

Ordination sous contraintes

Résumé

L'Analyse en Composantes Explicatives (ACE) d'Obadia (1978 - *Revue de Statistique Appliquée* : 24, 5-28) est la version ACP/ACP d'une méthode très générale de couplages de tableaux qu'on peut mettre en œuvre sur tout type de variables dans le module Projectors. La version AFC/ACP est l'Analyse Canonique des Correspondances ou AFCVI ou encore AFC sous contraintes. Diverses illustrations introduisent à la plasticité de ces méthodes et en soulignent les limites. Le module Projectors permet ainsi des analyses sous contraintes variées, en particulier les analyses canoniques des correspondances partielles, les analyses des covariances partielles, LONGI. Toutes les variantes sont exécutables sur variables qualitatives, quantitatives et floues.

Plan

1 — L'exemple de Lebreton & Coll. (1991)	2
1.1 — Données traitées (Carte Banyuls).....	2
1.2 — L'ACP de X2	4
1.3 — L'AFC de X1.....	5
1.3 — L'AFCVI de X1 par rapport à X2	7
2 — Éliminations d'effets dans un plan d'observation.....	11
2.1 — Les données de J. Blondel & B. Frochot.....	11
2.2 — Élimination d'effet et analyse intra-classes	15
2.3 — Sous-espaces et plan d'observations	17
2.4 — L'effet conjoint végétation-région AxB	19
2.5 — L'effet additif A+B.....	21
2.6 — L'effet partiel B sachant A.....	23
2.7 — Contraintes inter-intra	24
3 — L'ACP sur variables instrumentales.....	27
3-1 — Définition du sous-espace de projection.....	29
3-2 — Test de permutation.....	30
3-3 — L'option @PCAVI.....	30
3-4 — Bilan.....	39
Références	39

D. Chessel

D. Chessel

1 — L'exemple de Lebreton & Coll. (1991)

L'Analyse Canonique des Correspondances¹ (ACC) ou Analyse Factorielle des Correspondances sur Variables Instrumentales² (AFCVI) est la version ACP/AFC d'une méthode de couplage de deux tableaux dont la version ACP/ACP est l'ACP sur variables Instrumentales³ (ACPVI) ou encore Analyse en Composantes Explicatives⁴ (ACE) ou encore Analyse des Redondances⁵ (AR).

Le lien, important au niveau théorique, est clairement explicité dans l'article de Lebreton & Coll⁶. Ce texte présente, en outre, un exemple pédagogique remarquable proposé aux auteurs par R. Prodon, qui nous permettra de comparer les résultats obtenus avec ADE-4 et Biomeco 3.9⁷. On se référera à ce travail par le symbole L&C91. L'utilisation d'une copie de cet article est vivement recommandée.

1.1 — Données traitées (Carte Banyuls)

Data	067	Banyuls (12-22)+(12-4)
10110000000000000000		1483, 23, 3, 30
11111000000000000000		777, 29, 2, 20
0011111111110000000000		298, 35, 11, 25
0001110000010000000000		875, 57, 1, 25
0111100111100000000000		509, 57, 11, 20
0011100110001000000000		542, 60, 4, 20
0011110110101000000000		382, 62, 12, 20
0111100111101100000000		272, 66, 6, 27
0110100111101111100000		214, 80, 10, 18
0010010111001111011000		105, 85, 12, 10
0011000001100010001100		148, 100, 17, 5
0001000111101011001011		62, 123, 16, 15
Gradient of vegetation on rocky slopes facing the sea. data from Prodon R. pers. com. to Lebreton J.D. & coll 1991. 12 samples-22 plant species 12 samples-4 environmental variables 1-salt- Salinity (g/m ²) 2-dist- Dist. to sea (m) 3-soil- Soil depth (cm) 4-slope- (°)		
		camp poly dauc sonc plan thym crit fest heli crep dact arme rosm onon ulex aspa diam alis brac rubu plpd rubi

Reprenons strictement les notations utilisées dans L&C91. X1 est un tableau de 12 échantillons et 22 espèces végétales. X2 est un tableau de 12 échantillons (les mêmes) et 4 variables environnementales (salinité en g/m², distance à la mer en m, profondeur du sol en cm et pente en degrés). Implanter deux fichiers binaires X1 (12-22) et X2 (12-4) et vérifier que leur contenu est conforme au tableau 1 de L&C91 page 86.

List BIN file			
Bin file		H1	12 22
Print format (default = %8.4f)		%2.0f	
List BIN file			
Bin file		H2	12 4
Print format (default = %8.4f)		%4.0f	

Input file: X1													Input file: X2									
Row: 12 Col: 22													Row: 12 Col: 4									
1		1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2		1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3		0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
4		0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5		0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
6		0	0	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
7		0	0	1	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
8		0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0
9		0	1	1	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0
10		0	0	1	0	0	1	0	1	1	0	0	1	1	1	0	1	1	0	0	0	0
11		0	0	1	1	0	0	0	0	1	1	0	0	0	1	0	0	0	1	1	0	0
12		0	0	0	1	0	0	0	1	1	1	0	1	0	1	0	0	1	0	1	1	1
1		1483	23	3	30																	
2		777	29	2	20																	
3		298	35	11	25																	
4		875	57	1	25																	
5		509	57	11	20																	
6		542	60	4	20																	
7		382	62	12	20																	
8		272	66	6	27																	
9		214	80	10	18																	
10		105	85	12	10																	
11		148	100	17	5																	
12		62	123	16	15																	

Retrouver la matrice de corrélations des variables de milieu (L&C91, tableau 2, p. 86) avec le module MatAlg :

Diagonal Inner product C=H'DY

Input file for H matrix		H2	12	4
Option for H matrix (default=none)		3		
Input file for Y matrix		H2	12	4
Option for Y matrix (default=none)		3		

```

X input file: X2
--- Number of rows: 12, columns: 4
Y input file: X2
--- Number of rows: 12, columns: 4
Diagonal inner product: uniform weighting
XtDY output file: screen
--- Number of rows: 4, columns: 4
Input file: screen
--- Number of rows: 4, columns: 4
-----
[ 1] 1.0000e+00 -7.5109e-01 -7.6207e-01 6.5870e-01
[ 2] -7.5109e-01 1.0000e+00 7.2945e-01 -7.2721e-01
[ 3] -7.6207e-01 7.2945e-01 1.0000e+00 -7.1022e-01
[ 4] 6.5870e-01 -7.2721e-01 -7.1022e-01 1.0000e+00
-----

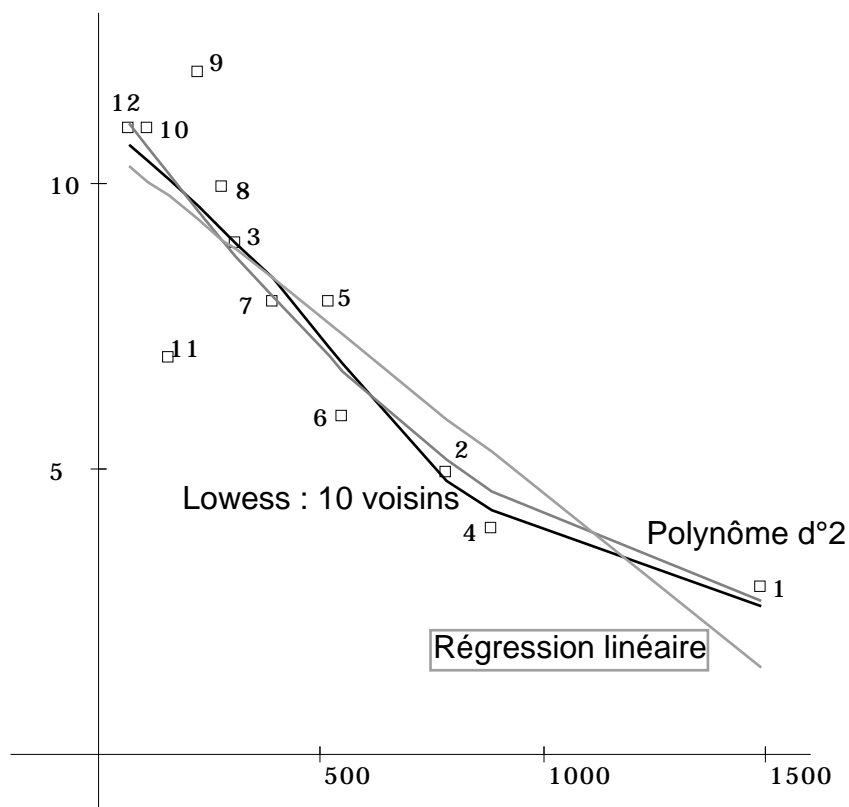
```

On a utilisé une pondération uniforme des relevés qui sera plus tard abandonnée. Toute la question est celle du couplage des tableaux X1 et X2 ou plus précisément la capacité de X2 à prédire le tableau X1. La relation entre la richesse et la salinité proposée par les auteurs (figure 1 p. 87) s'obtient par EcoMargins :

Ecological table margins

Input file (Bin)	H1	12	22
-------------------------	-----------	-----------	-----------

La richesse est la première colonne de X1.ecoli. La salinité est la première colonne de X2. Utiliser Curves et CurveModels :



La régression polynomiale suggère fortement un changement de variable en racine (x), mais on conservera les variables pour retrouver les résultats de l'article cité. En fait, cette première opération consistant à séparer un effet richesse, qui va disparaître dans l'ACC du couple de tableaux, est un choix fondamental. L'AFCVI ignorant par essence les marges du tableau floristique, l'utiliser implique d'examiner par ailleurs ces marges. L'ACPVI n'éliminant pas, dans sa version classique (centrage par colonnes), la marge relevés va au contraire permettre de discuter cet élément dans l'analyse elle-même. Nous comparerons les deux optiques.

1.2 — L'ACP de X2

La discussion du modèle d'équicorrélation des variables de milieu (L&C91 § 2.2) est très instructive. Pour retrouver le tableau 4 qui donne les poids des variables dans la définition du facteur 1 de l'ACP de X2 utiliser PCA :

```

Correlation matrix PCA
Matrix input file H2 12 4

--- Correlation matrix ---
[ 1] 1000
[ 2] -751 1000
[ 3] -762 729 1000
[ 4] 659 -727 -710 1000

```

puis DDUtil :

```

Add normed scores
Input file H2.cnup 4 2

```

X2.cnc1 créé contient les coefficients des variables dans la constitution de la première coordonnée :

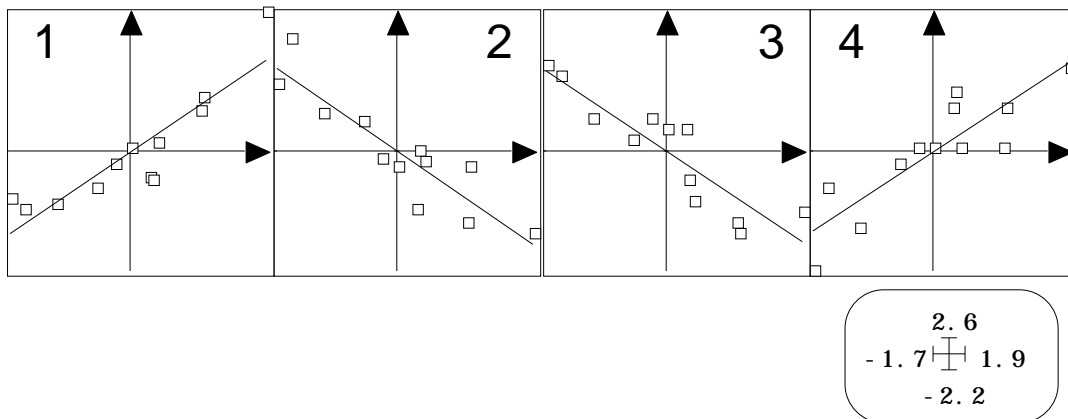
```

Input file: X2.cnc1
Row: 4 Col: 1
 1 | 0.5007 |
 2 | -0.5065 |
 3 | -0.5055 |
 4 | 0.4871 |

```

On retrouve l'équation $y = .50 \text{ sal} + .51 (-\text{dist}) + .51 (-\text{soil}) + .49 \text{ slope}$. La figure rendant compte de cette situation est le graphe canonique dessiné avec la coordonnée de variance 1 dans CurveModels :

Polynomials			
X file (default = 1, 2, 3, ..., n)		H2.cnc1	12 1
X file column number (default = 1)			
Y file (no default)		H2.cnta	12 4
Order of polynomial (default = 1) ?		1	



Chacune des variables est équivalente aux trois autres et, quitte à n'en utiliser qu'une, le meilleur est de prendre la coordonnée factorielle de rang 1 de l'ACP normée de X2.

1.3 — L'AFC de X1

L'article dont nous reproduisons les résultats est le premier, à notre connaissance, à montrer le rôle du centrage implicite de l'analyse des correspondances. La tradition⁸ a longtemps voulu que l'AFC traite un tableau non transformé qui donne une valeur propre parasite (égale à 1) qu'on élimine volontairement. Escoufier⁹ a rendu explicite le centrage caché dans l'élimination de la valeur propre 1 et le module d'ADE utilise explicitement l'algorithme décrit dans L&C91 tableau 5, p. 90.

