

Projectors

Projectors : Combine 2 Orthonormal Bases.....	2
Projectors : Intersection of 2 Subspaces.....	8
Projectors : One Categ Var->Orthonormal Basis.....	11
Projectors : Orthogonal PCAIV.....	14
Projectors : PCA on Instrumental Variables.....	20
Projectors : Subspace Test.....	26
Projectors : Table Projection.....	32
Projectors : Table->Orthonormal Basis.....	34
Projectors : Triplet Inertia Decomposition.....	36
Projectors : Triplet->Orthonormal Basis.....	39
Projectors : Two Categ Var->Orthonormal Bases.....	41
Résumé de l'organisation générale.....	46

Projectors : Combine 2 Orthonormal Bases



Utilitaire de construction de sous-espaces de projection.



Deux sous-espaces vectoriels X et Y sont connus par des bases orthonormées contenus dans des tableaux **X** et **Y**. On calcule des bases orthonormées de sous-espaces associés à X et Y. Les fichiers utilisés doivent contenir des bases orthonormées pour la même pondération des mêmes individus (lignes).

Si on note \mathbf{x}^j la colonne j du tableau **X** et \mathbf{y}^k la colonne k du tableau **Y** et \mathbf{a}_i la composante i du vecteur **a**, le sous-espace $X \times Y$ est engendré par les vecteurs de composantes $\mathbf{x}_i^j \mathbf{y}_i^k$ (modèles multiplicatifs, ou effet d'interaction). On s'en sert essentiellement pour croiser des variables qualitatives dans des plans d'observations avec des répétitions. Le sous-espace $X \bullet Y$ est le complémentaire orthogonal dans le sous-espace $X \times Y$ du sous-espace $X + Y$.

Le sous-espace $X + Y$ est engendré par la famille des vecteurs \mathbf{x}^j réunie avec la famille des vecteurs \mathbf{y}^k (modèles additifs dits effet $X + Y$). Le sous espace Y/X est la partie de $X + Y$ orthogonale à X, ensemble des variables du type $X + Y$ non corrélées à X (modèles dits effet X sachant Y). Ces sous-espaces définissent les analyses partielles, en particulier l'analyse canonique des correspondances partielles :

The only difference with canonical correspondence analysis is that the matrix of explanatory variables is replaced by the matrix of residuals of a multivariate multiple regression of Z_2 on Z_1 (Ter Braak, C.J.F. (1988) Partial Canonical Analysis. In : Classification and related methods of data analysis. Bock, H.H. (Ed.) North Holland. 551-558, p. 555).

Enfin, le sous espace X/Y est la partie de $X + Y$ orthogonale à Y, ensemble des variables du type $X + Y$ non corrélées à Y (modèles dits effet Y sachant X). La décomposition la plus classique est celle de l'analyse de variance à deux facteurs du type $X + Y$, X, Y/X et $X + Y$, Y, X/Y .



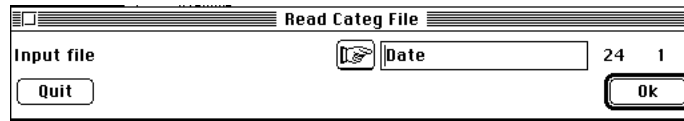
L'option utilise une seule fenêtre de dialogue :

Combine 2 Orthonormal Bases	
X: .@ob file	<input type="text"/>
Y: .@ob file	<input type="text"/>
Output file name	<input type="text"/>
I = Save XxY subspace	<input type="checkbox"/>
I = Save X.Y subspace	<input type="checkbox"/>
I = Save X+Y subspace	<input type="checkbox"/>
I = Save X/Y subspace	<input type="checkbox"/>
I = Save Y/X subspace	<input type="checkbox"/>

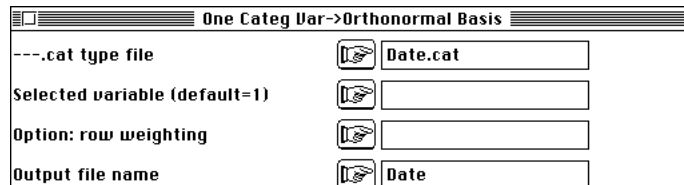
- Nom du fichier de type .@ob qui définit le sous-espace X.
- Nom du fichier de type .@ob qui définit le sous-espace Y.
- Nom générique des fichiers de sortie (création).
- Option de sauvegarde du sous-espace $X \times Y$ (Taper 1 pour oui).
- Option de sauvegarde du sous-espace $X \bullet Y$ (Taper 1 pour oui).
- Option de sauvegarde du sous-espace $X + Y$ (Taper 1 pour oui).
- Option de sauvegarde du sous-espace X/Y (Taper 1 pour oui).
- Option de sauvegarde du sous-espace Y/X (Taper 1 pour oui).



Utiliser les dossiers de travail créés par les cartes Méaudret et Méaudret+2 de la pile ADE-4•Data. Déplacer tous les fichiers créés dans un même dossier de travail. Lire le fichier Date avec CategVar : Read Categ File et faire de même pour le fichier Plan.



L'unique variable qualitative de Date sépare les relevés en 4 groupes (saisons). Définir le sous-espace associé à cette variable qualitative par Projectors : One Categ Var->Orthonormal Basis (voir la page de cette option) :



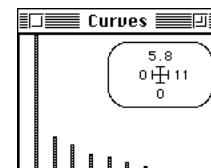
```

*-----*
Subspace from one categorical variable
-----
Input file: Date
It has 24 rows and 1 columns
Generic output file name: Date
Selected variable n°: 1
-----

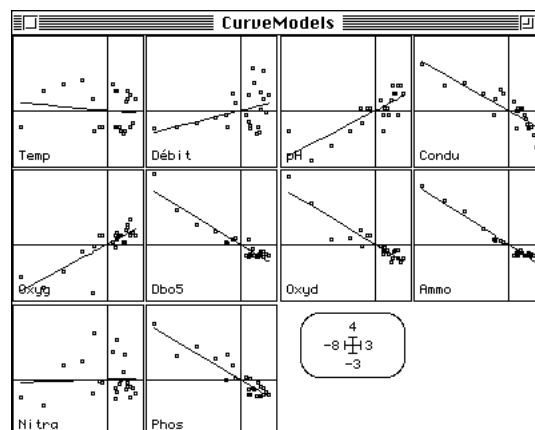
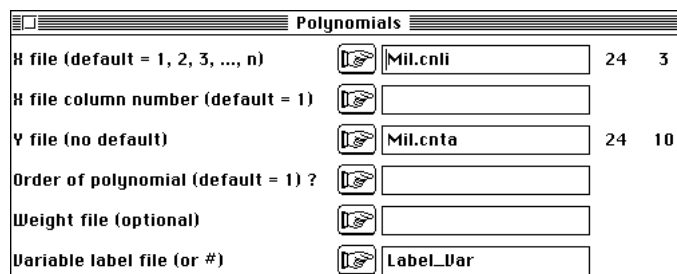
```

File Date.@ob contains an orthonormal basis of the subspace generated by the selected variable
It has 24 rows and 3 columns

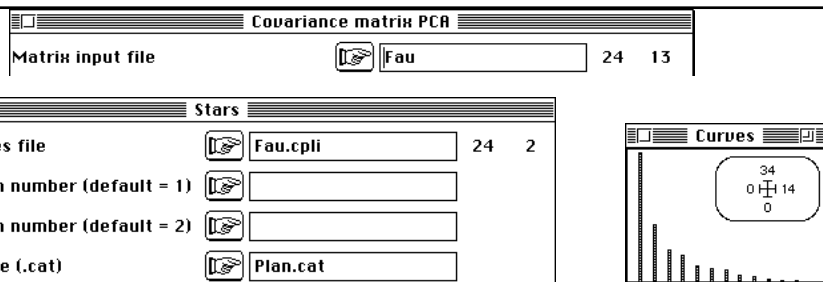
Faire l'ACP normée du tableau Mil :



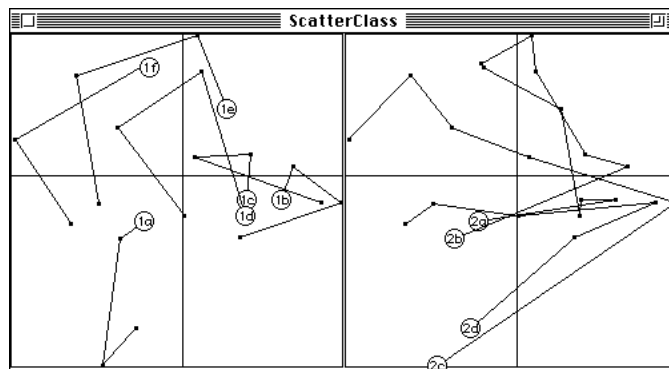
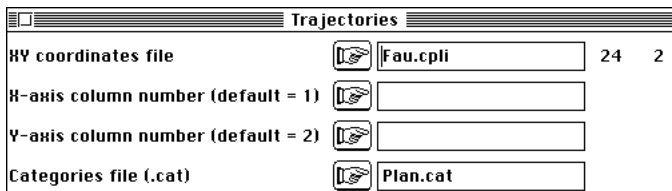
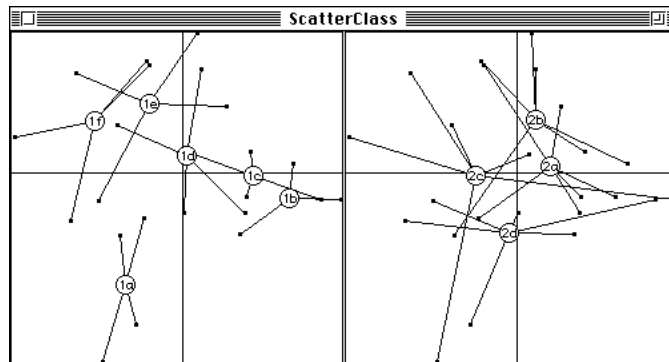
Observer la prépondérance du facteur 1, interprété comme un indice global de pollution :



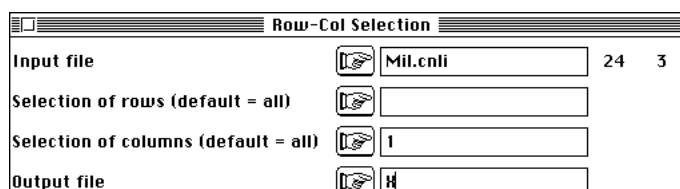
Faire l'ACP centrée du tableau Fau :



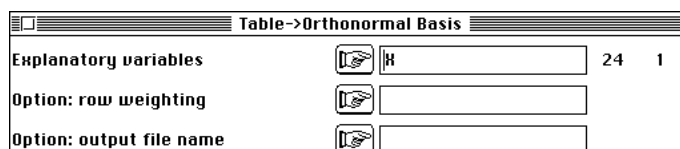
Observer le mélange des effets temporels et stationnels dans le plan 1-2 :



On veut donc étudier le rôle de la saison et de la pollution sur la faune. Définir le sous-espace de régression engendré par la première coordonnée factorielle de l'ACP normée de Mil (régression sur composantes). Extraire du fichier Mil.cnli la première colonne (FilesUtil : Row-Col Selection) :



Définir le sous-espace de projection associée (Projectors : Table->Orthonormal Basis) :



Orthonormalization: subspace generated by quantitative variables

Explanatory variable file: X
It has 24 rows and 1 columns

Orthonormal basis: X.@ob
It has 24 rows and 1 columns
Row weight file: X.@pl
Uniform row weight = 0.041667
Coordinates of the vectors of the orthonormal basis
in the initial basis in : X.@co
File X.@co has 1 rows and 1 columns

Associer les deux sous-espace de projection respectivement associé à la variable qualitative date et à la variable quantitative pollution :

Combine 2 Orthonormal Bases			
H: .@ob file		H.@ob	24 1
Y: .@ob file		Date.@ob	24 3
Output file name		HD	
I = Save HxY subspace			
I = Save H•Y subspace			
I = Save H+Y subspace		I	
I = Save H/Y subspace		I	
I = Save Y/H subspace		I	

X&YSpaces

First orthonormal basis: X.@ob
It has 24 rows and 1 columns
Second orthonormal basis: Date.@ob
It has 24 rows and 3 columns

File XD_X+Y.@ob contains an orthonormal basis of the subspace X+Y
It has 24 rows and 4 columns
File XD_X/Y.@ob contains an orthonormal basis of the subspace X/Y
It has 24 rows and 1 columns
File XD_Y/X.@ob contains an orthonormal basis of the subspace Y/X
It has 24 rows and 3 columns

On peut ainsi étudier la capacité prédictive du couple date-pollution (Projectors : Triplet Inertia Decomposition) :

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		HD_H+Y.@ob	24 4
Dependant variables: .**ta		Fau.cpta	24 13

Tot	3.6628e+01	4.3108e+01	7.9736e+01	4593	5406
-----	------------	------------	------------	------	------

46 % de la variabilité totale du tableau faunistique est modélisable en terme effet date + effet pollution. Dans ces 46 %, 20 % sont attribués à la pollution et 26 % à l'effet date sachant pollution :

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		H.@ob	24 1
Dependant variables: .**ta		Fau.cpta	24 13

Tot	1.6042e+01	6.3694e+01	7.9736e+01	2011	7988
-----	------------	------------	------------	------	------

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		HD_Y/H.@ob	24 3
Dependant variables: .**ta		Fau.cpta	24 13

Tot	2.0586e+01	5.9150e+01	7.9736e+01	2581	7418
-----	------------	------------	------------	------	------

De même, dans ces 46 %, 19 % sont attribués à la date et 27 % à l'effet pollution sachant date :

Triplet Inertia Decomposition					
Explanatory variables: .@ob file	<input type="checkbox"/>	Date.@ob	24	3	
Dependant variables: .**ta	<input type="checkbox"/>	Fau.cpta	24	13	

Tot	1.5097e+01	6.4639e+01	7.9736e+01	1893	8106
-----	------------	------------	------------	------	------

Triplet Inertia Decomposition					
Explanatory variables: .@ob file	<input type="checkbox"/>	RD_H/V.@ob	24	1	
Dependant variables: .**ta	<input type="checkbox"/>	Fau.cpta	24	13	

Tot	2.1531e+01	5.8205e+01	7.9736e+01	2700	7299
-----	------------	------------	------------	------	------

Il y a bien effet additif à deux composantes indépendantes. La synthèse de la modélisation sous-jacente est obtenue par une ACPVI (Projectors : PCA on Instrumental Variables) :

