R.ivc1"/>
Column number (default = 1)	<input type="text" value=""/>
Y-axis position file	<input type="text" value="R.ivls"/>
Column number (default = 1)	<input type="text" value=""/>
H-axis: Ordination (1) or Ranking (2)	<input type="text" value="1"/>
Y-axis: Ordination (1) or Ranking (2)	<input type="text" value="2"/>



Labels	
Data file (no default)	<input type="text" value="R.ivls"/>
Rows label file (default = #)	<input type="text" value="#"/>
Variable label file (or #)	<input type="text" value=""/>
Vertical (1) or horizontal (2) graphs	<input type="text" value="2"/>

Les positions moyennes par relevés sont dans le fichier ---.ivls :

File A.ivls contains row scores (dependant variables before projection)
It has 12 rows and 2 columns

```
File :A.ivls
|Col.| Mini | Maxi |
|----|-----|-----|
| 1 | -1.684e+00 | 9.063e-01 |
| 2 | -7.768e-01 | 1.178e+00 |
|----|-----|-----|
```

Le nouveau score des relevés a une certaine variance, qui ne peut dépasser la première valeur propre de l'AFC (optimale de ce point de vue). Les espèces ont été positionnées non pour que la variance des positions par relevés soit maximale mais pour que la **variance expliquée** par régression multiple sur les variables de Mil soit maximale. Cette variance expliquée est le produit de la variance par le carré de corrélation (R^2). Cette décomposition fondamentale est simplement indiquée dans :

Num	Variance	R2	Eig. Val.
1	0.4275	0.9691	0.4143
2	0.2423	0.8339	0.202

La prédiction est dans le fichier ---.ivli :

File A.ivli contains predicted row scores (dependant variables after projection)
It has 12 rows and 2 columns

```
File :A.ivli
|Col.| Mini | Maxi |
|----|-----|-----|
| 1 | -1.906e+00 | 9.503e-01 |
| 2 | -9.145e-01 | 1.023e+00 |
|----|-----|-----|
```

On pourra utiliser diverses techniques de représentation pour étudier la régression sous-jacente. En normalisant ce dernier fichier on retourne au début de la discussion et la méthode est clairement une méthode aux vecteurs propres.



Une difficulté majeure survient dans cette analyse. Elle est associée à la régression sous-jacente et aux difficultés ordinaires de la régression multiple et de l'analyse discriminante. Dès que les variables de milieu sont nombreuses (en particulier après une ACM) la prédictibilité de n'importe quel score est très bonne et la partie régression de CANOCO n'a pas de sens. **On ne s'en aperçoit pas**, car dans ce cas on fait simplement l'AFC du tableau à expliquer, ce qui conserve toujours un sens. L'alternative de l'analyse de co-inertie (CoInertia : Coinertia analysis) est alors vivement recommandée.



Chessel, D., Lebreton, J.D. & Yoccoz, N. (1987) Propriétés de l'analyse canonique des correspondances. Une utilisation en hydrobiologie. *Revue de Statistique Appliquée*, 35, 4, 55-72.

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. (1987) The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, 69, 69-77.

Ter Braak, C.J.F. (1988) CANOCO - a FORTRAN program for Canonical community ordination by [partial][detrended][canonical] correspondence analysis and redundancy analysis. Software documentation. Version 2.1, TNO Institute of Applied Computer Science, Wageningen.

CCA: Initialize explanatory variables



Utilitaire de préparation du tableau des variables de milieu dans une Analyse Canonique des Correspondances.



L'option utilise une seule fenêtre de dialogue :

Initialize explanatory variables

Explanatory variables

Option: output file name

Quit Ok

 Nom du fichier ---.##ta associée à l'analyse d'un tableau. Le module PCA, doit avoir été utilisé sur le tableau des variables de milieu pour obtenir ce fichier (en utilisant comme pondération des lignes le poids des lignes de l'AFC du tableau des espèces).

 Option : nom générique des fichiers de sortie. Par défaut, on reprend celui du tableau d'entrée.



Utiliser le dossier de travail créé par la carte Banyuls de la pile ADE-4•Data. Exécuter l'AFC (COA : CORrespondence Analysis) du tableau Veg (12-22) :

CORrespondence Analysis

Data file 12 22

Exécuter l'ACP normée (PCA : Correlation matrix PCA) du tableau Mil (12-4) en utilisant la pondération de l'analyse précédente :

Correlation matrix PCA

Matrix input file 12 4

Row weights (default=1/n)

Column weights (default=1)

Option: file for row weighting 12 1

Calculer le sous-espace associé à ce tableau normalisé (CCA : Initialize explanatory variables) :

Initialize explanatory variables

Explanatory variables

Option: output file name

Quit Ok

On obtient le listing d'exécution suivant :

```
Orthonormalization: subspace generated by a statistical triplet
-----
Explanatory variable file: Mil.cnta
It has 12 rows and 4 columns
-----
Orthonormal basis: Mil.@ob
It has 12 rows and 4 columns
Row weight file: Mil.@pl
(the same as Mil.cnpl)
Coordinates of the vectors of the orthonormal basis
in the basis of columns of Mil.cnta in : Mil.@co
File Mil.@co has 4 rows and 4 columns
-----
```

CCA: Projected inertia decomposition



Décomposition de l'inertie dans une Analyse Canonique des Correspondances.



L'option utilise une seule fenêtre de dialogue :

Nom du fichier .@ob (obtenu par CCA : CCA).

Nom du fichier ---.##ta associée à l'AFC du tableau florofaunistique.



Reprendre l'exemple de la carte Banyuls de la pile ADE-4•Data. Après l'exécution de l'option CCA : Initialize explanatory variables, on obtient la décomposition de l'inertie projetée par :

On obtient le listing suivant :

```
Projected inertia on a subspace
```

```
-----
Orthonormal basis: Mil.@ob
It has 12 rows and 4 columns
Dependant variable file: Veg.fcta
It has 12 rows and 22 columns
-----
```

	Subspace A	A Orthogo	Total	A+	A-
1	1.5177e-01	9.3622e-02	2.4539e-01	6184	3815
2	1.7937e-02	6.6594e-02	8.4530e-02	2121	7878
3	2.6990e-02	1.4448e-02	4.1438e-02	6513	3486
...					
15	4.9693e-02	9.7557e-03	5.9449e-02	8358	1641
16	2.8495e-02	2.7974e-02	5.6469e-02	5046	4953
17	5.3174e-03	6.7378e-02	7.2695e-02	731	9268
...					
22	5.5419e-02	2.4852e-02	8.0271e-02	6904	3095
Tot	8.5923e-01	9.3645e-01	1.7957e+00	4784	5215

Cette option donne pour chaque espèce du tableau floristique (Veg.fcta) le pourcentage de variance expliquée par la régression multiple de la colonne correspondante sur l'ensemble des variables de milieu. Ce pourcentage varie de 7.3% (taxon 17) à 83.6% (taxon 15). Au total 48% de l'inertie de Veg.fcta est prévisible par régression multiple et 52% n'est pas modélisable.