Si on exécute l'AFC de X1 :

COrrespondence Analysis			
Data file		H1	12 22

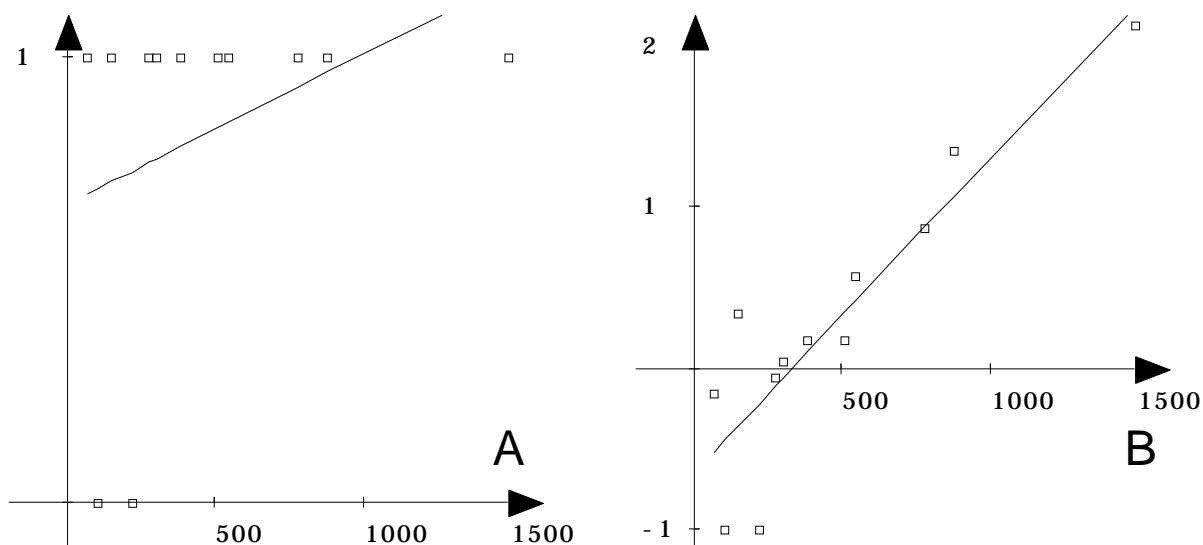
File X1.fcta contains the doubly centred table $DI-1 * P * DJ-1 - II * 1J'$
It has 12 rows and 22 columns

On peut examiner la relation entre salinité et abondance de l'espèce 4 (code Sonc). En utilisant les données brutes (0-1), on obtient la partie A de la figure ci-dessous :

Polynomials			
X file (default = 1, 2, 3, ..., n)	<input type="text" value="H2"/>	12	4
X file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="H1"/>	12	22
Order of polynomial (default = 1) ?	<input type="text" value="1"/>		

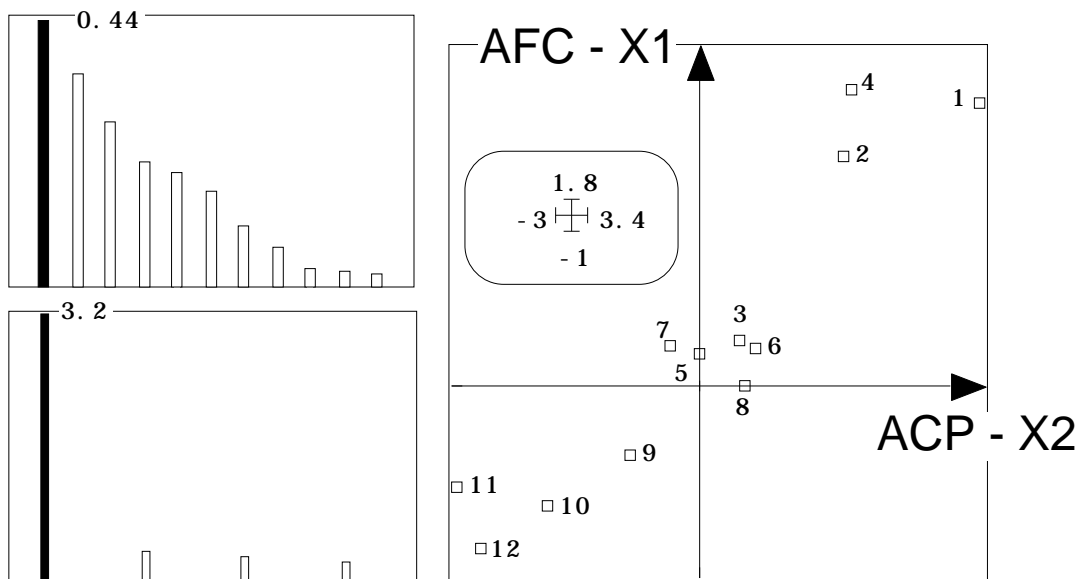
En utilisant les données transformées, on obtient la partie B :

Polynomials			
X file (default = 1, 2, 3, ..., n)	<input type="text" value="H2"/>	12	4
X file column number (default = 1)	<input type="text"/>		
Y file (no default)	<input type="text" value="H1.fcta"/>	12	22
Order of polynomial (default = 1) ?	<input type="text" value="1"/>		



La linéarisation de la relation est manifeste — L&C91, figure 2, p. 92 — lorsqu'on substitue aux données x_{ij} (à gauche) les données transformées $y_{ij} = x_{ij} / x_{i \cdot} x_{\cdot j} - 1$ (à droite). La présence d'une espèce prend une valeur qui dépend de toutes les autres : ici, la présence de l'espèce prend une valeur croissante avec la richesse qui décroît quand la salinité croît. Sa tolérance à la salinité est mise en évidence, par comparaison avec la tolérance des autres taxa. L'usage des valeurs corrigées qui figurent dans le tableau effectivement analysée par l'AFC est rarissime, alors qu'il serait utile lors de la relecture des données pour comprendre la position de certains objets.

On en arrive naturellement à comparer le facteur 1 de l'AFC et le facteur 1 de l'ACP (L&C91 figure 3, p. 93) :



Il est logique de ne pas tracer une droite de régression sur ce graphique, car la variable en abscisse est de moyenne nulle pour la pondération uniforme et la variable en ordonnée est de moyenne nulle pour la pondération marginale du tableau floristique. Pour associer directement les deux analyses, il convient de choisir une pondération commune.

Les fiches de ce fascicule ont souvent rappeler que faire une régression, c'est faire une projection. Pour projeter, il faut un produit scalaire et chacune des pondérations en définit un. Il faut choisir, soit importer l'une des deux pondérations dans l'analyse de l'autre, soit chercher un compromis. L'AFCVI choisit de transporter la pondération de l'AFC dans l'ACP normée.

1.3 — L'AFCVI de X1 par rapport à X2

Exécuter l'analyse conformément à la fiche Analyse Canonique des Correspondances. L'AFC de X2 est déjà faite. On a gardé trois axes. Refaire l'ACP de X2 en important la pondération de l'AFC de X1 :

Correlation matrix PCA			
Matrix input file	<input type="button" value="H2"/>	<input type="text" value="H2"/>	12 4
Row weights (default=1/n)	<input type="button" value="3"/>	<input type="text" value="3"/>	
Column weights (default=1)	<input type="button" value=""/>	<input type="text" value=""/>	
Option: file for row weighting	<input type="button" value="H1.fcpl"/>	<input type="text" value="H1.fcpl"/>	12 1
Option: file for column weighting	<input type="button" value=""/>	<input type="text" value=""/>	
1 = Save correlation matrix	<input type="button" value="1"/>	<input type="text" value="1"/>	

On garde deux axes pour des raisons pédagogiques, bien qu'un seul contienne toute l'information. Ce transport de pondération est certainement un des éléments les plus discutables de l'ACC (ou AFCVI). Il impose de calculer des moyennes, des variances et des corrélations entre variables de milieu qui dépendent du cortège floristique étudié. Pour un tableau faunistique, qui pourrait être divisé en groupes taxonomiques, on ne peut soutenir que la moyenne d'une variable de milieu devrait dépendre du groupe

auquel on s'intéresse. On peut prendre la chose comme une contrainte de fonctionnement, sans en être satisfait. D'autres choix sont possibles.

Définir alors la base orthonormée du sous-espace engendré par les colonnes de X2 dans le module Projectors :

Triplet->Orthonormal Basis			
Explanatory variables		X2.cnta	12 4
Option: output file name			

Orthonormalization: subspace generated by a statistical triplet

 Explanatory variable file: X2.cnta
 It has 12 rows and 4 columns

 Orthonormal basis: X2.@ob. It has 12 rows and 4 columns
 Row weighting file: X2.@pl (the same as X2.cnpl)
 Coordinates of the vectors of the orthonormal basis
 in the basis of columns of X2.cnta in : X2.@co
 File X2.@co has 4 rows and 4 columns

Etudier la décomposition de l'inertie de chaque colonne de X1 (espèces) dans sa régression multiple sur les 4 variables de milieu :




Triplet Inertia Decomposition			
Explanatory variables: .@ob file		X2.@ob	12 4
Dependant variables: .**ta		X1.fcta	12 22

Poids		Subspace A	A Orthogo	Total	A+	A-
0.021277	1	1.5177e-01	9.3622e-02	2.4539e-01	6184	3815
0.042553	2	1.7937e-02	6.6594e-02	8.4530e-02	2121	7878
0.10638	3	2.6990e-02	1.4448e-02	4.1438e-02	6513	3486
0.10638	4	4.2476e-02	1.5629e-02	5.8105e-02	7310	2689
0.085106	5	3.6084e-02	2.3948e-02	6.0033e-02	6010	3989
0.042553	6	8.2193e-03	9.3483e-02	1.0170e-01	808	9191
0.010638	7	6.2980e-02	3.7493e-02	1.0047e-01	6268	3731
0.085106	8	9.1803e-03	1.7330e-02	2.6510e-02	3462	6537
0.085106	9	9.1803e-03	1.7330e-02	2.6510e-02	3462	6537
0.074468	10	1.5500e-02	1.6334e-02	3.1835e-02	4868	5131
0.074468	11	2.3330e-02	1.3375e-02	3.6705e-02	6356	3643
0.010638	12	6.7329e-02	1.7203e-01	2.3936e-01	2812	7187
0.06383	13	1.7432e-02	2.8208e-02	4.5640e-02	3819	6180
0.031915	14	2.7986e-02	3.1513e-02	5.9499e-02	4703	5296
0.042553	15	4.9693e-02	9.7557e-03	5.9449e-02	8358	1641
0.031915	16	2.8495e-02	2.7974e-02	5.6469e-02	5046	4953
0.010638	17	5.3174e-03	6.7378e-02	7.2695e-02	731	9268
0.010638	18	2.9807e-02	5.0464e-02	8.0271e-02	3713	6286
0.031915	19	6.2081e-02	1.4229e-02	7.6310e-02	8135	1864
0.010638	20	5.6609e-02	7.5610e-02	1.3222e-01	4281	5718
0.010638	21	5.5419e-02	2.4852e-02	8.0271e-02	6904	3095
0.010638	22	5.5419e-02	2.4852e-02	8.0271e-02	6904	3095
	Tot	8.5923e-01	9.3645e-01	1.7957e+00	4784	5215

Ce tableau contient l'essentiel du tableau 6 de L&C91 p. 97. On a rajouté les poids dont les arrondis ont provoqué des approximations sensibles dans le travail cité. L'inertie de l'AFC est bien 1.7957 (comme indiqué dans l'annexe p. 109 par TRACE = 1.79568500 et non 1.821 comme indiqué dans le tableau 6) et l'inertie totale de l'AFCVI est bien 0.85923 (comme indiqué dans l'annexe p. 113 par TRACE = .859 et non 0.869 comme indiqué dans le tableau 6).

Dans le tableau ci-dessus, on notera l'inertie de chaque espèce dans l'AFC (par exemple 0.245 pour l'espèce 1), l'inertie projetée par régression sur le tableau X2 (0.152 pour l'espèce 1) et l'inertie projetée sur l'orthogonal ou inertie résiduelle ou encore erreur de prédiction (0.0936) obtenue par différence. Dans la colonne A+, on a donc des carrés de corrélation multiple (0.618 pour l'espèce 1) et les écarts à 1 (0.381) dans la colonne A-. Au total, l'inertie projetée (0.8593) représente 47.8 % de l'inertie d'origine (1.7957). L'importance d'une espèce en AFCVI sera donc un compromis entre son poids, sa distance à l'origine (qui augmentent son inertie initiale) et sa prédictibilité par les variables de milieu (qui en fait autant).

On obtient l'AFCVI par :

PCA on Instrumental Variables			
Explanatory variables: .@ob file		H2.@ob	12 4
Dependant variables: .**ta		H1.fcta	12 22
Output file name		A	

```

----- input -----
Orthonormal basis: X2.@ob
It has 12 rows and 4 columns
Dependant variable file: X1.fcta
It has 12 rows and 22 columns
----- output -----
File A.ivpc contains the column weights
It has 22 rows and 1 column
File A.ivpl contains the row weights
It has 12 rows and 1 column
Projected variable file: A.ivta
It has 12 rows and 22 columns

```

Le tableau A.ivta contient les positions moyennes des espèces sur les variables normalisées de X2.cnta. Le tableau 7 de L&C91 p. 98 donne les valeurs équivalentes sur les variables brutes de X2.

```

Inertia: 8.5923e-01
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

Les valeurs propres sont celles du listing de BIOMEKO (p. 113) qui ajoute leur racine carrée.

```

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

```

```

----- Minimum/Maximum:
Col.:  1 Mini = -2.3038  Maxi = 1.4764
Col.:  2 Mini = -2.0345  Maxi = 1.2913

```

Lister le fichier A.ivco : il est donné par BIOMEKO 3.9 (L&C91 p. 114). On notera encore une fois qu'un changement de signe de toutes les valeurs n'a aucune signification et dépend du hasard.

```

Input file: A.ivco
Row: 22 Col: 2
 1 | -2.3038 | 1.2594 |
 2 | -0.4523 | -0.2689 |
 3 | -0.4195 | 0.0499 |
 4 | -0.5291 | 0.2788 |
 5 | -0.5617 | -0.2237 |
 6 | -0.3095 | -0.2375 |
 7 | -0.3833 | -2.0345 |
 8 | 0.1038 | -0.2915 |
 9 | 0.1038 | -0.2915 |
10 | 0.3879 | -0.1663 |
11 | 0.2690 | -0.2262 |
12 | -1.4748 | 1.2913 |
13 | 0.2754 | -0.0591 |
14 | 0.3339 | -0.4230 |
15 | 0.9083 | 0.4925 |
16 | 0.8581 | 0.2575 |
17 | 0.3718 | -0.2189 |
18 | 0.7260 | 0.1064 |
19 | 1.0872 | 0.7296 |
20 | 1.0591 | 1.1974 |
21 | 1.4764 | 0.8850 |
22 | 1.4764 | 0.8850 |

```

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

```

----- Minimum/Maximum:
Col.:  1 Mini = -2.9614  Maxi = 1.4764
Col.:  2 Mini = -2.0345  Maxi = 2.2762

```

Lister le fichier A.ivl1 : il est donné par BIOMEKO 3.9 (L&C91 p. 114).