PCA on Instrumental Variables					
Explanatory variables: .@ob file	<input type="checkbox"/>	RD_X+Y.@ob	24	4	
Dependant variables: .**ta	<input type="checkbox"/>	Fau.cpta	24	13	
Output file name	<input type="checkbox"/>	W			

```
Instrumental variables
----- input -----
Orthonormal basis: XD_X+Y.@ob
It has 24 rows and 4 columns
Dependent variable file: Fau.cpta
It has 24 rows and 13 columns
----- output -----
Projected variable file: WW.ivta
It has 24 rows and 13 columns
Inertia: 3.6628e+01
File WW.ivpc contains the column weight
It has 13 rows and 1 column
File WW.ivpl contains the row weight
It has 24 rows and 1 column
```

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+2.5050E+01	+0.6839	+0.6839	02	+6.8167E+00	+0.1861	+0.8700
03	+4.0877E+00	+0.1116	+0.9816	04	+6.7388E-01	+0.0184	+1.0000

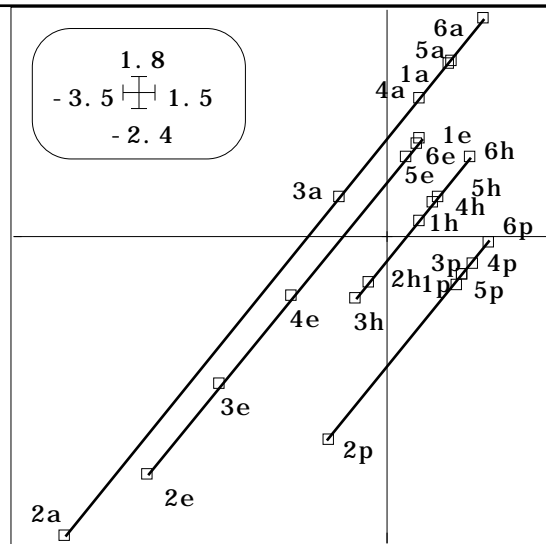
File WW.ivl1 contains the canonical row scores which are linear combination of the explanatory variables (unity norm)
It has 24 rows and 2 columns

```
File :WW.ivl1
```

Col.	Mini	Maxi
1	-2.338e+00	1.724e+00
2	-1.426e+00	1.583e+00

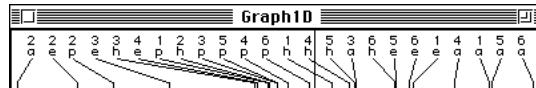
La variabilité totale du tableau faunistique comporte 46% d'explication par un modèle date+pollution dont 68% de ces 46% (soit 31%) par un seul modèle pour toutes les espèces. On n'aurait pas pu dépasser 41% (premier taux d'inertie de l'ACP de Fau). On retient pour le dépouillement la constitution d'un prédicteur unique (premier score canonique) dont on peut expliciter la constitution par les explicatives par (Curves : Lines) :

Lines	
X file (default = 1, 2, 3, ..., n)	H.eob
X file column number (default = 1)	
Y file (no default)	WW.iv11
Cumulated data (1=yes, 2=no)	
Variable label file (or #)	
Draw curves (1=yes, 2=no)	2
Draw points (1=yes, 2=no)	
Row label file (or #)	Label_rel



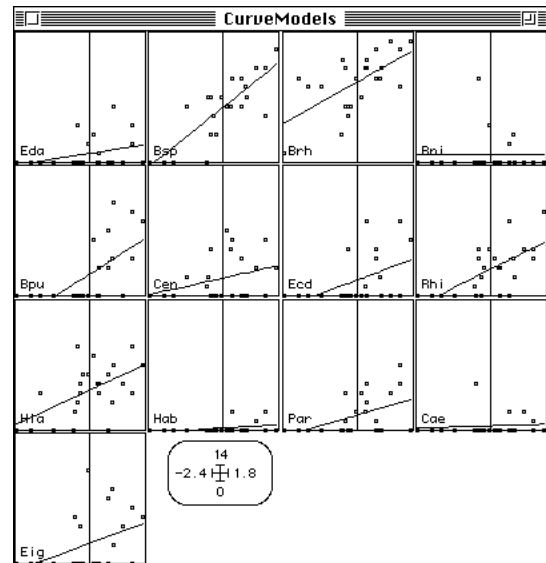
Ce prédicteur positionne les 24 relevés (Graph1D : Labels) :

Labels	
Data file (no default)	WW.iv11 24 2
Rows label file (default = #)	Label_rel



On représente les données traitées à l'aide de ce code (CurveModels : Polynomials) :

Polynomials	
X file (default = 1, 2, 3, ..., n)	WW.iv11
X file column number (default = 1)	
Y file (no default)	Fau
Order of polynomial (default = 1) ?	
Weight file (optional)	
Variable label file (or #)	Label_Esp



On obtient une bonne représentation de l'effet conjoint date-pollution sur l'abondance des organismes.



Takeuchi, K., Yanai, H. & Mukherjee, B.N. (1982) *The foundations of multivariate analysis. A unified approach by means of projection onto linear subspaces*. John Wiley and Sons, New York. 1-458.

Sabatier, R., Lebreton, J.D. & Chessel, J.D. (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. In : *Multivariate data analysis*. Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland. 341-352.

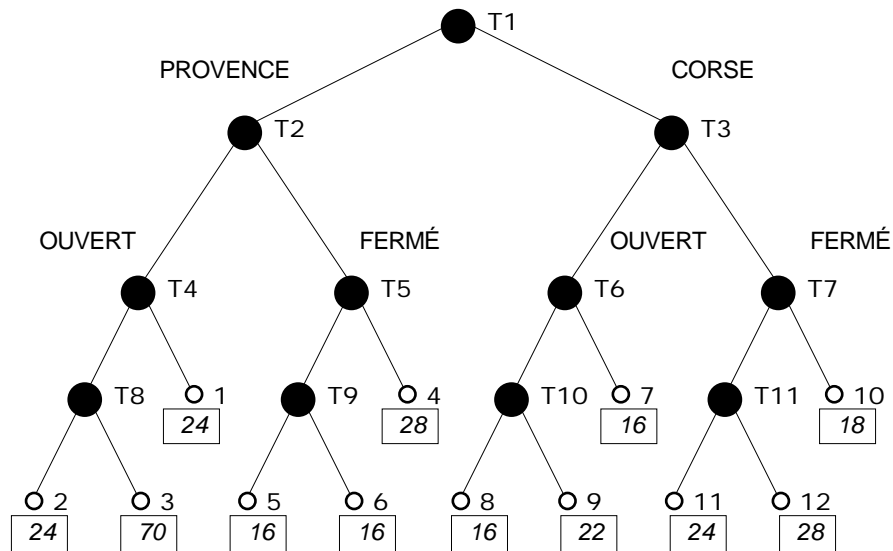
Projectors : Divisive hierarchy->Orthonormal Basis



Création d'une base orthonormale à partir d'un arbre de décision binaire.



Les méthodes de segmentation fournissent des arbres de décision binaire (voir une introduction dans ¹, section 3.5, p.302-318). On suppose ici que l'arbre est construit : on veut l'introduire comme information externe pour l'analyse d'un tableau de données. Un arbre de décision binaire est une information supplémentaire sur une simple variable qualitative. Si n objets sont répartis en m classes, cette information supplémentaire indique comment les classes sont construites. Par exemple, le plan d'observations de J. Blondel ¹ consigné dans la carte de données Provence_Corse+1 peut être représenté par :



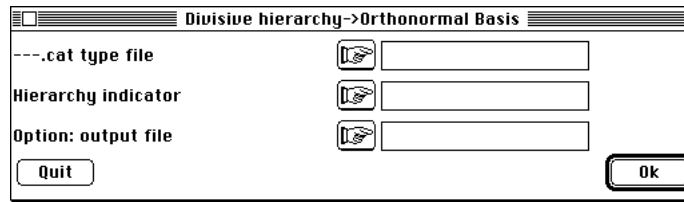
Il y a $n = 302$ relevés répartis en $m = 12$ classes (ou segments terminaux, cercles noirs) d'effectifs respectifs n_1, n_2, \dots, n_m (encadrés). Les segments terminaux sont construits par division successive de $m - 1 = 11$ classes intermédiaires (ou segments intermédiaires, cercles blancs). Appelons descriptif de la hiérarchie le tableau avec m lignes (segments terminaux) et $m - 1$ colonnes (segments intermédiaires) qui contient à la ligne i et à la colonne j la valeur -1 ou $+1$ suivant que la division du segment intermédiaire j a placé le segment final i dans un groupe ou dans l'autre. Si la division ne concerne pas le segment terminal on indique la valeur 0 :


	T										
	1	2	3	4	5	6	7	8	9	10	11
1	-1	-1	0	-1	0	0	0	0	0	0	0
2	-1	-1	0	1	0	0	0	-1	0	0	0
3	-1	-1	0	1	0	0	0	1	0	0	0
4	-1	1	0	0	-1	0	0	0	0	0	0
5	-1	1	0	0	1	0	0	0	-1	0	0
6	-1	1	0	0	1	0	0	0	1	0	0
7	1	0	-1	0	0	-1	0	0	0	0	0
8	1	0	-1	0	0	1	0	0	0	-1	0
9	1	0	-1	0	0	1	0	0	0	1	0
10	1	0	1	0	0	0	-1	0	0	0	0
11	1	0	1	0	0	0	1	0	0	0	-1
12	1	0	1	0	0	0	1	0	0	0	1


Le descriptif de la hiérarchie est alors un fichier binaire à m lignes et $m - 1$ colonnes. Ce descriptif permet de construire une base orthonormée du sous-espace engendré par les indicatrices de la variable qualitative donnant la partition finale en m classes. Ce sous-espace de dimension $m - 1$ est engendré par une base dans laquelle chaque vecteur correspond exactement à un segment intermédiaire.




L'option utilise une seule fenêtre de dialogue :



 Nom du fichier .cat qui renvoie à un fichier ne contenant qu'une seule variable qualitative (la partition étudiée).

 Nom du fichier binaire de sortie (création).

 Nom du fichier binaire de sortie (création).



Dialogues intermédiaires



Exemple



Particularités



Difficultés



Ancienne version



Informations connexes



Bibliographie

Projectors : Intersection of 2 Subspaces



Utilitaire de construction de sous-espaces de projection.



Deux sous-espaces vectoriels X et Y sont connus par des bases orthonormées contenues dans des tableaux **X** et **Y**. On calcule une bases orthonormée du sous-espace X (Y. Les fichiers utilisés doivent contenir des bases orthonormées pour la même pondération des mêmes individus (lignes).



L'option utilise une seule fenêtre de dialogue :

Nom du fichier de type .@ob qui définit le sous-espace X.

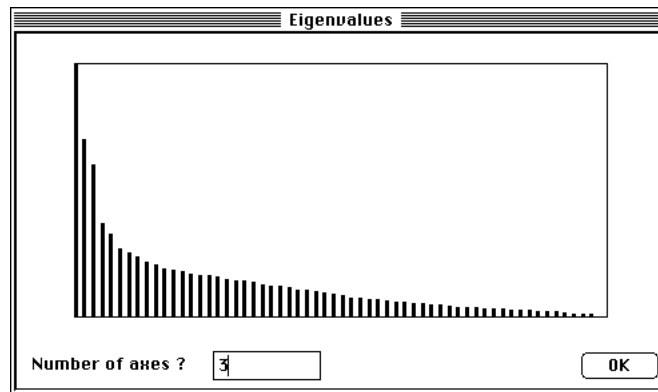
Nom du fichier de type .@ob qui définit le sous-espace Y.

Nom générique des fichiers de sortie (création).



Utiliser le fichier créé par la carte Provence_Corse de la pile ADE-4•Data. Lire le fichier Plan (302-2) par CategVar : Read Categ File :

Exécuter l'AFC du tableau Avi (302-60) par COA : COCorrespondence Analysis :






Définir les sous-espaces de projection associés au plan d'échantillonnage par Projectors : Two Categ Var->Orthonormal Bases (voir la fiche de l'option) :

Pour obtenir une base orthonormée du sous-espace A B , utiliser la propriété :

$$A \quad (A/B) = A \quad (A + B) \quad B = A \quad B$$

Cela permet d'éviter de chercher les vecteurs canoniques associés aux valeurs propres nulles de l'analyse canonique de A et B (comme proposé en ¹) en calculant les vecteurs canoniques associés aux valeurs propres unité de l'analyse canonique de A et A/B.

Intersection of 2 Subspaces			
X: .@ob file		P_A.@ob	302 1
Y: .@ob file		P_A/B.@ob	302 1
Output file name		W1	




Orthonormal basis of the intersection of two subspaces

 First orthonormal basis: P_A.@ob
 It has 302 rows and 1 columns
 Second orthonormal basis: P_A/B.@ob
 It has 302 rows and 1 columns

n°: 1 Eigenvalue: 9.58880901e-01

Dimension of X-Inter-Y is: 0
 Subspace X-Inter-Y is null

L'espace A (effet région, 2 modalités) est de dimension 1. L'espace A/B (effet région sachant végétation) est de dimension 1. L'intersection des deux est de dimension 0. La valeur propre de l'analyse canonique des deux sous-espaces (voir la synthèse sur la stratégie dans 2) est unique et non égale à 1.

Intersection of 2 Subspaces			
X: .@ob file		P_B.@ob	302 5
Y: .@ob file		P_B/A.@ob	302 5
Output file name		W2	

L'espace B (effet végétation, 6 modalités) est de dimension 5. L'espace B/A (effet végétation sachant région) est de dimension 5. L'intersection des deux est de dimension 4. Les valeurs propres de l'analyse canonique des deux sous-espaces (voir la synthèse sur la stratégie dans 2) sont au nombre de 5. Les 4 premières sont égales à 1 (à 10^{-6} près). Elles sont éditées pour contrôle.

Orthonormal basis of an intersection of subspaces

 First orthonormal basis: P_B.@ob
 It has 302 rows and 5 columns
 Second orthonormal basis: P_B/A.@ob
 It has 302 rows and 5 columns

n°: 1 Eigenvalue: 1.00000000e+00<--
 n°: 2 Eigenvalue: 1.00000000e+00<--
 n°: 3 Eigenvalue: 9.99999940e-01<--
 n°: 4 Eigenvalue: 9.99999940e-01<--
 n°: 5 Eigenvalue: 9.58880842e-01

Dimension of X-Inter-Y is: 4
 File W2_XiY.@ob contains an orthonormal basis of the subspace X_Inter_Y
 It has 302 rows and 4 columns

On voit donc que l'espace A ∩ B n'existe que sous certaines conditions. Quand il existe, c'est l'ensemble des variables centrées constantes par classe de A et centrées par classe de B (effet B et non A). Ce sous-espace a été défini par ³. L'ACPVI correspondante est utilisée dans ⁴.