```

Input file: A.ivl1
Row: 12 Col: 2
 1 | -2.9614 | 2.2762 |
 2 | -1.6462 | 0.2427 |
 3 | -0.3833 | -2.0345 |
 4 | -1.4748 | 1.2913 |
 5 | -0.4387 | 0.0570 |
 6 | -0.7201 | 0.3428 |
 7 | -0.1060 | -0.3133 |
 8 | -0.0960 | -1.1564 |
 9 | 0.3718 | -0.2189 |
10 | 0.7260 | 0.1064 |
11 | 1.0591 | 1.1974 |
12 | 1.4764 | 0.8850 |

```

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

```

----- Minimum/Maximum:
Col.:  1 Mini = -0.53105  Maxi = 0.3828
Col.:  2 Mini = -0.5041  Maxi = 1.2896

```

Le fichier A.ivfa contient les poids canoniques ou coefficient des variables dans la constitution des coordonnées de variance 1. Lister ce fichier et retrouver l'équation :

```

Input file: A.ivfa
Row: 4 Col: 2
 1 | -0.5310 | 1.2896 |
 2 |  0.3828 | 0.9871 |
 3 |  0.1746 | -0.0234 |
 4 | -0.0126 | -0.5041 |

```

$y = 0.531 \text{ SAL} - 0.383 \text{ DIST} - 0.175 \text{ SOIL} + 0.0126 \text{ SLOPE}$ (L&C91 p. 100). L'interprétation de l'AFCVI par BIOMECO 3.9 privilégie donc la démarche combinaisons de variables (A.ivfa) de milieu de variance unité (A.ivl1) qui maximise la variance des moyennes conditionnelles par espèce (A.ivco). On trouvera quelques remarques complémentaires dans la fiche Analyse Canonique des Correspondances.

2 — Éliminations d'effets dans un plan d'observation

On explicite ici les calculs relativement complexes qui sont en jeu dans une note¹⁰ donnant une solution à un problème posé par J. Blondel¹¹. Considérons un tableau faunistique T comportant n relevés et t taxons et supposons que les n relevés soient référencés à un plan d'observations à deux facteurs contrôlés. Le relevé i est caractérisé par la j^{ème} ligne du tableau T (profil faunistique) et les modalités qu'il prend pour deux variables qualitatives, la première notée A prenant m_A modalités, la seconde notée B prenant m_B modalités.

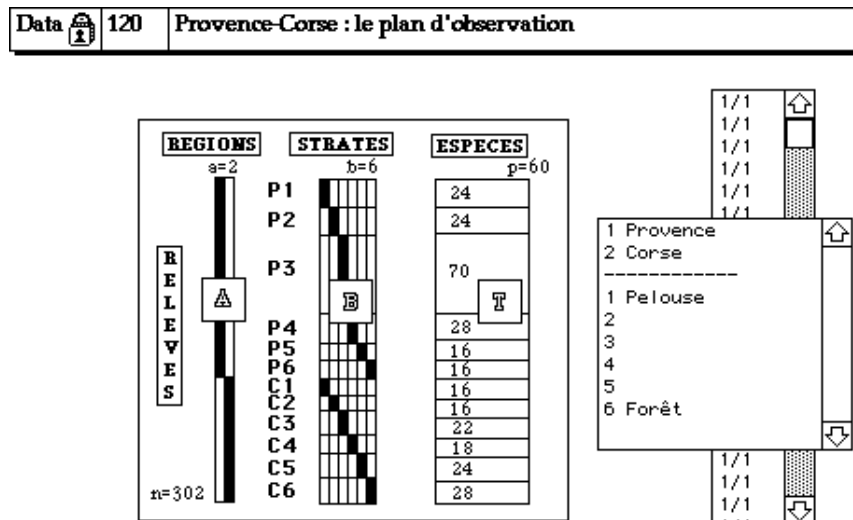
On suppose ici que *l'objectif de l'analyse est la description de la structure du tableau T induite par B et ne tenant pas compte de l'effet de A*. On peut utiliser trois Analyses Factorielles des Correspondances (AFC) sous contraintes croissantes de prise en compte et d'élimination d'effets.

2.1 — Les données de J. Blondel & B. Frochot

L'exemple traité est formé de n=302 relevés d'avifaune appartenant à deux régions différentes (A) et six stades d'une succession végétale (B). A prend les valeurs 1 (Provence) et 2 (Corse). B prend six valeurs de 1 (Mattoral bas haut de 0.5 m) à 6 (vieille futaie haute de 25 m). L'expérience est décrite dans l'ouvrage de J. Blondel¹². L'AFC simple montre la convergence des cortèges faunistiques en milieu fermé. Les différences importantes en milieux ouverts du cortège faunistique demande une élimination des effets régionaux afin de pouvoir comparer les structures insulaire et continentale. Le tableau faunistique est dans la carte Provence-Corse :

The screenshot shows a software window titled "Data 119 Provence-Corse 302-60 J.Blondel,B. Frochot Utilisation dans". The main area contains a grid of binary data (0s and 1s) and descriptive text: "Données de 302 relevés d'avifaune", "2 régions", "6 strates de végétation", "Descriptif complet dans l'article cité. Voir aussi", and "Plan d'observation sur la carte suivante (flèche à droite au clavier)". Below this is a "Char->Binary" window with two input fields: "Text input file" with a button labeled "Avi.Car" and "Binary output file" with a button labeled "Avi".

Le fichier Avi a 302 lignes et 60 colonnes. Le plan d'expérience est dans la carte Provence-Corse+1 :



Text->Binary

Text input file

Binary output file

Le fichier Plan a 302 lignes et 2 colonnes. Le lire par CategVar :

Read Categ File

Input file 302 2

Categorical variables: file Plan
 Rows: 302, Variables: 2, Categories: 8, Missing data: 0

Description of categories:

 Variable number 1 has 2 categories

 [1]Category: 1 Num: 178 Freq.: 0.589
 [2]Category: 2 Num: 124 Freq.: 0.411

 Variable number 2 has 6 categories

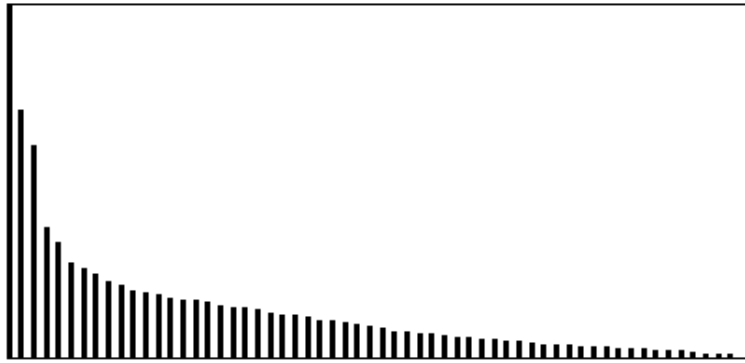
 [3]Category: 1 Num: 40 Freq.: 0.132
 [4]Category: 2 Num: 40 Freq.: 0.132
 [5]Category: 3 Num: 92 Freq.: 0.305
 [6]Category: 4 Num: 46 Freq.: 0.152
 [7]Category: 5 Num: 40 Freq.: 0.132
 [8]Category: 6 Num: 44 Freq.: 0.146

Faire l'AFC de Avi :

CORrespondence Analysis

Data file 302 60

On conserve trois facteurs particulièrement explicites :



Dépouiller les plans 1-2 et 1-3 en utilisant les 12 classes de relevés :

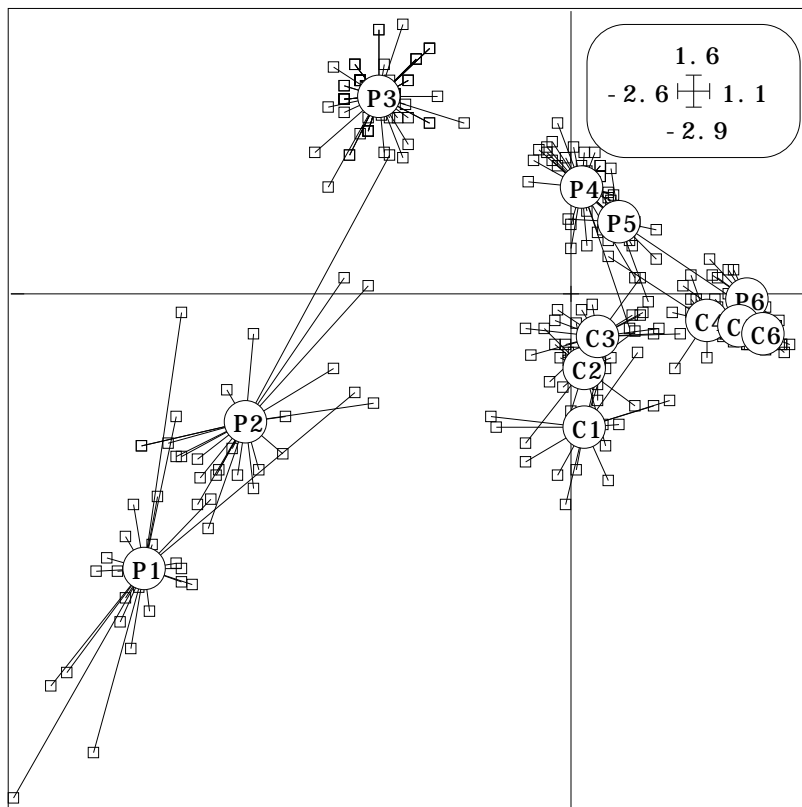
Create1Categ

Output file name

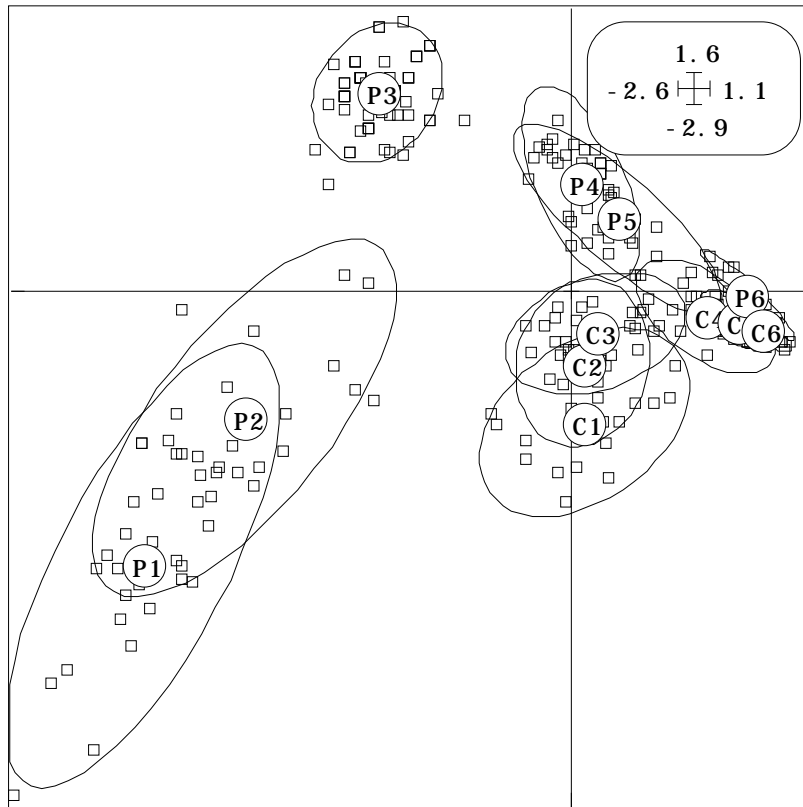
Block number

Enter value: ...

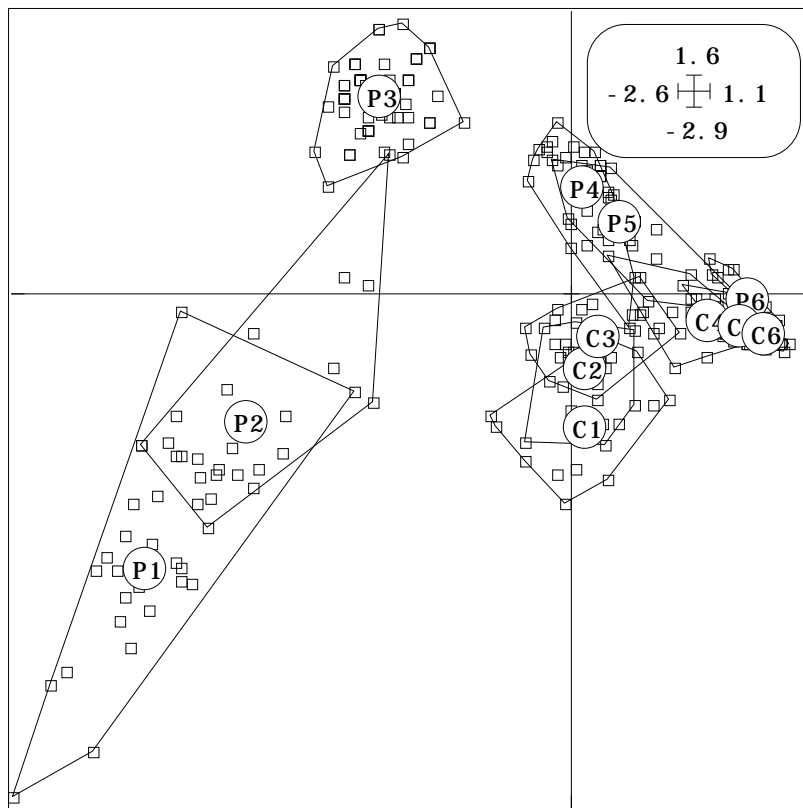
Modifier le fichier P12.123 pour rendre les étiquettes plus explicites (P1 à P6 et C1 à C6). Tracer la carte factorielle en couplant Avi.fcl et P12.cat dans ScatterClass :



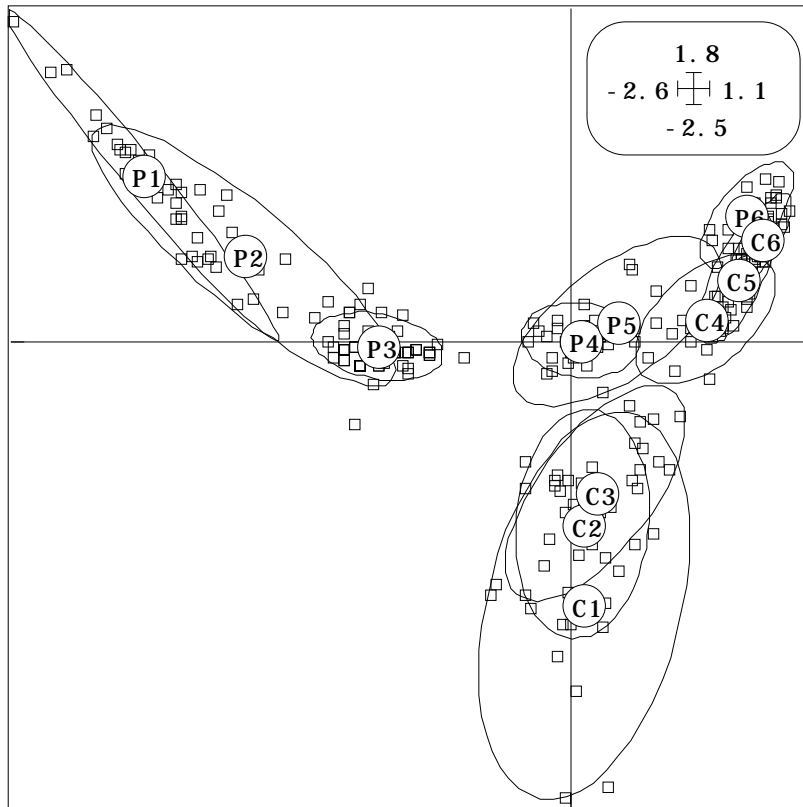
On peut préférer l'option Ellipses :



ou l'option polygones de contour :



Il y a convergence des cortèges faunistiques en milieu fermé, ce qui est confirmé sur le facteur 3 dans le plan 1-3 :



L'interprétation de cet effet de convergence des cortèges faunistiques a été largement explicité par J. Blondel (op. cit.) et ces données posent des questions de statistiques multidimensionnelles caractéristiques des problèmes d'ordination dans un contexte d'observations finalisées par des plans d'observations complexes. Si on rajoute des régions, l'observation de la convergence devient plus difficile du fait de l'hétérogénéité des cortèges régionaux. Nous avons donné une solution à cette question dans la fiche Analyse triadique partielle du fascicule 6, suite à l'article de Blondel et Farré¹³. Nous voulons ici comparer la structure végétation-avifaune en éliminant la différence inter-régionale au moyen des AFC sous contraintes utilisant des éliminations d'effets.

2.2 — Élimination d'effet et analyse intra-classes

Sans s'occuper de l'effet B (stades de végétation) il est possible d'éliminer l'effet A (région) par l'AFC intra-classes de Benzécri¹⁴ ou Escofier et Drouet¹⁵. L'analyse s'écrit :

$$\text{ACP} \left[Q_A (\mathbf{D}_n^{-1} \mathbf{P} \mathbf{D}_t^{-1} - \mathbf{1}_n \mathbf{1}_t'), \mathbf{D}_n, \mathbf{D}_t \right]$$

où ACP désigne l'analyse en composantes principales du triplet formé d'un tableau et de deux pondérations¹⁶, \mathbf{P} est la distribution bivariée de fréquence associée à T, \mathbf{D}_n la matrice diagonale des poids des lignes, \mathbf{D}_t la matrice diagonale des poids des colonnes, $\mathbf{1}_n$ et $\mathbf{1}_t$ des vecteurs de composantes toutes égales à 1. Q_A est alors le projecteur \mathbf{D}_n -orthogonal sur l'orthogonal du sous-espace vectoriel engendré par les indicatrices de A^{17 18}.

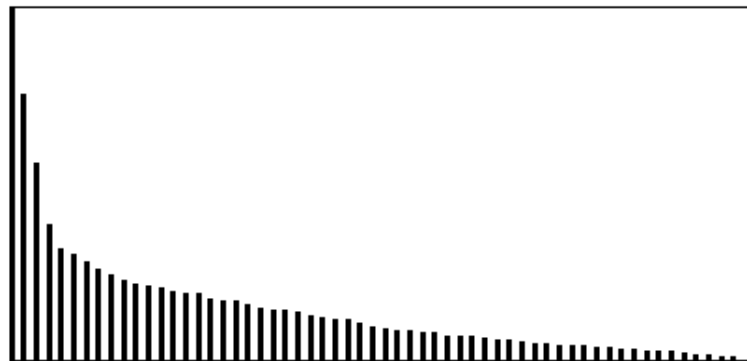
On peut faire cette analyse de plusieurs manières. La première est une option du module COA. Préparer un fichier binaire à une colonne et 2 lignes contenant le nombres de relevés par régions, soit respectivement 178 et 124 :