1 Pontier, J. & Pernin, M.O. (1987) Solution using "LONGI". In : *Data Analysis : Ins and Outs of solving real problems*. Janssen, J., Marcotorchino, F. & Proth, J.M. (Eds.) Plenum Publishing Corporation. 49-65.
 Pontier, J. & Pernin, M.O. (1989) Relations entre analyse canonique complète et méthode Longi. *Revue de Statistique Appliquée* : 37, 4, 67-82.

² Pages, J.P. & Tomassone, R. (1979) Analyse des données et modélisation. In : *Elaboration et justification des modèles : applications en Biologie*, tome II. Maloine SA ed. 407-439.

³ Afriat, S.N. (1957) Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences* : 53, 800-816.

⁴ Yoccoz, N. & Chessel, D. (1988) Ordination sous contraintes de relevés d'avifaune : élimination d'effets dans un plan d'observations à deux facteurs. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : III, 307 : 189-194.

Projectors : One Categ Var->Orthonormal Basis



Utilitaire de définition d'un sous-espace de projection.



Une variable qualitative sur n individus définit un sous-espace de l'espace euclidien \mathbb{R}^n . Quand on projette une variable sur ce sous-espace on obtient la variable où les valeurs initiales sont remplacées par les moyennes des classes. Quand on projette cette variable sur le sous-espace orthogonal on obtient la variable où les valeurs sont remplacées par les écarts aux moyennes de la classe. Le sous-espace a en général comme dimension le nombre de modalités - 1. Une ACPVI (Projectors : PCA on Instrumental Variables) sur ce sous-espace est exactement l'analyse inter-classe (Discrimin : Between analysis/Run). Une ACPVI orthogonale (Projectors : Orthogonal PCAIV) sur ce sous-espace est exactement une analyse intra-classe (Discrimin : Within Parameters). Un test de permutation sur ce sous-espace (Projectors : Subspace Test) est équivalent à un test de permutations sur l'inter-classe (Discrimin : Between analysis/Test). Les analyses inter et intra-classe sont des cas particuliers d'ACPVI et d'ACPVI orthogonale ¹.



L'option utilise une seule fenêtre de dialogue :

- Nom du fichier .cat créé par CategVar : Read Categ File.
- Numéro de la variable qualitative sélectionnée. Par défaut c'est la première.
- Option : fichier de pondération. Par défaut c'est la pondération uniforme.
- Nom générique des fichiers de sortie (création).



Utiliser les dossiers de travail créés par les cartes Méaudret et Méaudret+2 de la pile ADE-4•Data. Déplacer tous les fichiers créés dans un même dossier de travail. Lire le fichier Date avec CategVar : Read Categ File :

L'unique variable qualitative de Date sépare les relevés en 4 groupes (saisons). Définir le sous-espace associé à cette variable qualitative (dialogue ci-dessus). Lister les fichiers créés :

Date.@ob			
	1	2	3
1	1.7321	0.0000	0.0000
2	-0.5774	1.6330	0.0000
3	-0.5774	-0.8165	1.4142
4	-0.5774	-0.8165	-1.4142
5	1.7321	0.0000	0.0000
6	-0.5774	1.6330	0.0000
7	-0.5774	-0.8165	1.4142
8	-0.5774	-0.8165	-1.4142
9	1.7321	0.0000	0.0000
10	-0.5774	1.6330	0.0000
11	-0.5774	-0.8165	1.4142
12	-0.5774	-0.8165	-1.4142
13	1.7321	0.0000	0.0000
14	-0.5774	1.6330	0.0000
15	-0.5774	-0.8165	1.4142
16	-0.5774	-0.8165	-1.4142
17	1.7321	0.0000	0.0000
18	-0.5774	1.6330	0.0000
19	-0.5774	-0.8165	1.4142
20	-0.5774	-0.8165	-1.4142
21	1.7321	0.0000	0.0000
22	-0.5774	1.6330	0.0000
23	-0.5774	-0.8165	1.4142
24	-0.5774	-0.8165	-1.4142

Date.@pl	
	1
1	0.0417
2	0.0417
3	0.0417
4	0.0417
5	0.0417
6	0.0417
7	0.0417
8	0.0417
9	0.0417
10	0.0417
11	0.0417
12	0.0417
13	0.0417
14	0.0417
15	0.0417
16	0.0417
17	0.0417
18	0.0417
19	0.0417
20	0.0417
21	0.0417
22	0.0417
23	0.0417
24	0.0417

-----*

Subspace from one categorical variable

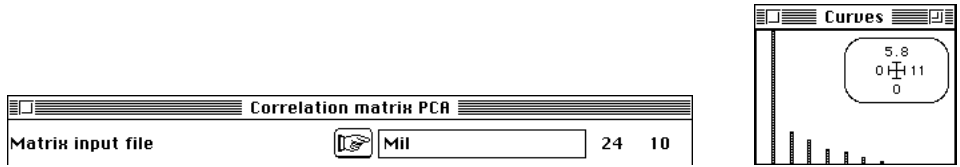
-----*

Input file: Date

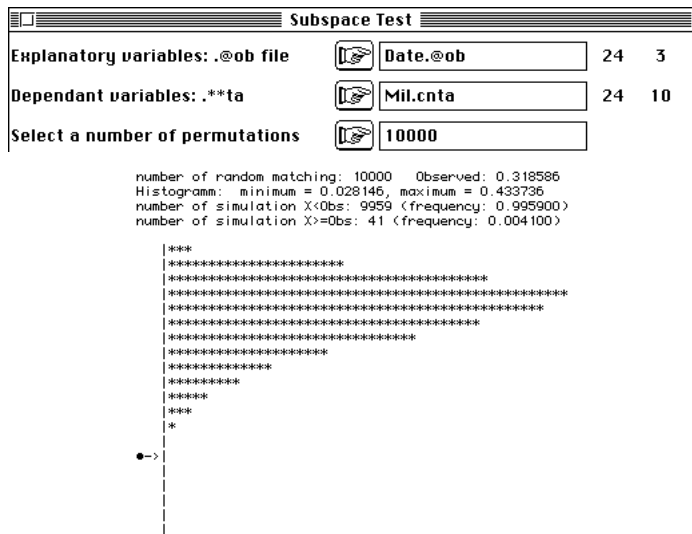
It has 24 rows and 1 columns
 Generic output file name: Date
 Selected variable n°: 1

 File Date.@ob contains an orthonormal basis of the subspace generated by the selected variable
 It has 24 rows and 3 columns

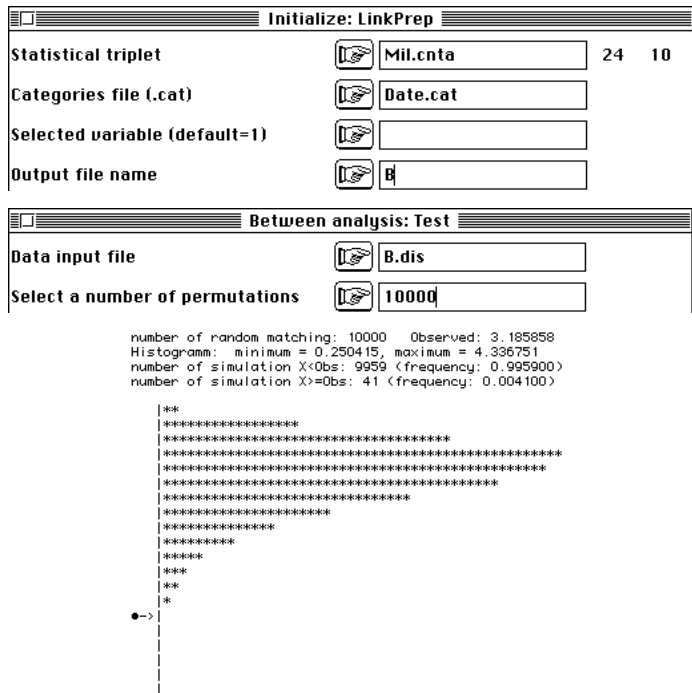
Faire l'ACP normée du tableau Mil (10 variables de milieu sur les 24 relevés) :



Tester l'ACPVI de cette analyse par rapport au sous-espace défini précédemment :



Tester l'ACP inter-date associée :



Exécuter l'ACPVI de cette analyse par rapport au sous-espace défini précédemment :

PCA on Instrumental Variables			
Explanatory variables: .@ob file	<input type="text" value="Date.@ob"/>	24	3
Dependant variables: .**ta	<input type="text" value="Mil.cnta"/>	24	10
Output file name	<input type="text" value="A"/>		

Exécuter l'ACP inter-date associée :

Between analysis: Run	
--.dis input file	<input type="text" value="B.dis"/>

Comparer point par point :

Projected variable file: A.ivta
 It has 24 rows and 10 columns
 Inertia: 3.1859e+00
between-class inertia 3.185858 (ratio: 0.318586)

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.5551E+00	+0.4881	+0.4881	02	+1.0390E+00	+0.3261	+0.8143
03	+5.9176E-01	+0.1857	+1.0000				

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.5551E+00	+0.4881	+0.4881	02	+1.0390E+00	+0.3261	+0.8143
03	+5.9176E-01	+0.1857	+1.0000	04	+0.0000E+00	+0.0000	+1.0000

...

File :A.ivli

Col.	Mini	Maxi
1	-1.186e+00	1.821e+00
2	-1.352e+00	1.208e+00

File :B.beli

Col.	Mini	Maxi
1	-1.821e+00	1.186e+00
2	-1.208e+00	1.352e+00

...



1 Dolédec, S. & Chessel, D. (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. *Acta Oecologica, Oecologia Generalis* : 8, 3, 403-426.

Dolédec, S. & Chessel, D. (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique II- Prise en compte et élimination d'effets dans un tableau faunistique. *Acta Oecologica, Oecologia Generalis* : 10, 3, 207-232.

Projectors : Orthogonal PCAIV



Méthode d'analyse de données dite ACPVIO (ACP sur Variables Instrumentales Orthogonale) appartenant à la famille des ACL (Analyses sous Contraintes Linéaires). Le cas le plus connu est celui de l'analyse des covariances partielles ¹.



L'option Projectors : PCA on Instrumental Variables projette un tableau de données sur un sous-espace de référence, fait l'analyse du tableau projeté et propose diverses aides à l'interprétation. La présente option fait la même chose pour les sous-espaces complémentaires. Quand un sous-espace A définit une contrainte ou un modèle, le sous-espace A définit la contrainte inverse ou le résidu au modèle précédent.



L'option utilise une seule fenêtre de dialogue :

Explanatory variables: .@ob file		
Dependant variables: .**ta		
Output file name		

Nom du fichier .@ob contenant une base orthonormée du sous-espace de projection. Les options Projectors : Table->Orthonormal Basis, Triplet->Orthonormal Basis, One Categ Var->Orthonormal Basis, Combine 2 Orthonormal Bases, Two Categ Var->Orthonormal Bases, et Intersection of 2 Subspaces créent de tels fichiers.

Nom du fichier ---.##ta associée à l'analyse d'un tableau. Les modules PCA, HTA, MCA et COA créent de tels fichiers.



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.

Nom générique des fichiers de sortie (création).



Utiliser le dossier de travail créé par la carte Light_trap ² de la pile ADE-4•Data. Transformer le tableau faunistique (Bin->Bin : c*Log[a*x+b]) :

Input file		Fau	49	17
Output file		FauLog		
Selection of columns (default = all)				
Parameter a (default=1)		1		
Parameter b (default=0)		1		
Parameter c (default=1)		1		

Faire l'ACP centrée du tableau transformé (PCA : Covariance matrix PCA) :

Matrix input file		FauLog	49	17
-------------------	--	--------	----	----

Faire l'ACP normée du tableau des variables météorologiques :

Matrix input file		Meteo	49	11
-------------------	--	-------	----	----

Les deux tableaux sont fortement structurés. On se demande ce qui dans la variation totale du contenu faunistique dépend et ne dépend pas des conditions de piégeage. Le tableau Mil définit un espace de projections qui correspond à la régression multiple ordinaire sur l'ensemble des variables météorologiques (Projectors : Triplet->Orthonormal Basis) :

Triplet->Orthonormal Basis		
Explanatory variables	<input type="text" value="Meteo.cnta"/>	49 11
Option: output file name	<input type="text"/>	

Ce qui en dépend relève de l'ACPVI (Projectors : PCA on Instrumental Variables) :

PCA on Instrumental Variables		
Explanatory variables: .@ob file	<input type="text" value="Meteo.@ob"/>	49 11
Dependant variables: .**ta	<input type="text" value="FauLog.cpta"/>	49 17
Output file name	<input type="text" value="Posi"/>	

Ce qui n'en dépend pas relève de l'ACPVI orthogonale :

Orthogonal PCAIV		
Explanatory variables: .@ob file	<input type="text" value="Meteo.@ob"/>	49 11
Dependant variables: .**ta	<input type="text" value="FauLog.cpta"/>	49 17
Output file name	<input type="text" value="Nega"/>	

L'inertie totale du tableau faunistique de départ (somme des variances pour toutes les espèces) vaut 20.661 :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: FauLog.cpta
--- Number of rows: 49, columns: 17
-----
Total inertia: 10.6406
-----
```

Cette inertie se décompose pour une part en inertie expliquée par régression multiple (5.2433) :

```
Instrumental variables
----- input -----
Orthonormal basis: Meteo.@ob
It has 49 rows and 11 columns
Dependent variable file: FauLog.cpta
It has 49 rows and 17 columns
----- output -----
Projected variable file: Posi.ivta
It has 49 rows and 17 columns
Inertia: 5.2433e+00
```

Cette inertie se décompose pour une autre part en inertie résiduelle (5.397) :

```
Instrumental variables
----- input -----
Orthonormal basis: Temp.@ob
It has 49 rows and 17 columns
Dependent variable file: FauLog.cpta
It has 49 rows and 17 columns
----- output -----
Projected variable file: Nega.ivta
It has 49 rows and 17 columns
Inertia: 5.3973e+00
```

Le listing de l'ACPVI orthogonale est celui d'une ACPVI et s'interprète comme telle. La seule différence est que l'effet des variables explicatives est positif dans un cas (c'est pourquoi on a choisit le nom générique de fichiers Posi) et négatif dans l'autre (nom générique Nega). Géométriquement les deux sous-espaces de projection sont des complémentaires orthogonaux. Numériquement le tableau projeté est celui des prévisions dans un cas, celui des écarts aux prévisions dans l'autre.

Dans l'analyse directe, on vérifie qu'on peut effectivement parler d'effet météorologique sur le piégeage (Projectors : Subspace Test) :

Subspace Test		
Explanatory variables: .@ob file	<input type="button" value="👉"/>	Meteo.@ob 49 11
Dependant variables: .**ta	<input type="button" value="👉"/>	FauLog.cpta 49 17
Select a number of permutations	<input type="button" value="👉"/>	1000

```

*
*****
*****
*****
*****
*****
*****
*****
*****
*****
*****
*****
***
*
-----
●->|

```

File Posi.ivc1 contains column scores (dependent variables, unity norm)
 It has 17 rows and 2 columns
 File :Posi.ivc1

Col.	Mini	Maxi
1	-1.700e-02	6.003e-01
2	-4.922e-01	4.184e-01

Dans l'analyse positive, on cherche des poids des espèces :

Labels	
XY coordinates file	<input type="button" value="👉"/> Posi.ivc1
X-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="👉"/> Label.Esp
Draw vectors from origin (yes = 1)	<input type="button" value="👉"/> 1

Scatters plot showing vectors for species: Hsp, Hfo, Led, Set, Han, Hspc, Hsp, Hum, Pga, Std, Psy. A tooltip shows coordinates: 0.55, -0.2, 0.8, -0.55.

Ces poids définissent des scores des relevés :

File Posi.ivls contains row scores (dependent variables before projection)
 It has 49 rows and 2 columns

File :Posi.ivls

Col.	Mini	Maxi
1	-2.962e+00	8.064e+00
2	-2.735e+00	3.630e+00

Match two scatters	
XY coordinates file	<input type="button" value="👉"/> Posi.ivls
X-axis column number (default = 1)	<input type="button" value="👉"/>
Y-axis column number (default = 2)	<input type="button" value="👉"/>
Second XY coordinates file	<input type="button" value="👉"/> Posi.ivli

Scatters plot showing a dense cluster of vectors. A tooltip shows coordinates: 3.7, -3.4, 8.1, -2.8.

Ces scores des relevés sont de variance expliquée par la météorologie maximale :

Num	Variance	R2	EigenVal.
1	5.311	0.7336	3.896
2	1.65	0.3768	0.6216

Seul le premier supporte la prédiction. Les prédictions sont dans :

File Posi.ivli contains the predicted row scores (dependent variables after projection)
It has 49 rows and 2 columns

File :Posi.ivli

Col.	Mini	Maxi
1	-3.314e+00	6.314e+00
2	-1.735e+00	1.308e+00

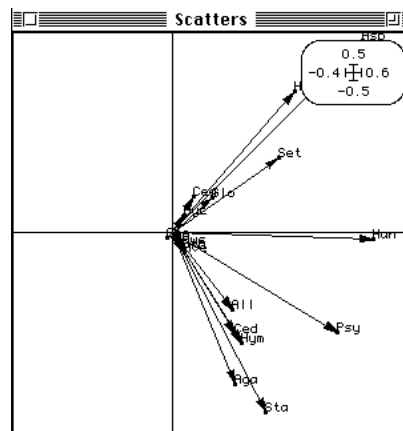
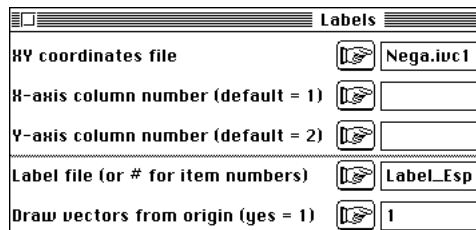
Le facteur météorologique est un fort effet taille, certaines conditions étant particulièrement favorable à l'éclosion et au vol des insectes.

Dans l'analyse négative, on cherche des poids des espèces :

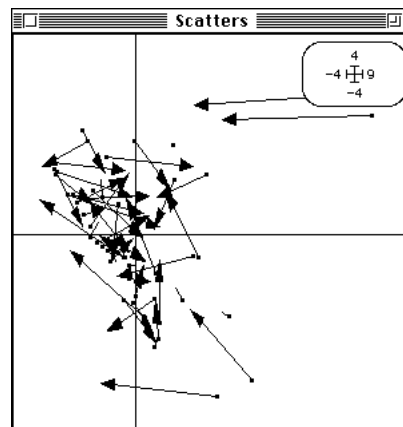
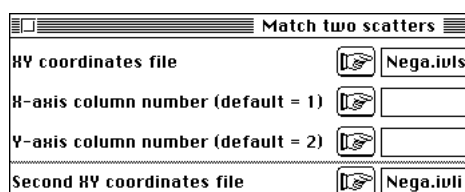
File Nega.ivc1 contains the column scores (dependent variables, unity norm)
It has 17 rows and 3 columns

File :Nega.ivc1

Col.	Mini	Maxi
1	-1.164e-02	5.047e-01
2	-4.502e-01	4.798e-01
3	-3.929e-01	6.829e-01



Ces poids définissent des scores des relevés :



Ces scores des relevés sont de variance indépendante de la météorologie maximale :

Num	Variance	R2	EigenVal.
1	4.911	0.3138	1.541
2	1.896	0.7302	1.384
3	1.294	0.5886	0.7616

L'exemple mérite d'être médité. Une ACPVI cherche une variance expliquée maximale (ici par l'orthogonal, c'est à dire une variance résiduelle après régression maximale). Une variance expliquée est un compromis entre variance et pourcentage d'explication. En perdant 70% de la variance de l'effet taille on obtient encore mieux qu'en gardant 73% de la variance du second code : l'effet taille est écrasant. Seul ici le second facteur mérite l'attention. Les scores et les prédictions sont dans :

File :Nega.ivls

Col.	Mini	Maxi
1	-2.658e+00	7.862e+00
2	-3.245e+00	2.810e+00
3	-2.760e+00	2.518e+00

File Nega.ivli contains the predicted row scores (dependent variables after projection)

It has 49 rows and 3 columns

File :Nega.ivli

Col.	Mini	Maxi
1	-3.093e+00	2.839e+00
2	-2.970e+00	2.568e+00
3	-2.021e+00	2.299e+00

File Nega.ivli contains the canonical row scores which are linear combinations of the explanatory variables (unity norm)

It has 49 rows and 3 columns

L'appariement par Scatters : Match two scatters (ci-dessus) est particulièrement explicite pour comprendre ce qui se passe. Une variance forte et une prédiction faible passent devant une variance faible et une prédiction forte dans la seconde analyse.

The screenshot shows the 'Lines' window with the following settings:

- X file (default = 1, 2, 3, ..., n): Nega.ivls
- Y file (no default): Posi.ivls

The 'Curves' window displays a scatter plot with data points. The 'Min/Max' window has the following settings:

- Min. abscissa: -4
- Max. abscissa: 9
- Min. ordinate: -4
- Max. ordinate: 9
- Window height: 300
- Window width: 300
- Options: Square drawing, Draw frame, Scale box

Les deux analyses ont les mêmes coordonnées car on ne peut faire mieux que 70% de la variance de l'effet taille dans un cas et on ne peut faire mieux que 30% de cette variance dans l'autre. On observera combien dans ce cas une maîtrise des principes en jeu s'impose pour éviter les pièges dans l'interprétation.

L'axe 2 de l'ACPVI orthogonale est consacrée à un effet inter-annuel.

Noter encore que les analyses intra-classes sont des ACPVI orthogonales, alors que les ACP inter-classes sont des ACPVI ordinaires. Les versions AFC de ces analyses sont disponibles dans les mêmes options, en changeant l'analyse préliminaires.



1 Lebart, L., Morineau, A. & Piron, M. (1995) *Statistique exploratoire multidimensionnelle*. Dunod, Paris. 1-439. Cf. pages 320-323.

2 Usseglio-Polatera, P. & Auda, Y. (1987) Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie* : 23, 1, 65-79.

Projectors : PCA on Instrumental Variables



Méthode d'analyse de données dite ACPVI (ACP sur Variables Instrumentales) ou ACL (Analyse sous Contraintes Linéaires). On peut exécuter avec ce programme tout ou partie de l'analyse en composantes explicatives ¹, l'analyse canonique des correspondances ², les ACP sur variables instrumentales ³, les analyses de redondance ⁴, l'analyse canonique des correspondances partielles ⁵.



L'option projette un tableau de données sur un sous-espace de référence, fait l'analyse du tableau projeté et propose diverses aides à l'interprétation. En fonction du tableau utilisé, de l'origine de l'espace de projection, de l'objectif fixé et de la structure des données, les résultats obtenus sont très variés.



L'option est un module complet de CANOCO ², à l'exception des options de detrending qui sont hors du champ du modèle euclidien.



L'option utilise une seule fenêtre de dialogue :

PCA on Instrumental Variables

Explanatory variables: .@ob file

Dependant variables: .**ta

Output file name

Nom du fichier .@ob contenant une base orthonormée du sous-espace de projection. Les options Projectors : Table->Orthonormal Basis, Triplet->Orthonormal Basis, One Categ Var->Orthonormal Basis, Combine 2 Orthonormal Bases, Two Categ Var->Orthonormal Bases, et Intersection of 2 Subspaces créent de tels fichiers.

Nom du fichier ---.##ta associée à l'analyse d'un tableau. Les modules PCA, HTA, MCA et COA créent de tels fichiers.



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.

Nom générique des fichiers de sortie (création).



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

COrrrespondence Analysis

Data file Veg 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 Mil 12 4

Row weights (default=1/n) 3

Column weights (default=1)

Option: file for row weighting Veg.fcpl 12 1

Calculer le sous-espace associé à ce tableau normalisé (Projectors : Triplet->Orthonormal Basis) :

Triplet->Orthonormal Basis

Explanatory variables Mil.cnta 12 4

Option: output file name

L'analyse canonique des correspondances du couple Mil/Veg est obtenue par :

PCA on Instrumental Variables			
Explanatory variables: .@ob file	<input type="text" value="Mil.@ob"/>	12	4
Dependant variables: .**ta	<input type="text" value="Veg.fcta"/>	12	22
Output file name	<input type="text" value="A"/>		

Instrumental variables
 ----- input -----
 Orthonormal basis: Mil.@ob
 It has 12 rows and 4 columns
 Dependent 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 weight
 It has 22 rows and 1 column
 File A.ivpl contains the row weight
 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 the canonical weights
 coefficients of linear combination of the explanatory variables (unity norm)
 It has 4 rows and 2 columns

A.ivfa		
	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 the canonical row scores which are
 linear combinations of the explanatory variables (unity 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 (dependent variables) with lambda norm
 It has 22 rows and 2 columns
 File :A.ivco

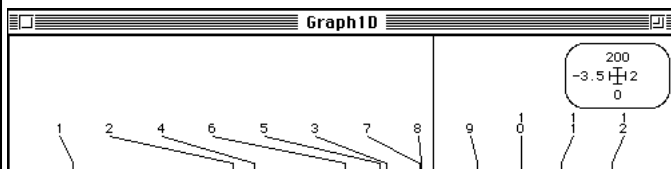
Col.	Mini	Maxi
1	-2.304e+00	1.476e+00
2	-2.035e+00	1.291e+00

Pour illustrer la démarche, transposer le tableau de végétation :

Transpose

Input file 12 22

Output file



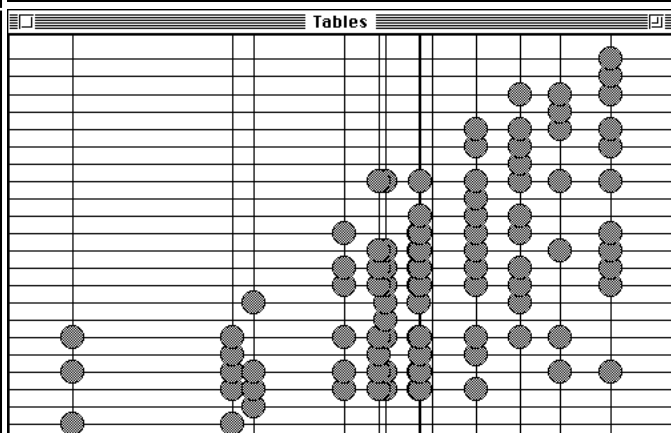
Labels

Data file (no default)

Rows label file (default = #)

Variable label file (or #)

Vertical (1) or horizontal (2) graphs



Values

Input table file

H-axis position file

Column number (default = 1)

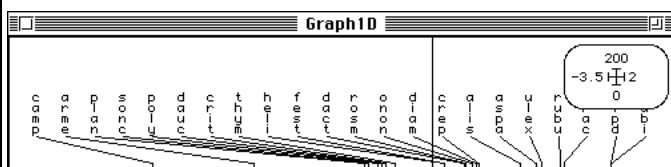
Y-axis position file

Column number (default = 1)

H-axis: Ordination (1) or Ranking (2)

Y-axis: Ordination (1) or Ranking (2)

Grid (yes = 1)



Labels

Data file (no default)

Rows label file (default = #)

Variable label file (or #)

Vertical (1) or horizontal (2) graphs

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.iv1s
      A.iv1i
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 (dependent variables, unity 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 :

The image shows three screenshots of software windows and two configuration panels. The top 'Graph1D' window shows a plot with a callout box containing a matrix: $\begin{pmatrix} 100 \\ -4 & 3 \\ 0 \end{pmatrix}$. The middle 'Tables' window shows a grid of points. The bottom 'Graph1D' window shows a plot with a callout box containing a matrix: $\begin{pmatrix} 200 \\ -4 & 3 \\ 0 \end{pmatrix}$. The 'Labels' panels are configured with 'Data file (no default)' as 'A.ivc1', 'Rows label file (default = #)' as 'Label_Esp', 'Variable label file (or #)' as empty, and 'Vertical (1) or horizontal (2) graphs' as '2'. The 'Values' panel is configured with 'Input table file' as 'Ueg', 'H-axis position file' as 'A.ivc1', 'Column number (default = 1)' as empty, 'Y-axis position file' as 'A.ivls', 'Column number (default = 1)' as empty, 'H-axis: Ordination (1) or Ranking (2)' as '1', and 'Y-axis: Ordination (1) or Ranking (2)' as '2'.

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

File A.ivls contains the row scores (dependent 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 (R²). Cette décomposition fondamentale est simplement indiquée dans :

Num	Variance	R ²	EigenVal.
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 the predicted row scores (dependent 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 discussions 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.



¹ Obadia, J. (1978) L'analyse en composantes explicatives. *Revue de Statistique Appliquée* : 24, 4, 5-28.

² 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. (1987) 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.

Ter Braak, C.J.F. (1987) *Unimodal models to relate species to environment*. Agricultural Mathematics Group, Box 100, NL-6700, AC Wageningen, The Netherlands. 1-152.

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.

Lebreton, J.D., Chessel, D., Prodon, R. & Yoccoz, N. (1988) L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. *Acta Oecologica, Oecologia Generalis* : 9, 1, 53-67.

Lebreton, J.D., Richardot-Coulet, M., Chessel, D. & Yoccoz, N. (1988) L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. II Variables de milieu qualitatives. *Acta Oecologica, Oecologia Generalis* : 9, 2, 137-151.

Lebreton, J.D., Sabatier, R., Banco, G. & Bacou, A.M. (1991) Principal component and correspondence analyses with respect to instrumental variables : an overview of their role in studies of structure-activity and species- environment relationships. In : *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers, J. & Karcher, W. (Eds.) Kluwer Academic Publishers. 85-114.

³ Rao, C.R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhya, A* : 26, 329-359.

Sabatier, R. (1983) *Approximations d'un tableau de données. Application à la reconstitution des paléoclimats*. Thèse de 3^o cycle, Université de Montpellier. 1-184.

Sabatier, R. (1987) *Méthodes factorielles en analyse des données : approximations et prise en compte de variables concomitantes*. Thèse de doctorat d'état. Université de Montpellier. 1-224.

Sabatier, R., Lebreton, J.D. & Chessel, J.D. (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. In : *Multivariate data analysis*. Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland. 341-352.

⁴ Wollenberg, A.L. (1977) Redundancy analysis, an alternative for canonical analysis. *Psychometrika* : 42, 2, 207-219.

Johansson, J.K. (1981) An extension of Wollenberg's redundancy analysis. *Psychometrika* : 46, 93-103.

Israels, A.Z. (1984) Redundancy analysis for qualitative variables. *Psychometrika* : 49, 661-346.

⁵ Ter Braak, C.J.F. (1988) Partial Canonical Analysis. In : *Classification and related methods of data analysis*. Bock, H.H. (Ed.) North Holland. 551-558.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology* : 73, 1045-1055.

Yoccoz, N. & Chessel, D. (1988) Ordination sous contraintes de relevés d'avifaune : élimination d'effets dans un plan d'observations à deux facteurs. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : III, 307 : 189-194.

Projectors : Subspace Test



Test de Monte Carlo sur l'inertie projetée.



A est un sous-espace de \mathbb{R}^n connu par une base orthonormée **A** pour la pondération **D**. **A** et **D** sont fixés *a priori* indépendamment du tableau **X**. **X** est un tableau de donnée à n lignes et p colonnes et **Q** est une métrique de \mathbb{R}^p . **Q** peut être définie par **X** mais doit être invariante par permutation des lignes de **X**. **X** subit une transformation préliminaire, ce qui donne un triplet $(\mathbf{X}_T, \mathbf{Q}, \mathbf{D})$. Le projecteur **D**-orthogonal sur le sous-espace A est noté \mathbf{P}_A . L'inertie totale du triplet $(\mathbf{P}_A(\mathbf{X}_T), \mathbf{Q}, \mathbf{D})$ est une fraction de l'inertie totale du triplet $(\mathbf{X}_T, \mathbf{Q}, \mathbf{D})$. Sous l'hypothèse nulle que chacune des $n!$ permutations des lignes de **X** est équiprobable, on sélectionne m permutations aléatoires et on compare le taux observé à la distribution des m valeurs obtenues par simulation.

Cette opération est acceptée pour les transformations associées aux triplets des options cp (ACP centrée), cn (ACP normée), cm (ACM), fl (ACF) et nc (ACP non centrée). Dans tous les cas **la pondération des lignes doit être fixée indépendamment du tableau**, en particulier la pondération uniforme ou une pondération fixée par ailleurs possible par l'option COA : Row weighted COA.



L'option utilise une seule fenêtre de dialogue :

Icon: Nom du fichier .@ob contenant une base orthonormée du sous-espace de projection. Les options Projectors : Table->Orthonormal Basis, Triplet->Orthonormal Basis, One Categ Var->Orthonormal Basis, Combine 2 Orthonormal Bases, Two Categ Var->Orthonormal Bases, et Intersection of 2 Subspaces créent de tels fichiers.

Icon: Nom du fichier ---.##ta associée à l'analyse d'un tableau. Les modules PCA, HTA, MCA et COA créent de tels fichiers.



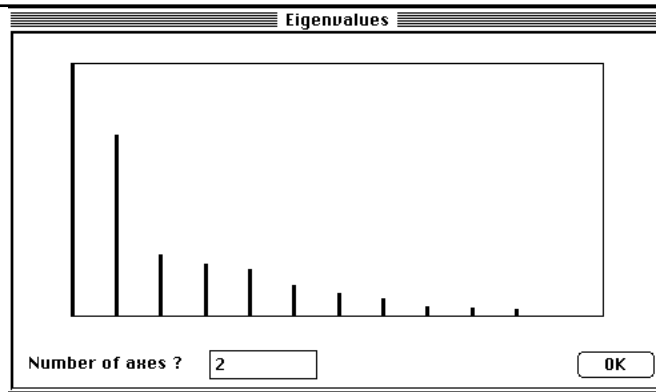
Le fichier .@ob est toujours associé à un fichier .@pl qui contient la pondération pour laquelle la base utilisée est orthonormée. Le fichier ---.##ta est toujours associé à un fichier ---.##pl qui contient la pondération des lignes dans l'analyse préliminaire utilisée. Une seule contrainte est imposée : ces deux pondérations doivent être identiques et **fixées indépendamment du tableau projeté**.

Icon: Nombre de permutations utilisées.



Utiliser le dossier de travail créé par la carte Light_trap de la pile ADE-4•Data. Transformer le tableau faunistique (Bin->Bin : $c \cdot \text{Log}[a \cdot x + b]$) :

Faire l'ACP normée du tableau des variables météorologiques :



DiagoRC: General program for two diagonal inner product analysis
 Input file: Cope.fcta
 --- Number of rows: 42, columns: 12

 Total inertia: 0.471293

L'inertie totale multipliée par 900 vaut 424.16 (valeur utilisée dans l'article cité).

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+1.7168E-01	+0.3643	+0.3643	02	+1.2291E-01	+0.2608	+0.6251
03	+4.1600E-02	+0.0883	+0.7133	04	+3.5995E-02	+0.0764	+0.7897
05	+3.1638E-02	+0.0671	+0.8568	06	+2.1497E-02	+0.0456	+0.9024
07	+1.6193E-02	+0.0344	+0.9368	08	+1.1930E-02	+0.0253	+0.9621
09	+7.1140E-03	+0.0151	+0.9772	10	+5.7645E-03	+0.0122	+0.9894
11	+4.9729E-03	+0.0106	+1.0000	12	+0.0000E+00	+0.0000	+1.0000

Les résultats sont ceux du tableau 2 (page 1145).

Implanter un plan d'expérience complet à deux variables (TextToBin : Create2Categ) :

Toute la suite tient sur le fait que la pondération de l'AFC est uniforme à cause des précautions prises par les expérimentateurs (30 mesures, exactement, par lignes du tableau).

Implanter les sous-espaces de projection (Projectors : Two Categ Var->Orthonormal Bases) :

Subspaces from two categorical variables

 Input file: DurTyp
 It has 42 rows and 2 columns
 Generic output file name: DT
 Crossing variable A (n° 1) and B (n° 2)

 File DT_AxB.@ob contains an orthonormal basis of the subspace AxB
 It has 42 rows and 41 columns
 File DT_A+B.@ob contains an orthonormal basis of the subspace A+B
 It has 42 rows and 11 columns
 File DT_A•B.@ob contains an orthonormal basis of the subspace A•B
 It has 42 rows and 30 columns
 File DT_A.@ob contains an orthonormal basis of the subspace A
 It has 42 rows and 5 columns

File DT_B.@ob contains an orthonormal basis of the subspace B
 It has 42 rows and 6 columns
 File DT_A/B.@ob contains an orthonormal basis of the subspace A/B
 It has 42 rows and 5 columns
 File DT_B/A.@ob contains an orthonormal basis of the subspace B/A
 It has 42 rows and 6 columns

L'effet A est la durée et l'effet B est le type de mucus.

La participation des effets A et B à l'inertie totale sont (Projectors : Triplet Inertia Decomposition) :

Triplet Inertia Decomposition					
Explanatory variables: .@ob file		DT_A.@ob	42	5	
Dependant variables: .**ta		Cope.fcta	42	12	

Triplet Inertia Decomposition					
Explanatory variables: .@ob file		DT_B.@ob	42	6	
Dependant variables: .**ta		Cope.fcta	42	12	

Orthonormal basis: DT_A.@ob
 It has 42 rows and 5 columns
 Dependent variable file: Cope.fcta
 It has 42 rows and 12 columns

	Subspace A	A Orthogo	Total	A+	A-
1	8.4656e-03	6.2434e-02	7.0899e-02	1194	8805
2	1.1552e-02	3.9859e-02	5.1411e-02	2246	7753
3	1.0053e-02	2.4339e-02	3.4392e-02	2923	7076
...					
11	1.0819e-02	3.5756e-02	4.6575e-02	2322	7677
12	2.1610e-02	4.4898e-02	6.6508e-02	3249	6750
Tot	9.6401e-02	3.7489e-01	4.7129e-01	2045	7954

Orthonormal basis: DT_B.@ob
 It has 42 rows and 6 columns
 Dependent variable file: Cope.fcta
 It has 42 rows and 12 columns

	Subspace A	A Orthogo	Total	A+	A-
1	1.6578e-02	5.4321e-02	7.0899e-02	2338	7661
2	1.2934e-02	3.8477e-02	5.1411e-02	2515	7484
3	1.1861e-02	2.2531e-02	3.4392e-02	3448	6551
...					
11	2.1721e-02	2.4854e-02	4.6575e-02	4663	5336
12	2.3810e-02	4.2698e-02	6.6508e-02	3579	6420
Tot	1.7380e-01	2.9749e-01	4.7129e-01	3687	6312

Ce sont des parties de l'inertie expliquée par l'effet additif :

Triplet Inertia Decomposition					
Explanatory variables: .@ob file		DT_A+B.@ob	42	11	
Dependant variables: .**ta		Cope.fcta	42	12	

Orthonormal basis: DT_A+B.@ob
 It has 42 rows and 11 columns
 Dependent variable file: Cope.fcta
 It has 42 rows and 12 columns

	Subspace A	A Orthogo	Total	A+	A-
1	2.5044e-02	4.5855e-02	7.0899e-02	3532	6467
2	2.4486e-02	2.6925e-02	5.1411e-02	4762	5237
3	2.1914e-02	1.2478e-02	3.4392e-02	6371	3628
...					