```

Input file: BlocR
Row: 2 Col: 1
 1 | 178.0000 |
 2 | 124.0000 |

```

Internal COA			
Input file (---)	<input type="button" value="👉"/>	<input type="text" value="Avi"/>	302 60
Row indicator (Default = 1 variable)	<input type="button" value="👉"/>	<input type="text" value="BlocR"/>	2 1
Col indicator (Default = 1 variable)	<input type="button" value="👉"/>	<input type="text"/>	
Output file (default = ---)	<input type="button" value="👉"/>	<input type="text"/>	



On garde encore 3 facteurs et les nouvelles coordonnées des relevés sont dans le fichier Avi.wvli. La seconde est une option du module Discrimin :

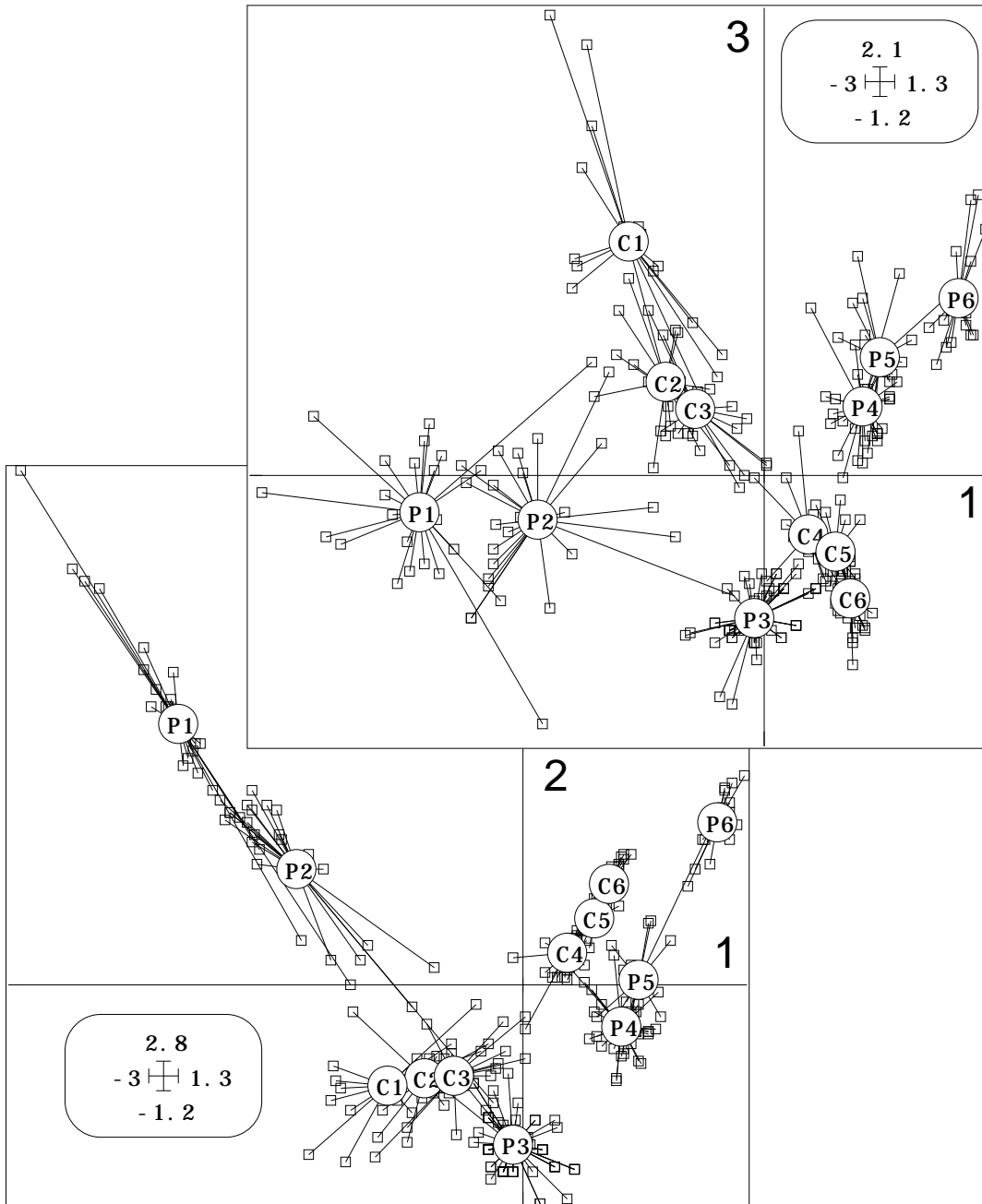
Initialize: LinkPrep			
Statistical triplet	<input type="button" value="👉"/>	<input type="text" value="Avi.fcta"/>	302 60
Categories file (.cat)	<input type="button" value="👉"/>	<input type="text" value="Plan.cat"/>	
Selected variable (default=1)	<input type="button" value="👉"/>	<input type="text" value="1"/>	
Output file name	<input type="button" value="👉"/>	<input type="text" value="Intra"/>	

Within Analysis: Run	
--.dis input file	<input type="button" value="👉"/> <input type="text" value="Intra.dis"/>

Les nouvelles coordonnées sont dans Intra.wvli. Le contenu des deux fichiers Avi.wvli et Intra.wvli sont identiques. Utiliser l'un ou l'autre pour obtenir la figure :

<pre> Input file: Avi.wvli Row: 302 Col: 3 1 -1.8824 1.1404 -0.4401 2 -1.6444 0.9030 0.0320 3 -2.9141 2.7812 -0.0777 </pre>	<pre> Input file: Intra.wvli Row: 302 Col: 3 1 -1.8824 1.1404 -0.4401 2 -1.6444 0.9030 0.0320 3 -2.9141 2.7812 -0.0777 </pre>
--	--






La figure montre le résultat de cette nouvelle ordination où les nuages sont centrés par région mais la non prise en compte de l'effet B y est marquée: les deux nuages sont imbriqués et le recentrage commun ne suffit pas à identifier les gradients d'ouverture du milieu. Ter Braak¹⁹ appelle "partial correspondence analysis" cette procédure. Cazes et coll.²⁰ l'appelle analyse des correspondances internes lorsqu'il y a élimination d'effets sur les deux marges. Le terme d'AFC intra-classes du à Benzécri (op. cit.) semble le plus clair.



La plus simple des analyses à éliminations d'effets, l'analyse intra-classes ne suffit pas à superposer les structures internes des deux ensembles régionaux. On veut intégrer le facteur B en éliminant le facteur A. La troisième façon d'exécuter l'analyse intra est un cas particulier des pratiques suivantes.

2.3 — Sous-espaces et plan d'observations

Les contraintes associées à un plan d'observation se manipulent à l'aide de sous-espaces dans la module Projectors:

Two Categ Var->Orthonormal Bases		
---.cat type file		Plan.cat
Variable A (default=1)		
Variable B (default=2)		
Option: row weighting		Avi.fcpl 302 1
Output file name		W

Subspaces from two categorical variables

```

-----
Input file: Plan
It has 302 rows and 2 columns
Generic output file name: W
Crossing variable A (n° 1) and B (n° 2)
-----
file W_AxB.@ob contains an orthonormal basis of subspace AxB
It has 302 rows and 11 columns
file W_A+B.@ob contains an orthonormal basis of subspace A+B
It has 302 rows and 6 columns
file W_A•B.@ob contains an orthonormal basis of subspace A•B
It has 302 rows and 5 columns
file W_A.@ob contains an orthonormal basis of subspace A
It has 302 rows and 1 columns
file W_B.@ob contains an orthonormal basis of subspace B
It has 302 rows and 5 columns
file W_A/B.@ob contains an orthonormal basis of subspace A/B
It has 302 rows and 1 columns
file W_B/A.@ob contains an orthonormal basis of subspace B/A
It has 302 rows and 5 columns

```

Chacun des fichiers créés contient une base orthonormée d'un sous-espace utilisés en analyse de variance. Les suffixes *.@ob* (*orthonormal basis*) désignent des sous-espaces les suffixes *_####* précisent la nature de ces sous-espaces. Sans entrer dans les détails, on considère :

- les effets simples *_A* (effet A = effet région, dimension 1, 2 classes) et *_B* (effet B = effet végétation, dimension 5, 6 classes),
- les effets combinés *_A+B* (effet additif = effet région + effet végétation, dimension 6 = 1 + 5) et *_AxB* (effet conjoint = effet régionxvégétation = effet de la partition en 12 classes région-végétation, dimension 11, 12 classes),
- les effets partiels *_A/B* (effet A sachant B = effet région dans l'effet région+végétation, dimension 1, 2 classes) et *_B/A* (effet B sachant A = effet végétation dans l'effet région+végétation, dimension 5, 6 classes),
- l'effet d'interaction *_A•B* (effet non additif dans l'effet conjoint, dimension 6 = 11 - 5).

Quand on manipule une seule variable expliquée, on se réfère à ces sous-espaces en analyse de variance pour décomposer la variance en parties additives et on utilise des tests F pour la signification de ces différentes parties. Quand on manipule un tableau, on additionne les variances décomposées des différentes variables pour décomposer l'inertie. Noter l'usage impératif de la pondération de l'AFC pour définir les bases orthonormales (orthonormales seulement pour le produit scalaire associé à la diagonale des poids de l'AFC).

L'utilisation de ces sous-espaces et de l'analyse de départ conduit à des pratiques très riches qui suppose un minimum de maîtrise statistique. L'usage de deux effets ne soulève cependant pas de difficultés majeurs. Une base orthonormale peut simplement

être comprise comme un paquet de variables explicatives orthogonales (non corrélées) qui décrit l'effet de tout ou partie d'une explicative qualitative.

Les sous-espace, en outre se recompose entre eux, pour former des ensembles cohérents qui permettent une exploration organisée du tableau de données.

2.4 — L'effet conjoint végétation-région AxB