```

...
  11 | 3.2540e-02 | 1.4035e-02 | 4.6575e-02 | 6986 | 3013 |
  12 | 4.5420e-02 | 2.1088e-02 | 6.6508e-02 | 6829 | 3170 |
-----|-----|-----|-----|-----|-----|
Tot  | 2.7020e-01 | 2.0109e-01 | 4.7129e-01 | 5733 | 4266 |
-----|-----|-----|-----|-----|-----|

```

D'où la décomposition de l'inertie totale en fonction de l'inertie de départ ::

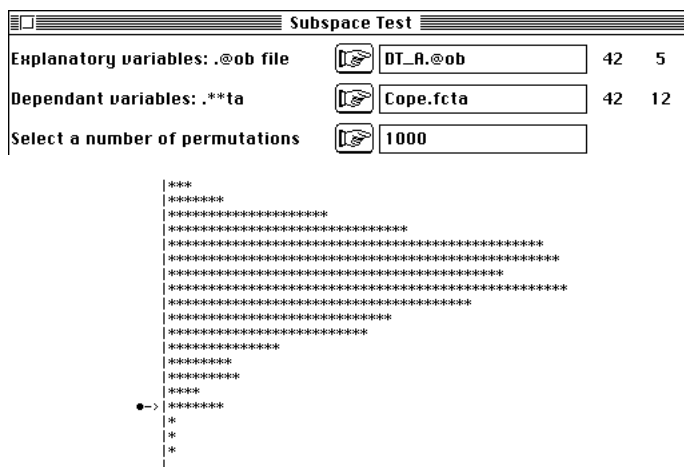
Effet A+B ---> 57.33%

Effet A ---> 20.45%

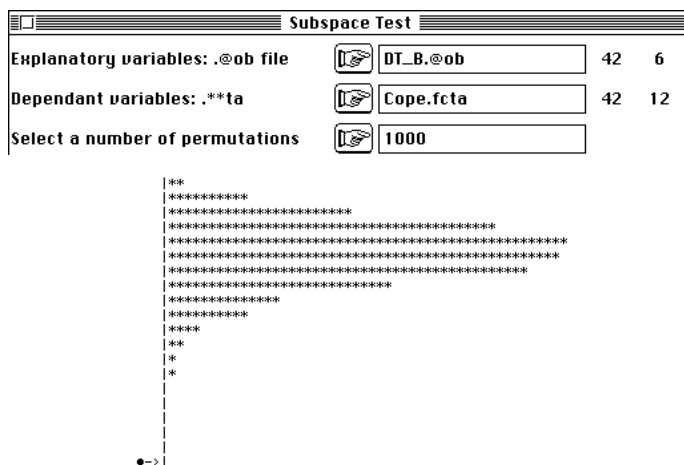
Effet B ---> 36.87%

Cette somme ne tient que parce que le plan est orthogonal (30 mesures par modalité croisée des deux facteurs). Les sous-espaces A, A/B et A B sont confondus (§4.4 p. 1150).

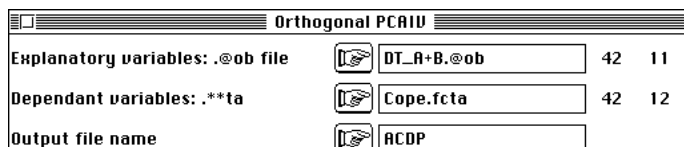
Le rôle de l'effet A est significatif au seuil de 5% (Projectors : Subspace Test) :



Le rôle de l'effet B est significatif au seuil de 1‰ (Projectors : Subspace Test) :



Les auteurs achèvent l'article en proposant l'ACPVI orthogonale sur le sous-espace A+B (Projectors : Orthogonal PCAIV) :



```

Instrumental variables
----- input -----
Orthonormal basis: Temp.@ob
It has 42 rows and 11 columns

```


Dependent variable file: Cope.fcta
 It has 42 rows and 12 columns
 ----- output -----
 Projected variable file: ACDP.ivta
 It has 42 rows and 12 columns
 Inertia: 2.0109e-01
 File ACDP.ivpc contains the column weight
 It has 12 rows and 1 column
 File ACDP.ivpl contains the row weight
 It has 42 rows and 1 column

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+6.4473E-02	+0.3206	+0.3206	02	+3.4618E-02	+0.1722	+0.4928
03	+2.7513E-02	+0.1368	+0.6296	04	+1.9788E-02	+0.0984	+0.7280
05	+1.6854E-02	+0.0838	+0.8118	06	+1.2595E-02	+0.0626	+0.8744
07	+8.8079E-03	+0.0438	+0.9182	08	+6.1054E-03	+0.0304	+0.9486
09	+5.0964E-03	+0.0253	+0.9739	10	+3.1421E-03	+0.0156	+0.9896
11	+2.1001E-03	+0.0104	+1.0000				

On retrouve les résultats du tableau 4 (p. 1151).

Les diverses propositions des auteurs pour le cas où le plan ne serait pas orthogonal sont en outre accessibles par les options du présent module.



¹ 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.

Projectors : Table Projection



Utilitaire de régression multiple.



L'objectif est le plus simple parmi ceux qui utilise un sous-espace. On veut simplement projeter les colonnes d'un tableau sur le sous-espace choisi.



L'option utilise une seule fenêtre de dialogue :

Table Projection	
Explanatory variables: .@ob file	<input type="text"/>
Dependant variables	<input type="text"/>
Output file name	<input type="text"/>

Nom du fichier binaire d'entrée.

Nom du fichier binaire de sortie (création).



Dialogues intermédiaires



Exemple : faire un modèle additif du tableau :

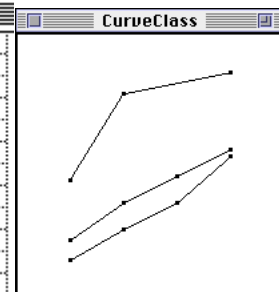
Exo.txt			
	1	2	3
1	1.2	1.8	3.6
2	2.1	2.9	•
3	2.9	3.7	6.2
4	4.3	4.5	6.8

La même information est séparée dans deux fichiers Exo (11-2) et data (11-1) :

Exo.txt			
	1	2	3
1	1.2	1	1
2	2.1	2	1
3	2.9	3	1
4	4.3	4	1
5	1.8	1	2
6	2.9	2	2
7	3.7	3	2
8	4.5	4	2
9	3.6	1	3
10	6.2	2	3
11	6.8	4	3

Exo		
	1	2
1	1.0000	1.0000
2	2.0000	1.0000
3	3.0000	1.0000
4	4.0000	1.0000
5	1.0000	2.0000
6	2.0000	2.0000
7	3.0000	2.0000
8	4.0000	2.0000
9	1.0000	3.0000
10	2.0000	3.0000
11	4.0000	3.0000

data	
	1
1	1.2000
2	2.1000
3	2.9000
4	4.3000
5	1.8000
6	2.9000
7	3.7000
8	4.5000
9	3.6000
10	6.2000
11	6.8000



Lire le fichier descriptif du plan (CategVar : Read Categ File) :

Read Categ File	
Input file	<input type="text" value="Exo"/> 11 2

Passer le fichier en disjonctif complet (CategVar : Categ->Disj) :

Categ->Disj	
---.cat type file	<input type="text" value="Exo.cat"/>

Orthonormaliser le tableau (Projectors : Table->Orthonormal Basis) :

Table->Orthonormal Basis	
Explanatory variables	<input type="text" value="Exo01"/> 11 7
Option: row weighting	<input type="text"/>
Option: output file name	<input type="text"/>

Projeter la variable :

Table Projection			
Explanatory variables: .@ob file	<input type="button" value="Browse"/>	Exo01.@ob	11 6
Dependant variables	<input type="button" value="Browse"/>	data	11 1
Output file name	<input type="button" value="Browse"/>	data+	

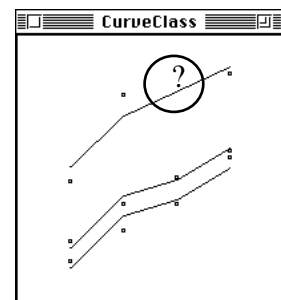
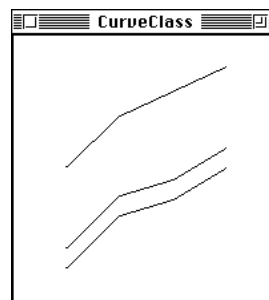
Orthonormal basis: Exo01.@ob
 It has 11 rows and 6 columns
 Dependent variable file: data
 It has 11 rows and 1 columns

	Subspace A	A Orthogo	Total		A+	A-
1	1.5900e+01	9.8586e-02	1.5998e+01		9938	61

File data+.mod contains the predicted variables
 It has 11 rows and 1 columns

File data+.res contains the (observed - predicted) values
 It has 11 rows and 1 columns

Lines	
X file (default = 1, 2, 3, ..., n)	<input type="button" value="Browse"/> Exo
X file column number (default = 1)	<input type="button" value="Browse"/>
Y file (no default)	<input type="button" value="Browse"/> data+.mod
Categories file (.cat)	<input type="button" value="Browse"/> Exo.cat
Categories file column number	<input type="button" value="Browse"/> 2



Reste encore un problème : comment obtenir la valeur du modèle pour la donnée manquante ?

Projectors : Table->Orthonormal Basis



Utilitaire d'orthonormalisation des colonnes d'un tableau.



Un tableau de données quelconque à n lignes peut définir par orthonormalisation une base de vecteurs orthonormés du sous-espace engendré par ses colonnes.



L'option utilise une seule fenêtre de dialogue :

Table->Orthonormal Basis		
Explanatory variables		vivo 34 4
Option: row weighting		
Option: output file name		

Nom du fichier binaire d'entrée.

Option : nom du fichier des poids des lignes. Par défaut, la pondération uniforme est utilisée.

Option : nom générique des fichiers créés. Par défaut, c'est le nom du fichier d'entrée.



Utiliser le dossier de travail créé par la carte ToxiCornée¹ de la pile ADE-4•Data.

```
Orthonormalization: subspace generated by quantitative variables
-----
```

```
Explanatory variable file: vivo
It has 34 rows and 4 columns
-----
```

```
Orthonormal basis: vivo.@ob
It has 34 rows and 4 columns
Row weight file: vivo.@pl
Uniform row weight = 0.029412
Coordinates of the vectors of the orthonormal basis
in the initial basis in : vivo.@co
File vivo.@co has 4 rows and 4 columns
-----
```

Lorsque le tableau est de rang plein (la dimension du sous-espace créé est égal au nombre de colonnes) le fichier ---.@co contient en lignes les coordonnées dans la nouvelle base des colonnes du tableau traité.

Table Projection		
Explanatory variables: .@ob file		vivo.@ob 34 4
Dependant variables		vitro 34 4
Output file name		VV

```
Orthonormal basis: vivo.@ob
It has 34 rows and 4 columns
Dependent variable file: vitro
It has 34 rows and 4 columns
```

	Subspace A	A Orthogo	Total	A+	A-
1	5.6421e+02	6.9498e+01	6.3371e+02	8903	1096
2	1.0126e+03	1.1465e+02	1.1273e+03	8982	1017
3	1.8646e+03	2.1682e+02	2.0815e+03	8958	1041
4	3.1276e+03	4.0987e+02	3.5375e+03	8841	1158

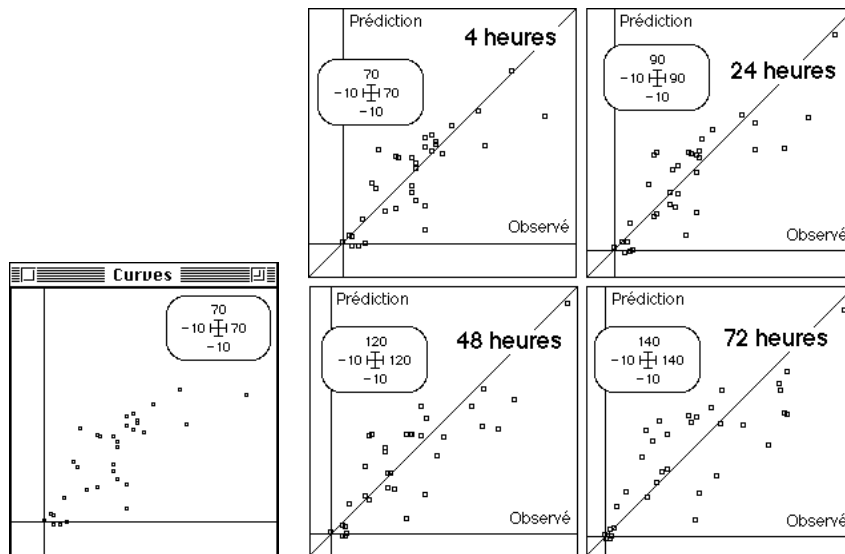
```
File VV.mod contains the predicted variables
It has 34 rows and 4 columns
```