```
Input file: Avi.fcta
--- Number of rows: 302, columns: 60
-----
Total inertia: 5.73908
-----
```




On se rappelle que l'inertie totale de l'AFC vaut 5.739. Elle se décompose dans l'analyse simple en une partie organisée dépouillée sur 3 facteurs (27.9 % de structure définie par les trois valeurs propres .6943, .4878 et .4170) et une partie résiduelle considérée comme inorganisée et sans intérêt (72.1 %).

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		W_AxB.@ob	302 11
Dependant variables: .**ta		Avi.fcta	302 60

affiche le rôle de l'effet conjoint :

---	-----	-----	-----	-----	-----
Tot	2.1110e+00	3.6281e+00	5.7391e+00	3678	6321
---	-----	-----	-----	-----	-----

L'inertie expliquée est exactement l'inertie totale de l'analyse inter-classes associée aux 12 classes région-végétation P1 à P6 et C1 à C6. De l'inertie totale (5.739) 36.8 % (2.111) est expliquée par l'effet conjoint et 63.2 % est résiduelle. L'inertie expliquée est exactement l'inertie totale de l'analyse inter-classes associée aux 12 classes région-végétation P1 à P6 et C1 à C6. L'inertie résiduelle est exactement l'inertie totale de l'analyse intra-classes associée aux 12 classes région-végétation P1 à P6 et C1 à C6. La troisième manière de faire des analyses intra-classes est donc celle de l'ACPVI orthogonale, par exemple :

Orthogonal PCAIV			
Explanatory variables: .@ob file		W_AxB.@ob	302 11
Dependant variables: .**ta		Avi.fcta	302 60
Output file name		Provi	

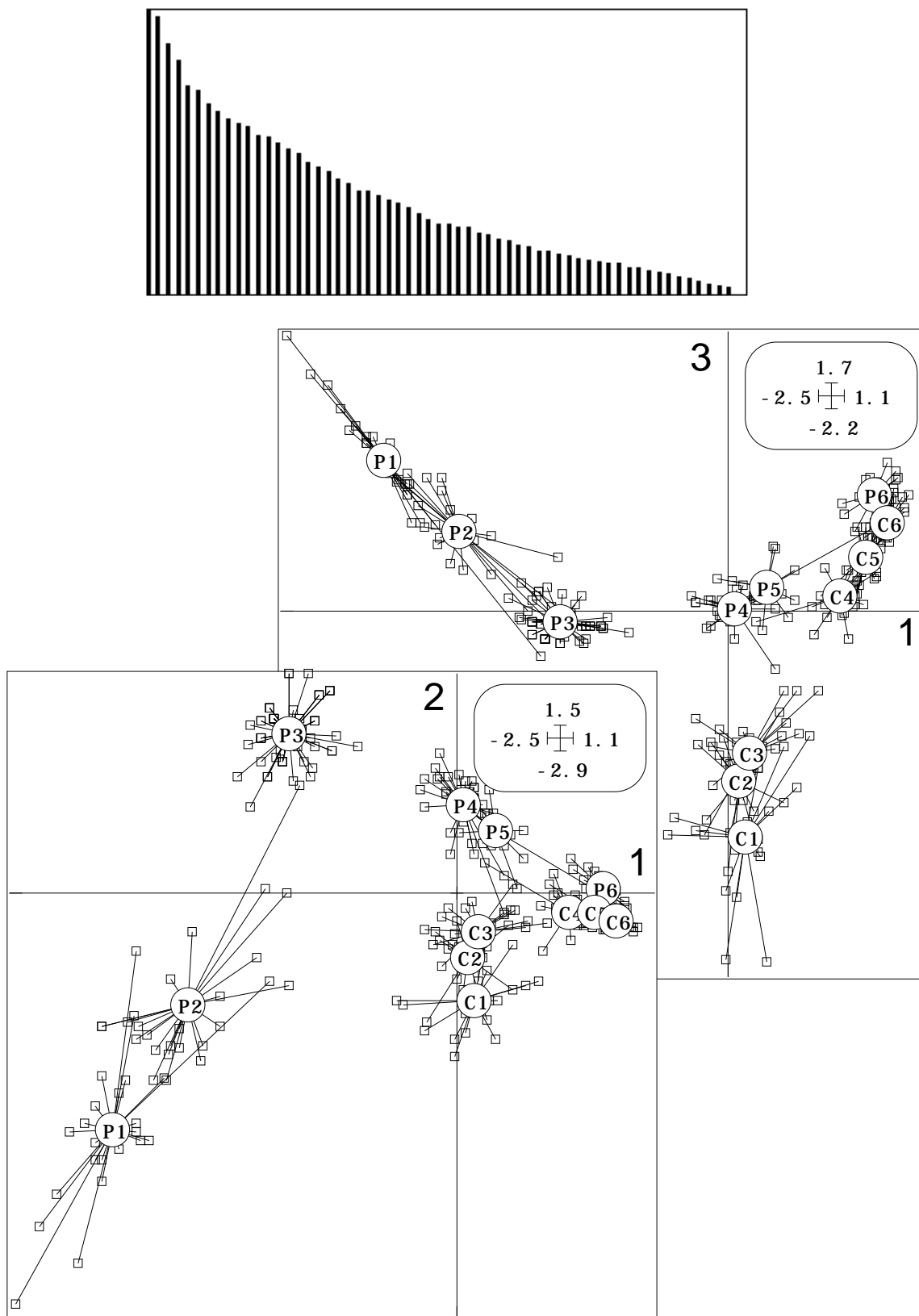


Figure 1 : Analyse inter-classes région-végétation (effet conjoint).

Cette analyse élimine totalement le plan d'observation et l'on voit immédiatement que dans l'inertie résiduelle qui est forte (3.6281) la structure est très faible (5.07 % sur

l'axe 1). Par contre l'inter-classes correspondante qui est l'ACPVI directe avec les mêmes paramètres :

PCA on Instrumental Variables			
Explanatory variables: .@ob file		W_AxB.@ob	302 11
Dependant variables: .**ta		Avi.fcta	302 60
Output file name		Provi	

L'inertie de cette analyse (2.111) est prise en compte à 70 % par trois valeurs propres (.6681, .4393 et .3728) qui sont de l'ordre de grandeur des valeurs propres de l'analyse simple. Les cartes sont très voisines de celle de l'analyse simple (figure 1).

On retiendra simplement la décomposition :

Effet	Dimension	Inertie	Proportion
AxB	11	2.1110	36.8 %
Résiduel	49	3.6281	63.2 %
Total	60	5.7391	

Il apparaît alors comme équivalent d'extraire du tableau de données sa structure directement par une analyse simple ou par une contrainte (projection sur AxB) puis une analyse. On peut augmenter le poids de la contrainte.

2.5 — L'effet additif A+B

Dans AxB on trouve l'espace A+B. L'inertie projetée diminue toujours avec la taille du sous-espace (sa dimension en terme algébrique, dimension qui définit les degrés de liberté de l'analyse de variance) :

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		W_A+B.@ob	302 6
Dependant variables: .**ta		Avi.fcta	302 60

Triplet Inertia Decomposition			
Explanatory variables: .@ob file		W_A•B.@ob	302 5
Dependant variables: .**ta		Avi.fcta	302 60

conduit à la décomposition :

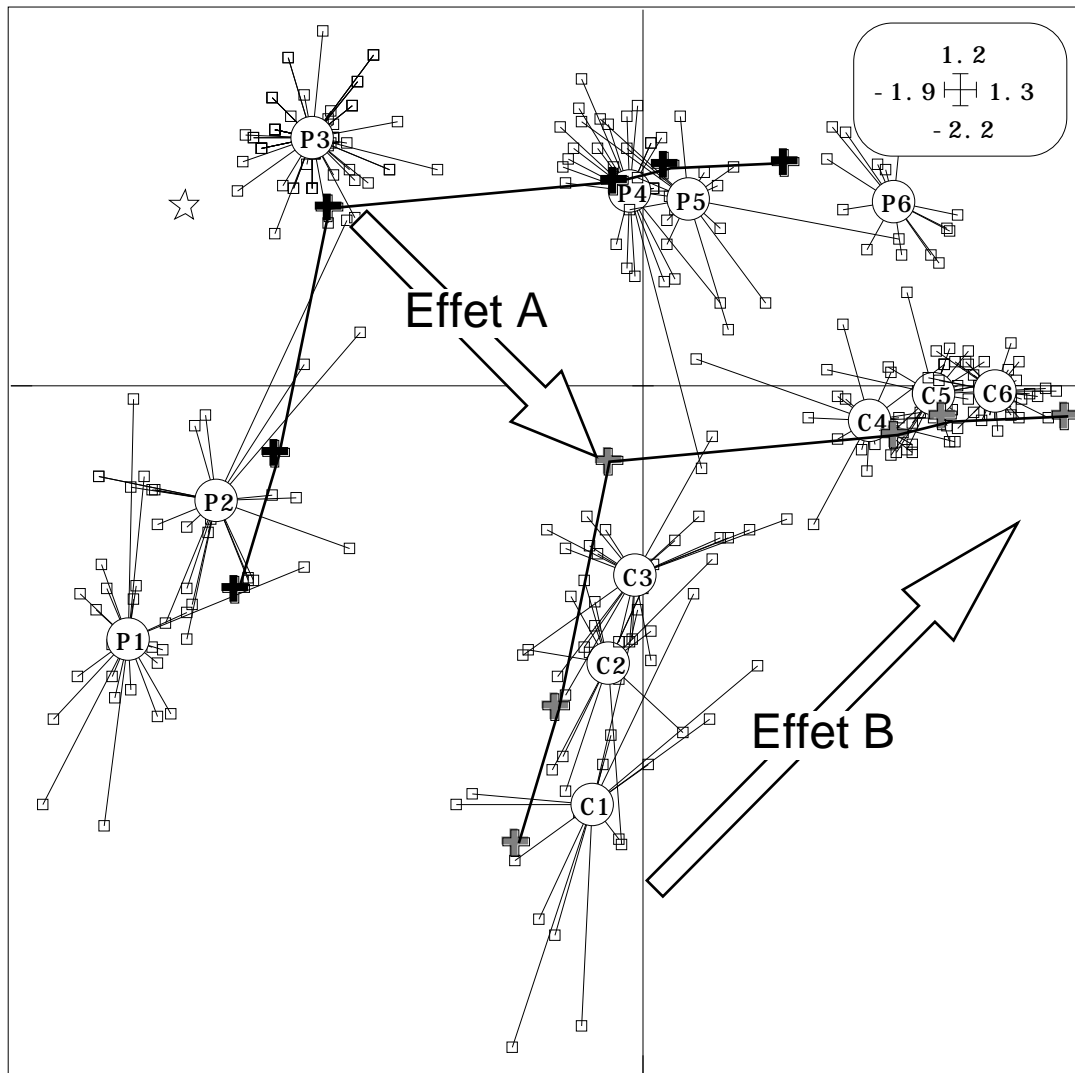
Effet	Dimension	Inertie	Proportion
A+B	5	1.4264	67.6 %
A•B	6	0.6846	32.4 %
AxB	11	2.1110	

L'espace A•B est la partie orthogonale complémentaire de A+B dans AxB. L'interaction est ce qui n'est pas additif dans le rôle des classes des variables croisées. C'est une notion délicate et fondamentale en statistique. L'analyse sous contrainte A+B cherche à fabriquer des coordonnées factorielles présentant l'addition d'un effet région et d'un effet végétation. Utiliser encore :

PCA on Instrumental Variables			
Explanatory variables: .@ob file	<input type="button" value="👉"/>	<input type="text" value="W_A+B.@ob"/>	302 6
Dependant variables: .**ta	<input type="button" value="👉"/>	<input type="text" value="Avi.fcta"/>	302 60
Output file name	<input type="button" value="👉"/>	<input type="text" value="Provi"/>	

On obtient la carte factorielle avec ScatterClass :

Stars			
HV coordinates file	<input type="button" value="👉"/>	<input type="text" value="Provi.iuls"/>	302 2
H-axis column number (default = 1)	<input type="button" value="👉"/>	<input type="text" value=""/>	
Y-axis column number (default = 2)	<input type="button" value="👉"/>	<input type="text" value=""/>	
Categories file (.cat)	<input type="button" value="👉"/>	<input type="text" value="P12.cat"/>	



L'effet A distingue les deux régions et l'effet B les strates de végétation. La carte est organisée pour expliciter un effet A+B. Pour récupérer les positions des modèles on a simplement utilisé pour la superposition :



Noter, bien qu'on ne discute pas ici du rôle des espèces dans ces analyses pour des raisons de place, que le principe d'une telle analyse est celui d'une AFCVI : 1 — positionner les espèces avec des codes numériques de variance unité, 2 — placer les relevés à la moyenne des espèces, 3 — modéliser la position des relevés ainsi obtenue par régression (ici un modèle additif A+B), l'objectif étant d'optimiser la prédictibilité (pour le type de modèle choisi) du code des relevés. On peut donc disposer les espèces pour voir l'effet régional, voir l'effet architecture de la végétation, voir les deux, n'en voir aucun, voir l'un en voyant le moins possible de l'autre. L'AFCVI est alors un véritable microscope euclidien qui met en évidence tel ou tel aspect de la structure en fonction des réglages choisis.

Ces opérations donnent ici des résultats variés qui sont pertinents essentiellement parce qu'on a 302 relevés et un plan d'expérience à 2 et 6 modalités de facteurs. Seules des groupes taxonomiques qui permettent d'obtenir un très grand nombre de relevés (avifaune ou végétation par exemple) autorisent à se poser de telles questions. Il est hors de question d'essayer des réglages au hasard pour vouloir tomber sur un résultat intéressant. Ce serait aussi stupide que de régler un microscope électronique au hasard ou de choisir au hasard un flacon pour faire une coloration. On appelle systématiquement Provi l'analyse pour écraser les fichiers créés.

2.6 — L'effet partiel B sachant A

Pour éliminer l'effet A dans la carte de l'effet A+B on utilise le sous-espace B/A. On a une nouvelle décomposition canonique de l'inertie en utilisant A+B, A et B/A qui sont deux parties complémentaires du précédent. On le vérifiera par :

Effet	Dimension	Inertie	Proportion
A	1	0.4498	31.5 %
B/A	5	0.9766	68.5 %
A+B	6	1.4264	

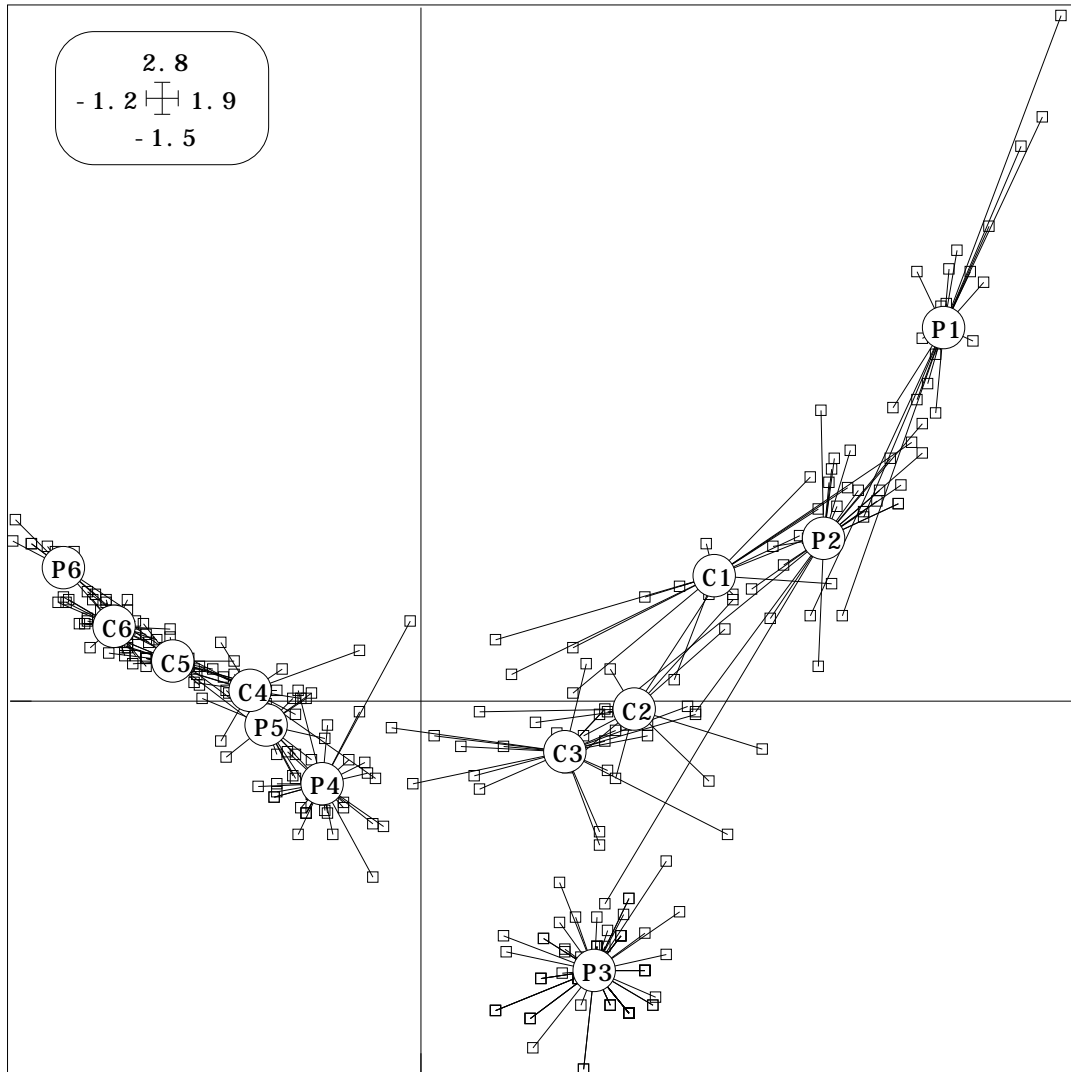
Si on veut éliminer l'effet régional en utilisant l'effet du facteur végétation on peut utiliser une Partial Canonical Correspondence Analysis (Ter Braak, op. cit.) qui est définie par :

$$\text{ACP} \left[P_{B/A} (\mathbf{D}_n^{-1} \mathbf{P} \mathbf{D}_t^{-1} - \mathbf{1}_n \mathbf{1}_t'), \mathbf{D}_n, \mathbf{D}_t \right]$$

introduite comme régression partielle sur variables qualitatives par Daudin²¹ et rendue disponible par Ter Braak comme généralisation de l'analyse canonique des correspondances²² qu'on peut considérer comme AFC sur variables instrumentales.

$P_{B/A}$ désigne le projecteur \mathbf{D}_n orthogonal sur l'intersection du sous-espace A+B classique en analyse de variance multivariée et du sous-espace orthogonal de A utilisé dans l'analyse précédente. Pour éviter la confusion avec la notion de dépendance partielle de Daudin²³ et l'analyse canonique partielle de Yanai²⁴ on peut parler d'AFC conditionnelles conformément au terme d'effet B sachant A traditionnellement attaché à l'espace B/A. Pour obtenir la nouvelle carte :

Stars	
HY coordinates file	<input type="text" value="Provi.iuls"/> 302 2
H-axis column number (default = 1)	<input type="text"/>
Y-axis column number (default = 2)	<input type="text"/>
Categories file (.cat)	<input type="text" value="P12.cat"/>