```
File VV.coe contains the canonical weights
coefficients of linear combination of explanatory variables
It has 4 rows (explanatory v.) and 4 columns (dependent v.)
```

File VV.res contains the (observed - predicted) values
It has 34 rows and 4 columns

On vient de faire la régression par l'origine des variables (sans terme constant) des variables du fichier vitro sur les quatre variables du fichiers vitro. Représenter les données et les modèles des régressions qui précèdent (assembler la figure dans un grapheur) :

Lines		Min/Ma	
X file (default = 1, 2, 3, ..., n)	<input type="text" value="vitro"/>	Min. abscissa:	<input type="text" value="-10"/>
X file column number (default = 1)	<input type="text" value="1"/>	Max. abscissa:	<input type="text" value="70"/>
Y file (no default)	<input type="text" value="VV.mod"/>	Min. ordinate:	<input type="text" value="-10"/>
Cumulated data (1=yes, 2=no)	<input type="text" value=""/>	Max. ordinate:	<input type="text" value="70"/>
Variable label file (or #)	<input type="text" value=""/>	Window height:	<input type="text" value="200"/>
Draw curves (1=yes, 2=no)	<input type="text" value="2"/>	Window width:	<input type="text" value="200"/>



La qualité des prédictions est remarquablement stable de même que les coefficients des équations de régression :

	1	2	3	4
1	0.3810	0.5861	0.8849	0.9846
2	0.0117	-0.0933	-0.1596	-0.0086
3	0.4484	0.5421	0.6603	0.7471
4	-0.2384	-0.1980	-0.2077	-0.2400

ce qui donne (V pour mesure in vivo) :

$$\text{Vitro 04 h} = 0.3810 * V1/2h + 0.0117 * V1h + 0.4484 * V2h + -0.2384 * V4h$$

$$\text{Vitro 24 h} = 0.5861 * V1/2h - 0.0933 * V1h + 0.5421 * V2h + -0.1980 * V4h$$

$$\text{Vitro 48 h} = 0.8849 * V1/2h - 0.1596 * V1h + 0.6603 * V2h + -0.2077 * V4h$$

$$\text{Vitro 72 h} = 0.9846 * V1/2h - 0.0086 * V1h + 0.7471 * V2h + -0.2400 * V4h$$

La seconde variable explicative est systématiquement disqualifiée par une valeur aberrante. Le rôle de la quatrième laisse à penser à des différences de processus entre les deux types de mesure.



Si on veut faire de la régression avec terme constant, ajouter le vecteur des constantes aux tableaux des explicatives (FilesUtil : Add column 1n).



¹ Jacobs, G.A. & Martens, M.A. (1990) Quantification of eye irritation based upon in vitro changes of corneal thickness. *ATLA* : 17, 255-262.

Projectors : Triplet Inertia Decomposition



Décomposition de l'inertie dans une projection de tableau.



Edition des normes (en général les variances) expliquées et résiduelles dans la projection sur un sous-espace.



L'option utilise une seule fenêtre de dialogue :

Triplet Inertia Decomposition		
Explanatory variables: .@ob file	<input type="text" value="Fau"/>	49 11
Dependant variables: .**ta	<input type="text" value="FauLog"/>	49 17

Nom du fichier .@ob contenant une base orthonormée du sous-espace de projection. Les options Projectors : Table->Orthonormal Basis, Triplet->Orthonormal Basis, One Categ Var->Orthonormal Basis, Combine 2 Orthonormal Bases, Two Categ Var->Orthonormal Bases, et Intersection of 2 Subspaces créent de tels fichiers.

Nom du fichier ---.##ta associée à l'analyse d'un tableau. Les modules PCA, HTA, MCA et COA créent de tels fichiers.



Utiliser le dossier de travail créé par la carte Light_trap de la pile ADE-4•Data. Transformer le tableau faunistique (Bin->Bin : c*Log[a*x+b]) :

c*Log[a*x+b]		
Input file	<input type="text" value="Fau"/>	49 17
Output file	<input type="text" value="FauLog"/>	
Selection of columns (default = all)	<input type="text"/>	
Parameter a (default=1)	<input type="text" value="1"/>	
Parameter b (default=0)	<input type="text" value="1"/>	
Parameter c (default=1)	<input type="text" value="1"/>	

Faire l'ACP normée du tableau des variables météorologiques et garder deux facteurs :

Correlation matrix PCA		
Matrix input file	<input type="text" value="Meteo"/>	49 11

Eigenvalues	
Number of axes ?	<input type="text" value="2"/>
OK	

Créer le sous-espace associé aux tableaux des deux coordonnées (technique de régression sur composantes) :

Table->Orthonormal Basis		
Explanatory variables	<input type="text" value="Meteo.cnli"/>	49 2
Option: row weighting	<input type="text"/>	
Option: output file name	<input type="text" value="Meteo2Fad"/>	

Faire l'ACP centrée du tableau transformé (PCA : Covariance matrix PCA) :

Covariance matrix PCA		
Matrix input file	<input type="text" value="FauLog"/>	49 17

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		Meteo2Fac.@ob	49 2
Dependant variables: .**ta		FauLog.cpta	49 17

Projected inertia on a subspace

 Orthonormal basis: Meteo2Fac.@ob
 It has 49 rows and 2 columns
 Dependent variable file: FauLog.cpta
 It has 49 rows and 17 columns

	Subspace A	A Orthogo	Total		A+	A-
1	6.5147e-04	2.6963e-02	2.7615e-02		235	9764
2	1.9130e-03	2.5702e-02	2.7615e-02		692	9307
3	3.8324e-02	4.8426e-01	5.2258e-01		733	9266
4	3.5103e-04	5.1276e-02	5.1627e-02		67	9932
5	1.1491e+00	9.8873e-01	2.1378e+00		5375	4624
6	1.4264e-01	5.8890e-01	7.3154e-01		1949	8050
7	3.4181e-02	1.0776e-01	1.4194e-01		2408	7591
8	1.3103e-02	1.0668e-01	1.1978e-01		1093	8906
9	1.8846e-02	5.8375e-02	7.7221e-02		2440	7559
10	1.7958e-01	4.2226e-01	6.0184e-01		2983	7016
11	4.6449e-01	5.9629e-01	1.0608e+00		4378	5621
12	2.5429e-02	3.6384e-01	3.8927e-01		653	9346
13	1.3375e-01	1.1351e+00	1.2688e+00		1054	8945
14	1.4967e-01	5.8081e-01	7.3048e-01		2048	7951
15	4.2420e-01	1.0922e+00	1.5164e+00		2797	7202
16	1.3723e-02	7.1209e-02	8.4932e-02		1615	8384
17	2.6544e-01	8.8491e-01	1.1503e+00		2307	7692
Tot	3.0554e+00	7.5852e+00	1.0641e+01		2871	7128

On lit pour chaque espèce (1 à 17 en lignes), la variance de l'abondance du taxon (colonne Total), la variance expliquée par la régression sur les deux composantes (colonne Subspace A) et la variance résiduelle (colonne A Orthogo), le pourcentage de variance expliquée (colonne A+) qui est donc un R^2 (carré de corrélation multiple) et le pourcentage de variance résiduelle (colonne A-). La capacité de prédiction varie de 0 à 54%. En bas totaux pondéré correspondant respectivement aux inertie totales de l'ACPVI (colonne Subspace A), de l'ACPVI orthogonale (colonne A Orthogo) et de l'analyse initiale de FauLog.cpta (colonne Total). A droite pourcentage d'inertie expliquée (colonne A+) et d'inertie résiduelle (colonne A-). On retrouve ces valeurs dans :

PCA on Instrumental Variables			
Explanatory variables: .@ob file		Meteo2Fac.@ob	49 2
Dependant variables: .**ta		FauLog.cpta	49 17
Output file name		Posi2Fac	

Projected variable file: Posi2Fac.ivta
 It has 49 rows and 17 columns
 Inertia: 3.0554e+00

Orthogonal PCAIV			
Explanatory variables: .@ob file		Meteo2Fac.@ob	49 2
Dependant variables: .**ta		FauLog.cpta	49 17
Output file name		Nega2Fac	

Projected variable file: Nega2Fac.ivta
 It has 49 rows and 17 columns
 Inertia: 7.5852e+00

Subspace Test			
Explanatory variables: .@ob file		Meteo2Fac.@ob	49 2
Dependant variables: .**ta		FauLog.cpta	49 17
Select a number of permutations		1000	

Projected inertia on a subspace

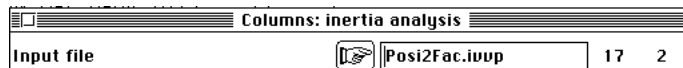
Orthonormal basis: Meteo2Fac.@ob
It has 49 rows and 2 columns
Dependent variable file: FauLog.cpta
It has 49 rows and 17 columns

number of permutations: 1000 Observed: 0.287144
Histogram: minimum = 0.008450, maximum = 0.287144
number of simulations X<Obs: 1000 (frequency: 1.000000)
number of simulations X>=Obs: 0 (frequency: 0.000000)

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

•->

Noter la cohérence d'ADE-4. L'ACPVI directe est une analyse d'inertie du tableau projeté et donc est une analyse d'inertie (DDUtil : Columns/Inertia analysis) :



Input file: Posi2Fac.ivta
Number of rows: 49, columns: 17

Inertia: Two diagonal norm inertia analysis
Total inertia: 3.05537 - Number of axes: 2

-----Relative contributions-----

Num	Fac 1	Fac 2	Remains	Weight	Cont.
1	7973	2026	0	10000	2
2	9322	677	0	10000	6
3	2487	7512	0	10000	125
4	23	9976	0	10000	1
...					
12	7567	2432	0	10000	83
13	9548	451	0	10000	437
14	9580	419	0	10000	489
15	9875	124	0	10000	1388
16	8290	1709	0	10000	44
17	9430	569	0	10000	868

C'est le facteur 1 de l'ACP de Mil qui fait la quasi totalité de la prédiction.



Usseglio-Polatera, P. & Auda, Y. (1987) Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie* : 23, 1, 65-79.

Voir aussi un large usage de ces techniques dans :

Poizat, G. (1993) *Echelle d'observation et variabilité des abondances de juvéniles de poissons dans un secteur aval du Rhône*. Thèse de Doctorat, Université Lyon 1. 1-217.

Projectors : Triplet->Orthonormal Basis



Utilitaire de création d'une base orthormée d'un sous-espace de \mathbb{R}^n à partir d'un triplet statistique dans lequel le tableau a n lignes.



L'objectif est le même dans l'option Projectors : Table->Orthonormal Basis mais utilise systématiquement la pondération des lignes du triplet pour définir la métrique de \mathbb{R}^n .



L'option utilise une seule fenêtre de dialogue :

Triplet->Orthonormal Basis	
Explanatory variables	<input type="text"/>
Option: output file name	<input type="text"/>

Nom du fichier ---.##ta associée à l'analyse d'un tableau. Les modules PCA, HTA, MCA et COA créent de tels fichiers.

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



Utiliser la carte Tricho de la pile ADE-4•Data. Lire le fichier des stratégies écologiques (FuzzyVar : Read Fuzzy File) :

Read Fuzzy File	
Fuzzy variables: input file (---)	<input type="text"/> 42 24
Category indication file	<input type="text"/> 4 1
Output file name (default = ---F)	<input type="text"/>

En faire l'analyse des correspondances floues (MCA : Fuzzy Correspondence Analysis) :

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

Utiliser le centrage particulier de cette analyse pour en faire un tableau de variables explicatives (sous-espace de projection) :

Triplet->Orthonormal Basis	
Explanatory variables	<input type="text"/> 42 24
Option: output file name	<input type="text"/>

Cette illustration est ici pour montrer qu'on gère dans le module tous les problèmes de rang de matrice sans bricolage. En effet, pour contourner les questions liées aux modules de régression qui ne supportent pas les explicatives redondantes, on propose souvent d'enlever une indicatrice par variable qualitative. Pour les variables floues cette opération serait impossible.

Ajouter dans le dossier de travail le fichier Fau de la carte Tricho+1 de la pile ADE-4•Data et en faire l'ACP centrée par ligne-espèce (HTA : Row centring) :

Row centring	
Matrix input file	<input type="text"/> 42 15

Tester la possibilité de prédire le cortège faunistique par les stratégie des espèces (Projectors : Subspace Test) :

Subspace Test	
Explanatory variables: .@ob file	<input type="text"/> 42 20
Dependant variables: .**ta	<input type="text"/> 42 15
Select a number of permutations	<input type="text"/>

L'opération conduit à un refus :



L'option ne reconnaît pas le centrage par lignes. L'opération est cependant possible. Préparer le régression orthogonale (OrthoVar : Initialize) :

Initialize		
X file: explanatory variables	<input type="text" value="StraF.@ob"/>	42 20
Y file: dependent variables	<input type="text" value="Fau.cltA"/>	42 15
Y transformation (default = none)	<input type="text"/>	
Option: row weighting	<input type="text"/>	
Output file name	<input type="text" value="R"/>	

Exécuter un tests de permutation valide car le centrage par ligne est conservé dans chaque simulation (OrthoVar : Subspace test) :

Subspace test	
Input file	<input type="text" value="R.0Upa"/>
Selection of columns (default = all)	<input type="text"/>
Number of permutations	<input type="text" value="1000"/>

```

----> Explanatory variables: StraF.@ob
----> Dependent variable file: Fau.cltA
----> Transformation used: 0
      0 = None 1 = D-centring, 2 = D-standardization, 3 = D-normalization
----> Row weight file: Uniform_weight
----> Number of random permutation: 1000
----> Selection of explanatory variables: 1a20

```