La figure montre l'amélioration importante introduite par la nouvelle contrainte. En dépit de la faible proportion d'espèces communes aux deux régions (25 sur un total de 60) les deux ordinations sur l'ouverture du milieu sont très voisines, la séparation des classes 3 montrant qu'elle n'est pas parfaite, soit par le biais méthodologique soit par impossibilité biologique.

2.7 — Contraintes inter-intra




Pour trancher l'ambiguïté précédente il est possible d'utiliser l'intersection du sous espace B (effet inter-strates) et de l'orthogonal du sous-espace A (effet intra-régions) défini théoriquement par Afriat²⁵. Le sous-espace impose un modèle des coordonnées centré par région et constant par classes de végétation. Les travaux de Pontier²⁶ utilise le sous-espace dans une analyse canonique appelée LONGI, sans donner une méthode

convaincante de calcul pour cet espace. Cette contrainte n'est possible que si m_A (ici 2) est plus petit que m_B (ici 6). On peut alors définir :

$$ACP \left[P_{B/A} (D_n^{-1} P D_t^{-1} - \mathbf{1}_n \mathbf{1}'_t), D_n, D_t \right]$$

qui pourrait s'appeler AFC inter-intra car elle exécute simultanément l'AFC intra-classes associée à A et l'AFC inter-classes associée à B.

Pour obtenir ce sous-espace, utiliser :

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

Orthonormal basis of an intersection of subspaces




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





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

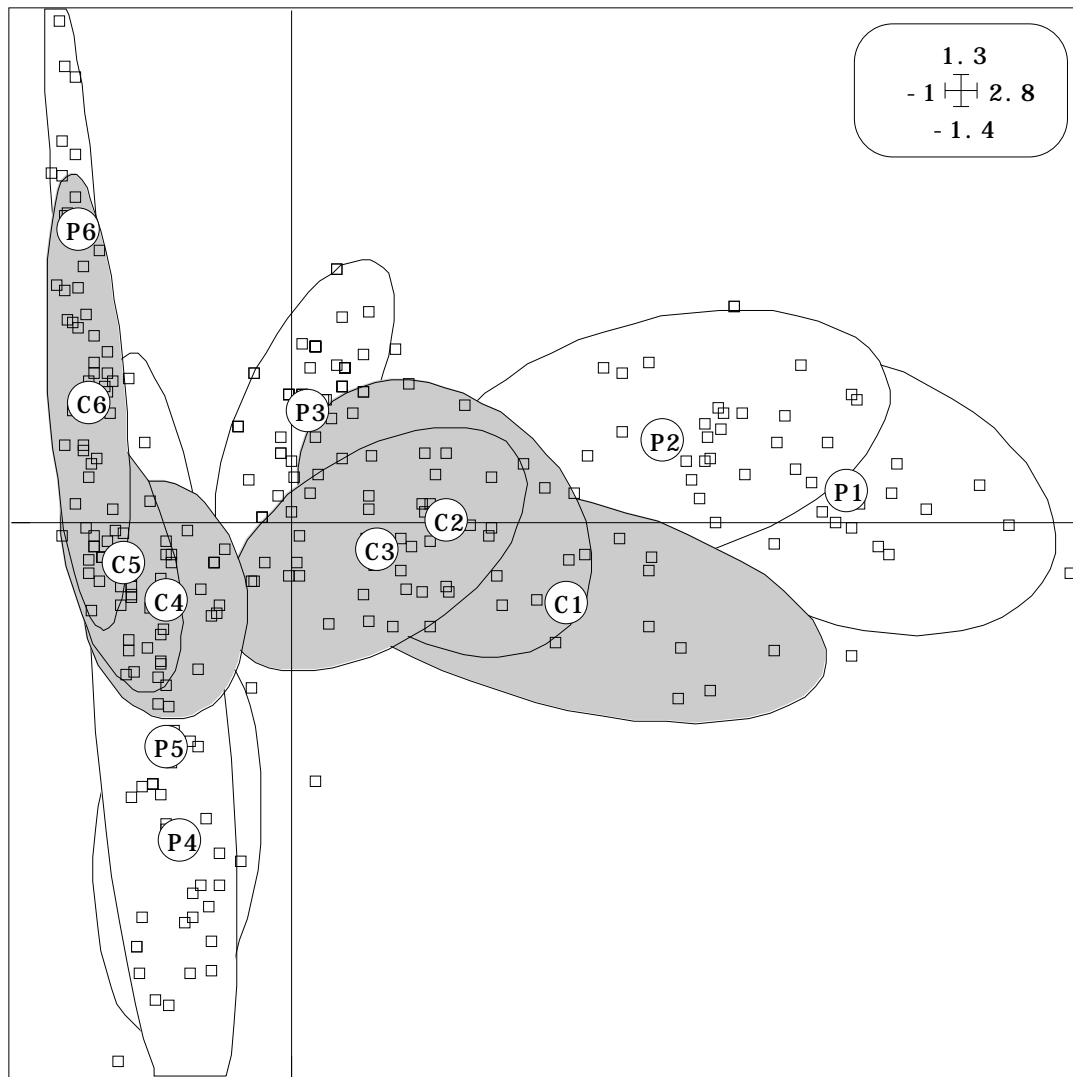
Dimension of X-Inter-Y is: 4
 file Z_XiY.@ob contains an orthonormal basis of subspace
 X_Inter_Y
 It has 302 rows and 4 columns

Le module permet d'obtenir des bases orthonormales de sous-espaces intersection de deux sous-espaces donnés ici B et B/A dont l'intersection est $B \cap A$. On utilise les valeurs propres égales à 1 de l'analyse canonique des deux sous-espaces, ce qui est souligné par le listing d'exécution et est numériquement bien meilleur que l'usage des valeurs propres nulles décrit par Pontier.

L'analyse se poursuit comme les précédentes :

PCA on Instrumental Variables		
Explanatory variables: .@ob file		Z_XiY.@ob 302 4
Dependant variables: .**ta		Avi.fcta 302 60
Output file name		Provi

Ellipses		
H-Y coordinates file		Provi.iuls 302 2
H-axis column number (default = 1)		
Y-axis column number (default = 2)		
Categories file (.cat)		P12.cat



Le résultat consigné sur la figure 3 est surprenant de signification écologique : **l'ordination simultanée est parfaite et les relevés de Corse viennent se placer à l'intérieur du nuage des points de Provence, image de la diminution de l'originalité des relevés en milieu insulaire liée à la dilatation des niches écologiques.** La carte des espèces associée consoliderait la discussion développée par J. Blondel (op. cit.).

Les méthodes d'ordination sous contrainte linéaire renouvellent en profondeur les capacités d'analyse des structures écologiques complexes. Elles ne sont pas propres à l'AFC choisie ici comme analyse de base en relation avec les propriétés des tableaux ornithologiques²⁷ et le caractère unimodal des courbes de réponse des espèces au gradient d'ouverture de la végétation²⁸. En hydrobiologie où les variations de richesse et d'abondance sont liées aux perturbations du milieu on pourra faire un autre choix. Ces méthodes sont importantes car elles autorisent l'introduction d'objectifs expérimentaux affirmés avant l'analyse (prise en compte et/ou élimination d'effets). L'approche des structures spatio-temporelles de la faune en milieu aquatique dans la logique des analyses sous contraintes des variables de milieu est désormais possible. Les problèmes d'identification des procédures utilisées exige une définition mathématique stricte comme le souligne Kenkel et Orloci²⁹, les notions de schéma de dualité et de projecteurs³⁰ apparaissent de ce point de vue remarquablement efficaces. Le dépouillement des résultats est strictement dans la logique des méthodes d'ordination linéaires bien connues.

3 — L'ACP sur variables instrumentales

Les analyses sur variables instrumentales sont d'un type quelconque : leur point commun est d'utiliser un sous-espace de projection dans une optique AFC comme ACP. Pour illustrer cette communauté de raisonnement nous reprenons l'exemple central de la fiche "Régression linéaire" de présent fascicule.

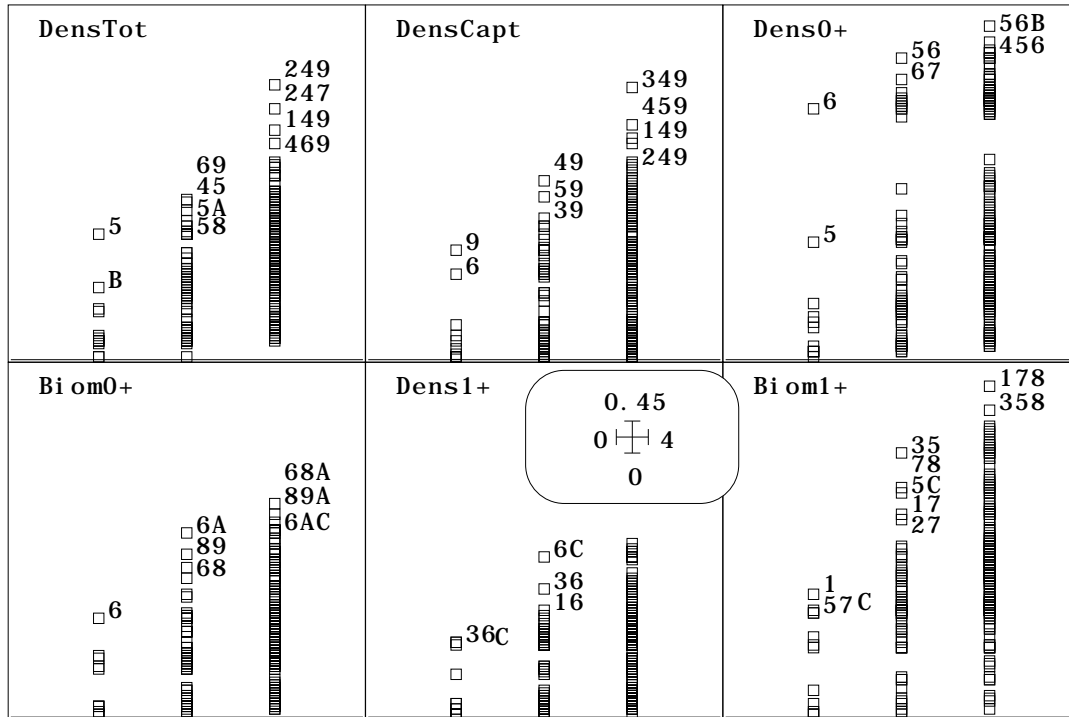


Figure 2 : Carrés de corrélation des 6 variables de Taille sur les combinaisons de 1, 2 ou 3 variables de Mil. Une élimination de certaines variables instrumentales ne semble pas nécessaire.

Le tableau Mil contient des variables explicatives de l'abondance contenue dans l'unique colonne du tableau Taille. On a un problème classique de régression qui peut être abordé sans difficultés particulières. Un peu moins classique, pour l'utilisateur occasionnel de la statistique est la question de la prévision du tableau Forme qui contient 6 variables par le tableau Mil qui en contient 12. Comment trouver dans le premier ce qui se modélise par tout ou partie du second ?

L'option MultCorCoeff (figure 2) donne une première idée :

MultCorCoeff			
Explanatory variables	<input type="checkbox"/>	Mil	33 12
Dependent variables	<input type="checkbox"/>	Forme	33 6
Maximal number of dependent	<input type="checkbox"/>	3	
Option: row weighting	<input type="checkbox"/>		
Output file name	<input type="checkbox"/>	Forme/Mil	

Il n'est plus question de construire des modèles pour chaque variable : on sait qu'elles présentent une redondance certaine. On ne cherche plus, d'ailleurs de modèles prédictifs au sens strict (comme précédemment où l'usage de point supplémentaires permettrait d'évaluer une biomasse potentielle dans les unités d'origine), mais plutôt une description des relations entre deux ensembles de variables dont le second (Forme)

est formé déjà de résidus de régression. Ceci signifie que, par exemple, la variable Biom+1 ne définit pas la biomasse des truites 1+, mais l'écart de cette biomasse avec ce qu'on attend, en moyenne, pour le système étudié, de la biomasse de cette catégorie. Il n'est désormais plus question de parler de la densité du peuplement mais de ses caractéristiques et ce indépendamment, *a priori*, de cette densité.

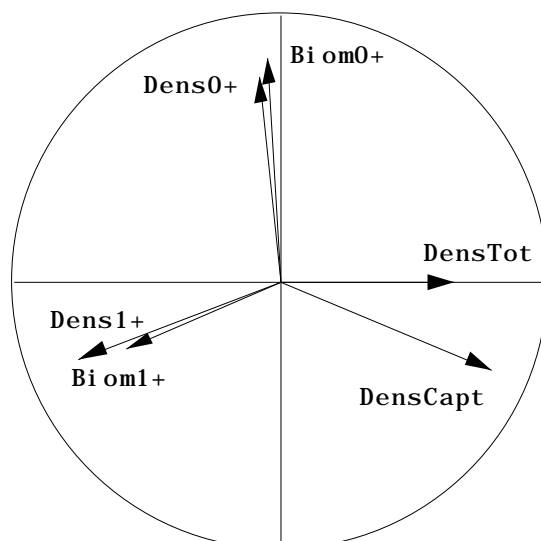
Les unités étant différentes dans les variables des deux groupes, elles sont toutes normalisées. Les variables instrumentales sont normalisées par l'option Centring de BinToBin) :

Centring	
Input file	<input type="button" value="👉"/> Mil 33 12
Option: file for row weighting	<input type="button" value="👉"/> <input type="text"/>
Option for H matrix (no default)	<input type="button" value="👉"/> 3
Output file	<input type="button" value="👉"/> Mil0

Les poids des lignes sont définitivement les poids uniformes (1/33) dans tout ce qui suit. Les variables dépendantes sont préparées par ACP normée :

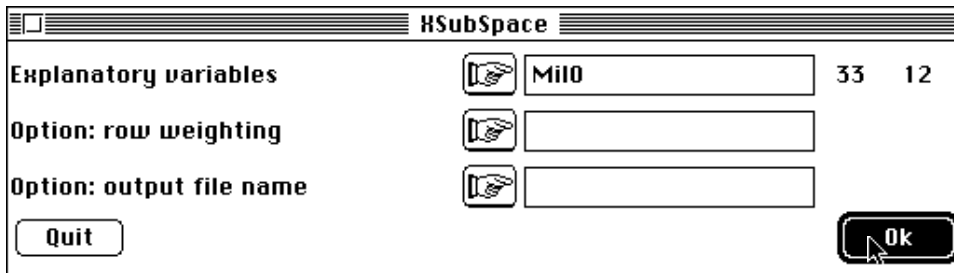
Correlation matrix PCA	
Matrix input file	<input type="button" value="👉"/> Forme 33 6
Row weights (default=1/n)	<input type="button" value="👉"/> <input type="text"/>
Column weights (default=1)	<input type="button" value="👉"/> <input type="text"/>
Option: file for row weighting	<input type="button" value="👉"/> <input type="text"/>
Option: file for column weighting	<input type="button" value="👉"/> <input type="text"/>
1 = Save correlation matrix	<input type="button" value="👉"/> 1

On obtient un plan 1-2 qui est conforme à la figure 8 de la fiche "Régression linéaire". Rappelons, car c'est une question souvent posée, qu'on ne gagne, très généralement, rien à refaire des analyses en enlevant des variables pour faire passer dans le premier facteur des éléments apparus dans les facteurs lointains (ou en enlevant un point qui a trop d'importance). Ici, en enlevant l'axe 1 de TLog (sous la forme de régression orthogonale à sa première variable), on obtient rien d'autre que le plan 2-3 de la figure 8 :



3-1 — Définition du sous-espace de projection

Dans Projectors, définissons le sous-espace de projection associé aux variables instrumentales :

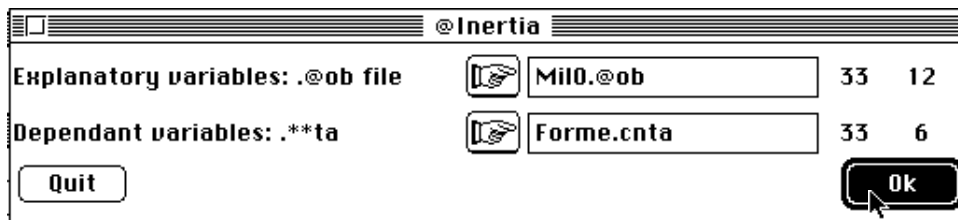


```
Orthonormalization: subspace generated by quantitative
variables
```

```
-----
Explanatory variable file: Mi10
It has 33 rows and 12 columns
```

```
-----
Orthonormal basis: Mi10.@ob
It has 33 rows and 12 columns
Row weighting file: Mi10.@pl
Uniform row weights = 0.030303
Coordinates of the vectors of the orthonormal basis
in the initial basis in : Mi10.@co
File Mi10.@co has 12 rows and 12 columns
-----
```

Le sous-espace de projection étant défini, la première question est celle de la position des variables dépendantes par rapport aux explicatives. L'option :

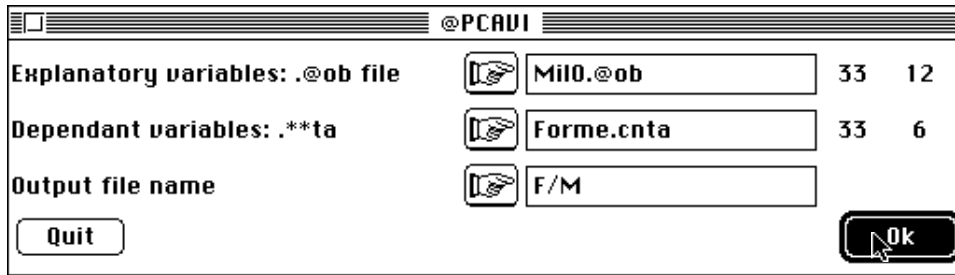


affiche :

```
Projected inertia on a subspace
```

```
-----
Orthonormal basis: Mi10.@ob
It has 33 rows and 12 columns
Dependant variable file: Forme.cnta
It has 33 rows and 6 columns
-----
```

	Subspace A	A Orthogo	Total	A+	A-
1	5.1765e-01	4.8235e-01	1.0000e+00	5176	4823
2	5.0665e-01	4.9335e-01	1.0000e+00	5066	4933
3	5.6979e-01	4.3021e-01	1.0000e+00	5697	4302
4	4.0213e-01	5.9787e-01	1.0000e+00	4021	5978
5	3.8677e-01	6.1323e-01	1.0000e+00	3867	6132
6	6.0044e-01	3.9956e-01	1.0000e+00	6004	3995
Tot	2.9834e+00	3.0166e+00	6.0000e+00	4972	5027



On obtient un listing très général qui a été explicité à plusieurs reprises. Ce module est un programme d'analyse en composantes principales sur variables instrumentales paramétré par les options d'entrée dont des cas particuliers sont les options *Redundancy Analysis* et *Canonical Correspondence Analysis* du programme CANOCO³¹.

```

*-----*
Instrumental variables
----- input -----
Orthonormal basis: Mil0.@ob
It has 33 rows and 12 columns
Dependant variable file: Forme.cnta
It has 33 rows and 6 columns
----- output -----
Projected variable file: F/M.ivta
It has 33 rows and 6 columns
Inertia: 2.9834e+00
File F/M.ivpc contains the column weights
It has 6 rows and 1 column
File F/M.ivpl contains the row weights
It has 33 rows and 1 column

```

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+1.1940E+00	+0.4002	+0.4002	02	+9.0633E-01	+0.3038	+0.7040
03	+5.1764E-01	+0.1735	+0.8775	04	+2.4122E-01	+0.0809	+0.9583
05	+1.0428E-01	+0.0350	+0.9933	06	+1.9998E-02	+0.0067	+1.0000

```

File F/M.ivvp contains the eigenvalues and relative inertia
for each axis
It has 6 rows and 2 columns

```

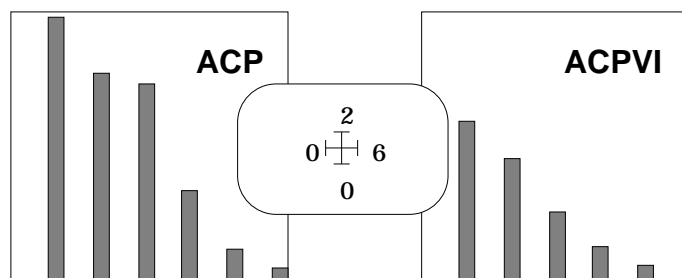


Figure 3

```

File F/M.ivco contains column scores (dependant variables)
with lambda norm
It has 6 rows and 2 columns
File : F/M.ivco
----- Minimum/Maximum:

```

```

Col.: 1 Mini = -0.62017 Maxi = 0.66665

```


Col.: 2 Mini = -0.39874 Maxi = 0.68497

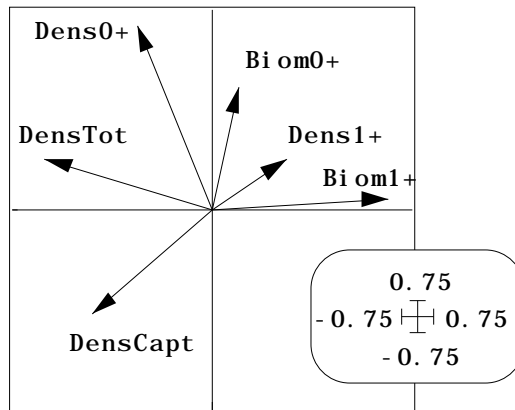


Figure 4

File F/M.ivli contains predicted row scores (dependant variables after projection)
It has 33 rows and 2 columns

File : F/M.ivli

----- Minimum/Maximum:
Col.: 1 Mini = -3.0174 Maxi = 1.7793
Col.: 2 Mini = -1.5343 Maxi = 2.3552

files F/M.ivta
F/M.ivpc
F/M.ivpl
F/M.ivpa
F/M.ivvp
F/M.ivco
F/M.ivli
are those of the complete analysis
of the projected table (DDUtil can be used)

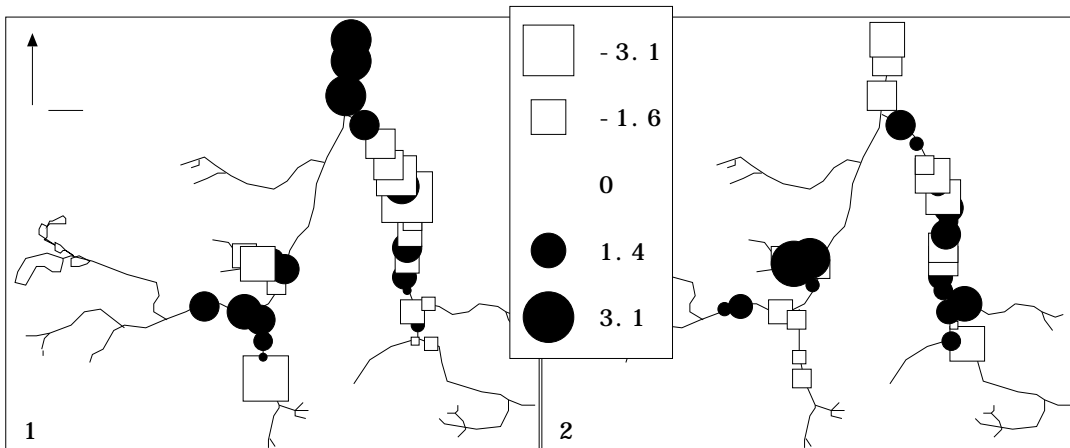


Figure 5

La première partie est une analyse standard d'un tableau particulier (F/M.ivta avec sa pondération uniforme des lignes dans F/M.ivpl, sa pondération unitaire des colonnes dans F/M.ivpc, les paramètres dans F/M.ivpa, les valeurs propres dans F/M.ivvp, les coordonnées des lignes dans F/M.ivli et les coordonnées des colonnes dans F/M.ivco). On obtient exactement le même résultat qu'avec une ACP centrée ordinaire du tableau des prévisions des variables de Forme par celle de Mil.

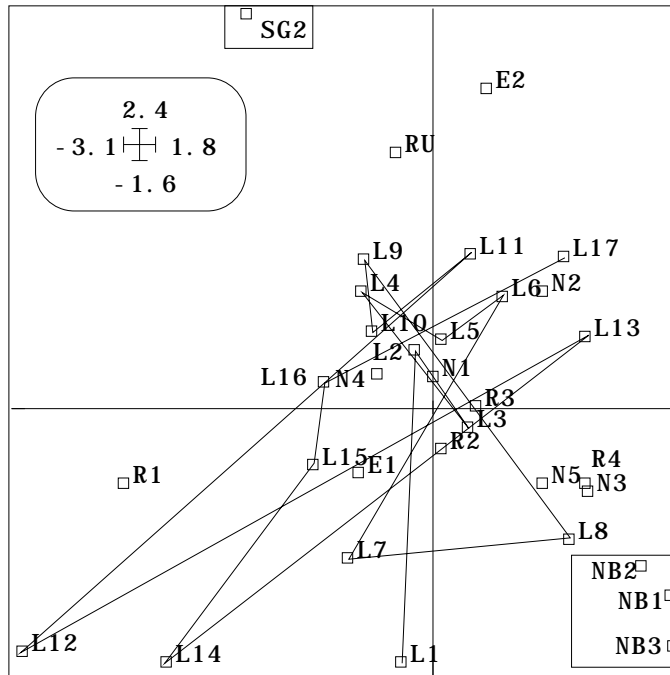


Figure 6

On a gardé deux axes parce que la décomposition de l'inertie n'est pas caractéristique (figure 3 : valeurs propres) et que les résultats doivent être interprétés sans excès. On sait avoir perdu dans la projection 50% de l'inertie et sans doute une bonne partie de la redondance entre les trois couples (Densité-Biomasse 0+, Densité-Biomasse 1+, Densité totale et capturable). C'est visible sur la figure 4 (carte des variables). La cartographie des coordonnées lignes (figure 5) et la figure 6 (carte factorielle) montre des oppositions localisées le long des cours d'eau et isole la zone aval et la station SG2.

Ce point de vue de l'analyse d'un tableau de modèles prévisionnels peut être utile. Ici, il donne guère d'interprétation signifiante par lui-même. Le listing indique seulement que c'est un point de vue possible. L'essentiel est dans les deux points de vue complémentaires exprimés par :

```

| files F/M. ivc1
|       F/M. ivls
|       F/M. ivli
| allow a convenient interpretation
|-----
| files F/M. ivfa
|       F/M. ivl1
|       F/M. ivco
| allow a convenient interpretation
|-----

```

Le premier part du principe qu'on cherche dans le tableau des variables dépendantes ce qui est au mieux prédictible. Au lieu de prédire chacune des variables cherchons la combinaisons des variables à prédire (dans `Forme.cnta`) donnant le meilleur résultat. On désire cependant ne pas constituer une combinaison qui n'aurait comme propriété que la prédictibilité. On sait en effet qu'en analyse canonique qui fait cela, on trouve facilement des carrés de corrélation optimaux avec des combinaisons de variables bizarres qu'on sait mal interpréter. On va donc maximiser non pas le pourcentage de variance prédite mais la variance prédite elle-même ce qui oblige en même temps à maximiser la variance comme dans une ACP normale. La valeur propre de cette analyse est une variance expliquée, donc le produit d'une variance et d'un carré de corrélation :

Num	Variance	R2	Ei g. Val.

1	1.839	0.6493	1.194
2	1.551	0.5844	0.9063
---	-----	-----	-----

On peut constituer une combinaison des variables biologiques de variance 1.839 (on ne pourrait dépasser l'optimum qui est de 1.969 (la première valeur propre de l'ACP de Forme) dont 64.93% peut être expliquée par une régression multiple sur Mil. D'où la question : comment sont constituées les combinaisons à prédire et par quoi sont-elles prédites ? Les coefficients utilisés dans la constitution de la combinaison à prédire sont dans :

```
File F/M.ivc1 contains column scores (dependant variables)
with unit norm
It has 6 rows and 2 columns
File : F/M.ivc1
----- Mini mum/Maxi mum:
Col. : 1 Mini = -0.56757 Maxi = 0.61011
Col. : 2 Mini = -0.41884 Maxi = 0.71949
File F/M.ivls contains row scores (dependant variables before
projection). It has 33 rows and 2 columns
```

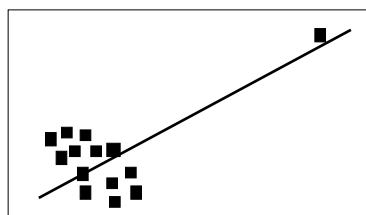
On trouve les valeurs de ces combinaisons de variables à expliquer dans :