```

-----
|VarY|r2 observ. | mean sim. | normal I | X>Xobs | Frequency|
-----
| 1 | 4.357e-01 | 3.657e-01 | 9.291e-01 | 193 | 1.930e-01|
| 2 | 6.329e-01 | 4.821e-01 | 1.324e+00 | 97 | 9.700e-02|
| 3 | 6.211e-01 | 4.844e-01 | 1.252e+00 | 110 | 1.100e-01|
| 4 | 4.972e-01 | 4.918e-01 | 4.904e-02 | 488 | 4.880e-01|
| 5 | 2.808e-01 | 4.568e-01 | -1.792e+00 | 967 | 9.670e-01|
| 6 | 3.773e-01 | 4.568e-01 | -7.608e-01 | 771 | 7.710e-01|
| 7 | 3.765e-01 | 4.859e-01 | -1.004e+00 | 851 | 8.510e-01|
| 8 | 5.091e-01 | 4.857e-01 | 2.143e-01 | 409 | 4.090e-01|
| 9 | 6.911e-01 | 4.811e-01 | 1.907e+00 | 29 | 2.900e-02|
|10 | 4.822e-01 | 4.116e-01 | 7.354e-01 | 241 | 2.410e-01|
|11 | 4.672e-01 | 4.487e-01 | 1.748e-01 | 439 | 4.390e-01|
|12 | 4.270e-01 | 4.633e-01 | -3.357e-01 | 626 | 6.260e-01|
|13 | 5.821e-01 | 4.870e-01 | 8.808e-01 | 197 | 1.970e-01|
|14 | 3.581e-01 | 3.584e-01 | -3.843e-03 | 498 | 4.980e-01|
|15 | 4.274e-01 | 3.475e-01 | 9.490e-01 | 172 | 1.720e-01|
-----

```

Globalement, il y a trop d'explicatives pour obtenir une prédiction significative.



Les sous-espaces générés dans ce module permettent les régressions orthogonales du module OrthoVar.

Projectors : Two Categ Var->Orthonormal Bases



Utilitaire de construction de sous-espaces de projection.



Deux sous-espaces vectoriels A et B sont définis par des variables qualitatives contenus dans un fichier lu par CategVar : Read Categ File. On calcule des bases orthonormées de sous-espaces associés aux sous-espaces A et B engendré par les indicatrices des classes de chacune des variables. La pondération **D** des individus (lignes) définissant le produit scalaire utilisé est laissée aux choix de l'utilisateur.



Les sous-espaces A et B sont **D**-centrés, c'est-à-dire sont exactement les intersections des sous-espaces engendrés par les indicatrices avec l'orthogonal du sous-espace engendré par le vecteur dont toutes les composantes égalent l'unité. Dans ces sous-espaces tous les vecteurs sont des variables **D**-centrées par construction.

A est le sous-espace des variables centrées et constantes par classe de la première variable qualitative. B est le sous-espace des variables centrées et constantes par classe de la seconde variable qualitative.

AxB est le sous-espace des variables centrées constantes par cellule du plan d'observation (couple d'une modalité de la première variable qualitative et d'une modalité de la seconde variable qualitative). Cet espace définit les modèles d'interaction. Le sous-espace $A \bullet B$ est le complémentaire orthogonal dans le sous-espace AxB du sous-espace $A+B$ lui-même est l'ensemble des variables centrées somme d'un code numérique constant par classe de la première variable et d'un code numérique de la seconde (modèles additifs).

Le sous espace B/A est la partie de A+B orthogonale à A, ensemble des variables du type A+B centrées par classe de A (modèles dits effet A sachant B). A/B est défini d manière symétrique. Ces sous-espace définissent les analyses partielles, en particulier l'analyse canonique des correspondances partielles.



L'option utilise une seule fenêtre de dialogue :

---.cat type file	Plan.cat
Variable A (default=1)	[]
Variable B (default=2)	[]
Option: row weighting	Avi.fcpl 302 1
Output file name	P

Nom du fichier de type .cat qui contient les deux variables qualitatives.

Numéro de la colonne qui définit le sous-espace A (par défaut, c'est 1).

Numéro de la colonne qui définit le sous-espace B (par défaut, c'est 2).

Fichier de pondération des lignes (par défaut, c'est la pondération uniforme qui est utilisée).

Nom générique des fichiers de sortie (création).

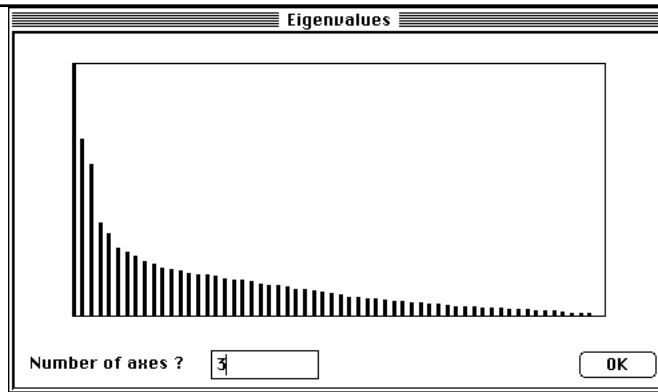


Utiliser le fichier créé par la carte Provence_Corse de la pile ADE-4•Data. Lire le fichier Plan (302-2) par CategVar : Read Categ File :

Input file	Plan 302 2
------------	------------

Exécuter l'AFC du tableau Avi (302-60) par COA : COrrrespondence Analysis :

Data file	Avi 302 60
-----------	------------



Définir les sous-espaces de projection associés au plan d'échantillonnage par Projectors : Two Categ Var->Orthonormal Bases (dialogue ci-dessus) :

Subspaces from two categorical variables

 Input file: Plan
 It has 302 rows and 2 columns
 Generic output file name: P
 Crossing variable A (n° 1) and B (n° 2)

File P_AxB.@ob contains an orthonormal basis of the subspace AxB
 It has 302 rows and 11 columns
 File P_A+B.@ob contains an orthonormal basis of the subspace A+B
 It has 302 rows and 6 columns
 File P_A•B.@ob contains an orthonormal basis of the subspace A•B
 It has 302 rows and 5 columns
 File P_A.@ob contains an orthonormal basis of the subspace A
 It has 302 rows and 1 columns
 File P_B.@ob contains an orthonormal basis of the subspace B
 It has 302 rows and 5 columns
 File P_A/B.@ob contains an orthonormal basis of the subspace A/B
 It has 302 rows and 1 columns
 File P_B/A.@ob contains an orthonormal basis of the subspace B/A
 It has 302 rows and 5 columns

Pour exécuter l'analyse canonique des correspondances partielle sur l'effet végétation sachant région utiliser simplement :

PCA on Instrumental Variables			
Explanatory variables: .@ob file	<input type="button" value="..."/>	P_B/A.@ob	302 5
Dependant variables: .**ta	<input type="button" value="..."/>	Avi.fcta	302 60
Output file name	<input type="button" value="..."/>	PCCA	

----- input -----

Orthonormal basis: P_B/A.@ob
 It has 302 rows and 5 columns
 Dependent variable file: Avi.fcta
 It has 302 rows and 60 columns

----- output -----

Projected variable file: PCCA.ivta
 It has 302 rows and 60 columns
 Inertia: 9.7658e-01
 File PCCA.ivpc contains the column weight
 It has 60 rows and 1 column
 File PCCA.ivpl contains the row weight
 It has 302 rows and 1 column

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+5.1742E-01	+0.5298	+0.5298	02	+2.7001E-01	+0.2765	+0.8063
03	+1.1369E-01	+0.1164	+0.9227	04	+4.7452E-02	+0.0486	+0.9713
05	+2.8003E-02	+0.0287	+1.0000				

File PCCA.ivvp contains the eigenvalues and relative inertia for each axis
 It has 60 rows and 2 columns

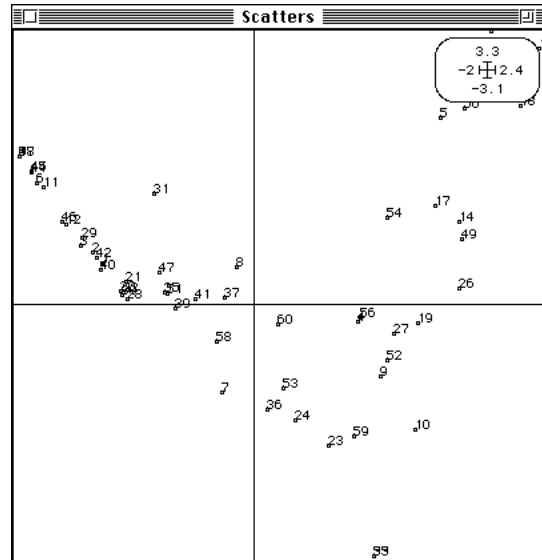
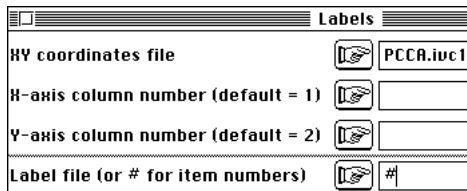
File PCCA.ivcl contains column scores (dependent variables, unity norm)

It has 60 rows and 2 columns

File :PCCA.ivc1

Col.	Mini	Maxi
1	-1.940e+00	2.345e+00
2	-3.035e+00	3.278e+00

Les espèces sont positionnées par des codes numériques de variances unités :

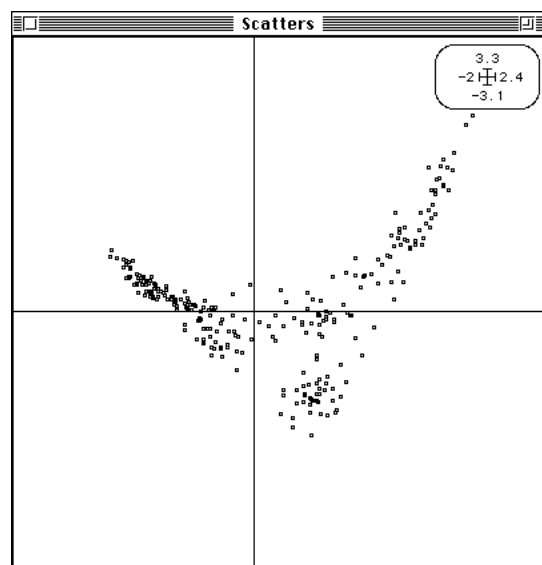
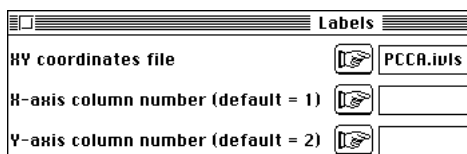


File PCCA.ivls contains the row scores (dependent variables before projection)
It has 302 rows and 2 columns

File :PCCA.ivls

Col.	Mini	Maxi
1	-1.187e+00	1.859e+00
2	-1.485e+00	2.761e+00

Les relevés sont à la moyenne des espèces qu'ils contiennent (averaging classique) :



Ces codes relevés sont de variance expliquée maximale en terme du sous-espace sélectionné. Ils maximisent le produit de la variance par le pourcentage de cette variance expliquée par le modèle choisi :

Num	Variance	R2	EigenVal.
1	0.5979	0.8654	0.5174
2	0.4301	0.6278	0.27

Préparer la variable qualitative qui définit les catégories Régions-Végétation :

Binary->Char

Binary input file: Plan 302 2
 Text output file: Plan_Label

LabelToCateg

Label input file: Plan_Label
 Output file name: P12

CateRowSum-Mean

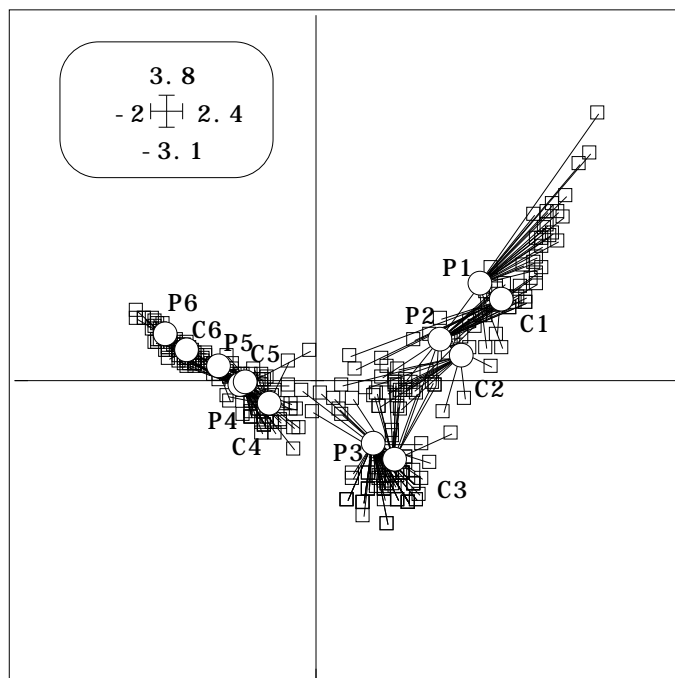
Input file: PCCA.ivli 302 2
 .cat file: P12.cat
 Column number for selection: 1
 Output file: R0
 Option: sum (1) or mean (2): 2

File PCCA.ivli contains the predicted row scores (dependent variables after projection)
 It has 302 rows and 2 columns

File :PCCA.ivli

Col.	Mini	Maxi
1	-1.004e+00	1.213e+00
2	-8.195e-01	1.001e+00

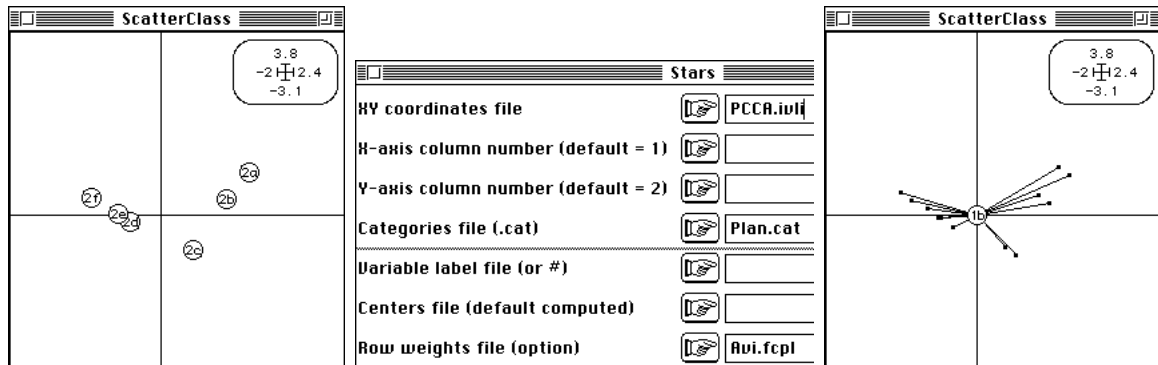
ScatterClass : Stars autorise le tracé d'étoiles qui relie les positions ---.ivls aux moyennes par classes des codes ---.ivli qui ne sont pas les centres de gravité des classes mais la position modèle attachée à tous les éléments d'une classe :



Stars

XY coordinates file: PCCA.ivls
 X-axis column number (default = 1):
 Y-axis column number (default = 2):
 Categories file (.cat): P12.cat
 Variable label file (or #):
 Centers file (default computed): R0

Les espèces ont été positionnés pour que les relevés soit disposés au plus près d'un modèle A + B, c'est à dire région + végétation, centré par région ce qu'on voit par :



On peut donc dire que le module prépare des réglages permettant de contraindre l'analyse de multiples façons. Pour les sous-espaces voir ¹. Pour le problème biologique, voir ². Pour l'utilisation des sous-espaces voir ³.



¹ Benzecri, J.P. & Coll. (1973) *L'analyse des données. II L'analyse des correspondances*. Bordas, Paris. 1-620, (p. 179-181).

Cailliez, F. & Pages, J.P. (1976) *Introduction à l'analyse des données*. SMASH, 9 rue Duban, 75016 Paris. 1-616, (p. 351-392).

Pontier, J., Jolicœur, P. & Pernin, M.O. (1987) Analyse canonique complète. *Statistique et Analyse des données* : 12, 1 et 2, 124-148.

² Blondel, J., Chessel, D. & Frochot, B. (1988) Niche expansion and density compensation of island birds in mediterranean habitats. A case study from comparison of two ecological successions. *Ecology* : 69, 6, 1899-1917.

Blondel, J. & Farre, H. (1988) The convergent trajectories of bird communities along ecological successions in european forests. *Œcologia (Berlin)* : 75, 83-93.

³ Sabatier, R., Lebreton, J.D. & Chessel, D. (1989) Principal component analysis with instrumental variables as a tool for modelling composition data. In : *Multivariate data analysis*. Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland. 341-352.

Sabatier, R. (1987) *Méthodes factorielles en analyse des données : approximations et prise en compte de variables concomitantes*. Thèse de doctorat d'état. Université de Montpellier. 1-224.

Lebreton, J.D., Sabatier, R., Banco, G. & Bacou, A.M. (1991) Principal component and correspondence analyses with respect to instrumental variables : an overview of their role in studies of structure-activity and species- environment relationships. In : *Applied Multivariate Analysis in SAR and Environmental Studies*. Devillers, J. & Karcher, W. (Eds.) Kluwer Academic Publishers. 85-114.

Résumé de l'organisation générale

Le module prend en entrée des tableaux de données (fichiers binaires quelconques) et des triplets statistiques (fichiers du type ---.##ta, --- étant un nom générique, ## les initiales de l'analyse et ta pour tableau) créés par les modules de la première couche (PCA, COA, HTA, MCA).

Quatre options créent des bases orthonormées de sous-espaces. On peut considérer une telle base orthonormée comme un ensemble de variables dérivée des variables d'origine ayant de bonnes propriétés pour la régression : lorsqu'elles sont centrées (cela dépend des cas) elles sont de variance 1 et de covariance nulle deux à deux. Les variables qui génèrent de telles bases sont dites explicatives, soit qu'on cherche à en tenir compte (explicatives positives), soit qu'on cherche à se débarrasser de leurs effets (explicatives négatives).

Projectors : Table->Orthonormal Basis crée une base à partir d'un tableau et d'une pondération externe. Si ce tableau contient des variables non centrées la base permet de faire des régression par l'origine (modèles sans terme constant). Si ce tableau contient des variables centrées, on retrouve la régression multiple classique comme projection sur ce sous-espace.

Projectors : Triplet->Orthonormal Basis crée une base à partir d'un tableau et d'une pondération associés dans un même triplet. Après une ACP normée, on retrouve la régression multiple classique comme projection sur ce sous-espace.

Projectors : One Categ Var->Orthonormal Basis crée la base associée à une variable qualitative. La projection sur ce sous-espace correspond au moyennage par classe et renvoie aux analyses inter et intra-classes.

Projectors : Two Categ Var->Orthonormal Bases crée les sous-espaces associés à deux variables qualitatives. La projection sur ces sous-espaces génèrent les modèles de l'analyse de variance multivariée (effet AxB pour les plans avec répétitions, effet A•B dit d'interaction dans le même cas, effet additif A+B, effet d'un facteur A et B, effets partiels A/B et B/A).

Deux options créent des sous-espaces à partir de sous-espaces créés par les options précédentes.

Projectors : Combine 2 Orthonormal Bases étend le calcul des effets AxB, A+B, A/B et B/A à des variables quelconques. On peut manipuler un paquet d'explicatives de variables quantitatives et un paquet d'explicatives qualitatives (analyse de covariance multiple).

Projectors : Intersection of 2 Subspaces calcule une base de l'intersection de deux sous-espaces et peut servir à l'usage du sous-espace A(B des effets A et Non B des contraintes fortes).

Cinq options utilisent ces bases pour étudier un autre tableau, dans la stratégie générale des variables instrumentales. Les trois premières sont des utilitaires :

Projectors : Table Projection fait la projection simple d'un tableau sur une base. Les utilisateurs peu habitués à ces pratiques statistiques du champ professionnel pourront vérifier qu'en projetant sur un espace A d'une variable qualitative on fait des moyennes par classe, qu'en projetant sur un paquet de quantitative on fait de la régression multiple ordinaire.

Projectors : Triplet Inertia Decomposition décompose l'inertie d'un triplet par rapport à une base orthonormale et édite les résultats par colonne. On peut décomposer progressivement de l'inertie pour voir le rôle des espaces emboîtés. Le triplet projeté (variables à expliquer) et la base de projection (variables explicatives) doivent être définis pour la même pondération.

Projectors : Subspace Test teste par randomisation l'effet des explicatives sur les expliquées. Utiliser de préférence des associations utilisant des pondérations uniformes, l'usage de pondération non uniforme étant soumise à des règles strictes.

Les deux dernières options sont des programmes très généraux.

Projectors : PCA on Instrumental Variables est un programme très général d'analyse sur variables instrumentales positives qui permet les pratiques d'ACPVI, d'analyse des redondances, de regression simultanée et, pour l'écologie, est un module CANOCO complet.

Projectors : Orthogonal PCAIV est un programme très général d'analyse sur variables instrumentales négatives qui permet les pratiques d'ACPVI orthogonale, d'analyse des covariances partielles et d'ordination avec élimination d'effets.

Ces modules de calcul génèrent des cartes factorielles sous contraintes qui sont autant de point de vue imposé pour explorer les données. La diversité des résultats potentiels est considérable et un minimum de connaissances théoriques sont requises pour un emploi efficace.