```
File : F/M.ivls
----- Mini mum/Maxi mum:
Col. : 1 Mini = -5.5966 Maxi = 2.4092
Col. : 2 Mini = -3.0161 Maxi = 4.4302
```

Les prévisions sont exactement les valeurs du fichier F/M.ivli. On porte donc dans la figure 7 les coefficients, les valeurs à prédire et les prédictions associées par l'option Match two graphs du module Scatters. On a rajouté les corrélations entre variables de F/M.ivls et celles de Forme.cnta. En effet, il est toujours bon de confronter les coefficients qui fabriquent une combinaison linéaire et les corrélations de cette combinaison linéaire avec les variables de départ. toute contradiction est l'indice d'une instabilité numérique et seul ce qui concorde est habituellement utilisé. Ici la cohérence est bonne.

La présence des points extérieurs impose de retourner aux données (Tableau 1). La signification statistique n'étant pas considérable, il faut vérifier la cohérence de la figure obtenue. La variable DensCapt (densité de truites pêchables corrigée par la biomasse totale présente une seule valeur extrême de 5.2 écarts-types. La station L12 est très particulière et présente des valeurs de Dens+1 et Biom+1 négatives, ce qui organise l'axe 1. La variable Dens0+ présente deux valeurs extrêmes (3.72 en SG2 et 2.94 en L9) ce qui éloigne les stations correspondantes sur le plan. La distribution n'est cependant pas aberrante.

L'essentiel de la signification statistique se concentre donc sur les variables Dens+0 et Biom+1. Il est remarquable de distinguer ici les composantes de la densité de la population en choisissant dans les deux groupes une variable de chaque type. La biomasse des 1+ inclut en effet la densité et les contraintes sur la croissance qui n'est pas sensible sur les 0+. La position des variables 1 et 2 sont des conséquences liées aux valeurs extrêmes qui ne sont pas bien prédites par les variables de milieu et donc limitent le R². On n'a donc pas mis en évidence les valeurs extrêmes pour caler une régression sans signification du type pathologique :



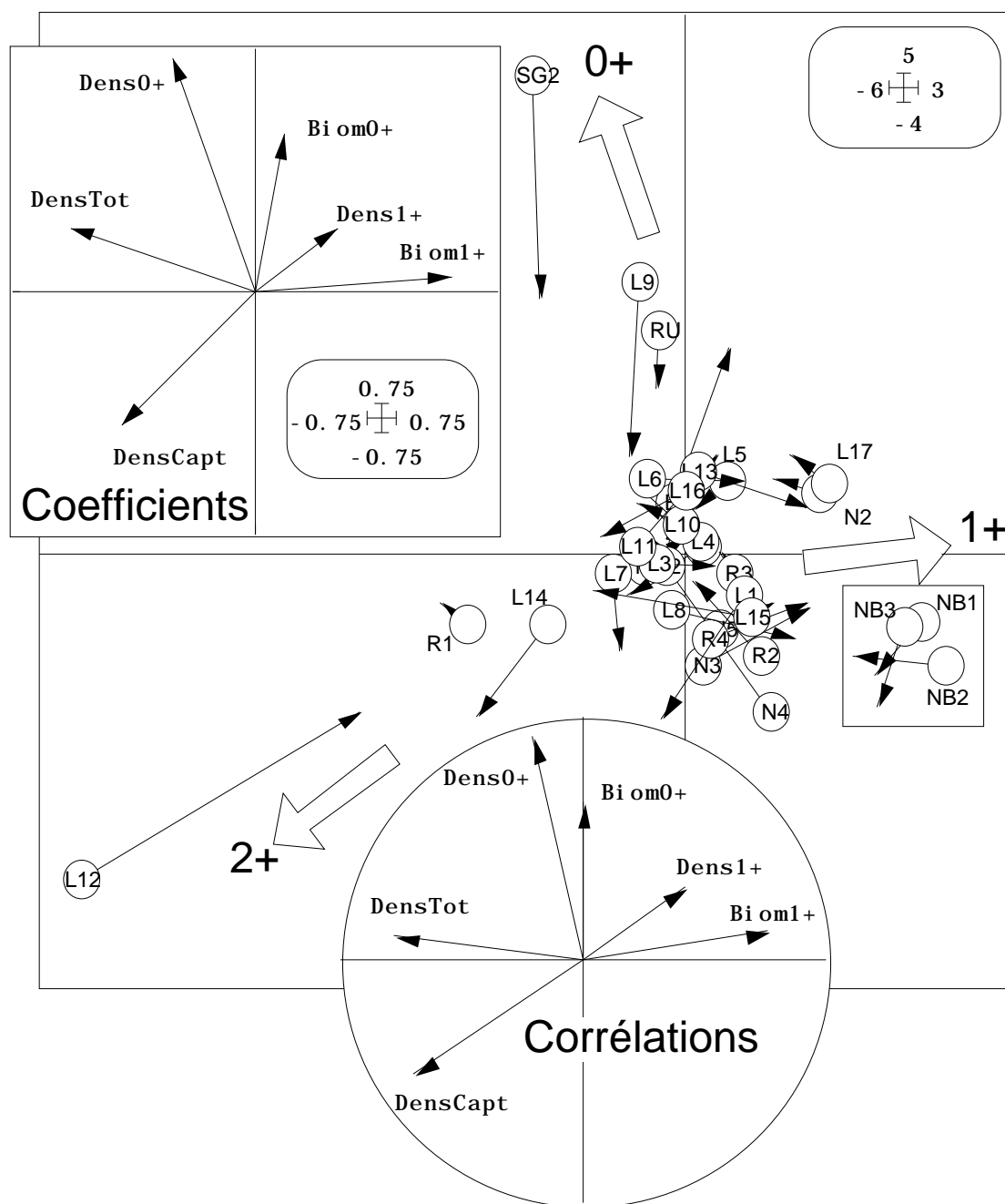


Figure 7 : Positions des stations sur deux codes combinaisons des variables biologiques (cercles). Coefficients des combinaisons et corrélations avec les variables de départ. La flèche désigne la prévision de la position à partir des variables de milieu.

Reste alors à interpréter les variables prédictrices. On utilise la dernière partie des fichiers créés. Les prédicteurs sont (à une constante près) des combinaisons linéaires de variance unité, appelés, comme en analyse discriminante, codes canoniques :

```
File F/M.ivl1 contains canonical row scores
linear combination of explanatory variables with unit norm
It has 33 rows and 2 columns
```

File : F/M.ivl1

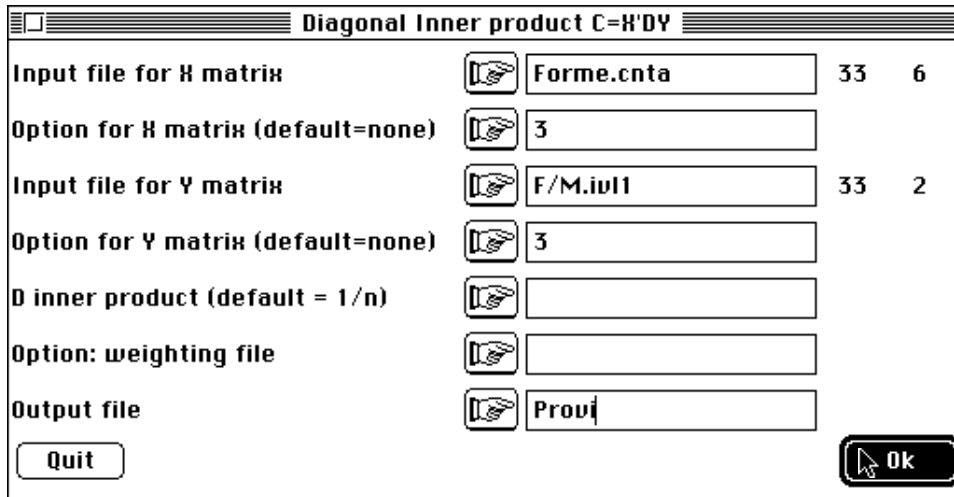
----- Minimum/Maximum:

Col. : 1 Mini = - 2. 7614 Maxi = 1. 6283
 Col. : 2 Mini = - 1. 6116 Maxi = 2. 4739

	DensTot	DensCapt	Dens0+	Biom0+	Dens1+	Biom1+
N1	0.305	-0.625	-0.136	-0.583	-0.234	-0.643
N2	0.226	-0.097	0.024	0.459	0.902	1.729
N3	-0.185	0.046	-0.442	-1.348	-0.147	0.195
N4	0.042	-0.871	-1.063	-2.629	0.896	0.325
N5	-0.653	-0.163	-0.421	-0.473	-0.547	-0.079
E1	0.114	0.046	0.171	-0.222	0.126	0.443
E2	0.451	-0.811	0.171	-0.353	0.477	-0.387
SG2	2.561	-0.663	3.720	0.993	2.643	-0.138
R1	1.051	0.879	0.358	-1.643	0.470	-1.546
R2	-0.224	-0.019	-0.803	-0.992	0.638	0.470
R3	0.037	-0.128	-0.390	-0.027	0.149	0.460
R4	-0.515	-0.177	-0.429	-0.822	-0.352	-0.126
L1	0.002	-0.460	-0.433	-0.781	0.460	0.328
L2	-0.749	-0.271	-0.562	1.080	-0.809	-1.236
L3	-0.848	-0.307	-0.504	1.142	-0.914	-1.435
L4	0.432	0.139	0.078	-0.245	0.589	0.515
L5	-0.415	-0.205	0.171	1.313	-0.507	0.169
L6	0.506	0.673	0.144	1.590	0.076	0.093
L7	0.281	0.515	0.065	-0.179	0.142	-0.526
L8	-0.962	-0.138	-0.485	0.514	-1.308	-0.944
L9	-0.119	-0.411	2.941	0.546	-0.043	0.072
L10	-0.399	-0.345	0.520	-0.089	-0.679	-0.162
L11	-0.035	-0.206	-0.122	0.447	-0.634	-0.782
L12	3.957	5.200	-1.083	-0.955	-1.567	-1.651
L13	-0.026	-0.258	0.304	1.009	-0.260	0.071
L14	-0.366	-0.118	-0.321	0.136	-2.040	-1.831
L15	-1.684	-0.429	-0.283	-0.053	-0.939	-0.590
L16	-0.116	-0.286	0.170	0.892	-0.309	-0.238
L17	-0.660	-0.214	0.440	1.021	-1.038	1.872
NB1	-0.629	0.385	-1.005	-0.332	2.404	1.850
NB2	-0.895	-0.630	-1.081	-1.473	1.429	1.874
NB3	-0.848	0.424	-0.998	0.134	1.207	1.874
RU1	0.362	-0.476	1.282	1.923	-0.284	-0.026

Tableau 3 : Valeurs normalisées du tableau Forme.

Ces combinaisons d'explicatives qui ont la propriété de prédire les meilleurs variables à prédire ont aussi la propriété de maximiser la somme de leurs coefficients de corrélation (au carré) avec les variables à prédire. Cette seconde propriété est facile à expliciter. Calculer les corrélations entre variables de Mil0 et Forme.cnta (MatAlg) :

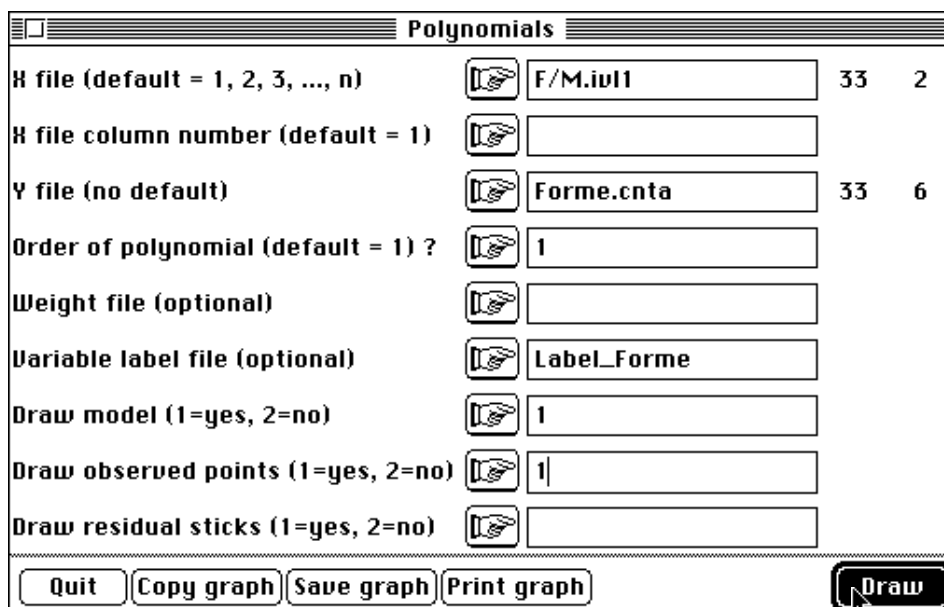


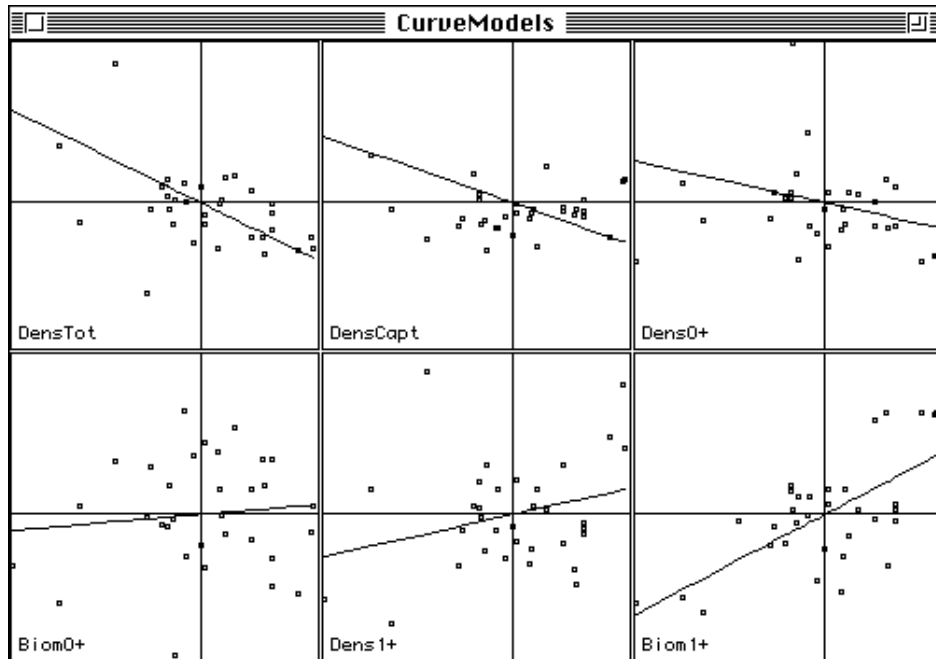
Éditer le fichier créé :



```
[ 1] -6.2017e-01  1.8505e-01
[ 2] -4.4574e-01 -3.9874e-01
[ 3] -2.7532e-01  6.8497e-01
[ 4]  1.0412e-01  4.5713e-01
[ 5]  2.8211e-01  1.8266e-01
[ 6]  6.6665e-01  3.9734e-02
```

Vérifier que la somme des carrés de la première colonne est exactement la première valeur propre (1.1940). Il n'est pas possible de faire mieux. Le critère Somme des carrés des corrélations avec un ensemble de variables est dit de Stewart et Love³². On sait que la première coordonnée d'une ACP normée maximise ce critère avec les variables du tableau traité, tandis qu'ici on maximise ce critère avec les variables de Forme en utilisant uniquement une combinaison des variables de Mil. On peut illustrer par CurveModels :





Les variables DensTot (distribution douteuse) et Biom+1 sont prédites à plus de 60%. Les coefficients de ces prédicteurs sont dans le dernier fichier :

```
File F/M.ivfa contains weights
coefficients of linear combination of explanatory variables
with unit norm
It has 12 rows and 2 columns
```

```
File : F/M.ivfa
----- Minimum/Maximum:
Col.: 1 Mini = -0.68609 Maxi = 0.88666
Col.: 2 Mini = -0.8649 Maxi = 0.77882
```

Nous pouvons alors retrouver le rôle des variables explicatives (figure 21) soit par les coefficients des combinaisons soit par les corrélations entre ces combinaisons (F/M.iv11) et les variables (Mil0). Contrairement au cas précédent (figure 20), la contradiction est flagrante. La variable module intervient avec un coefficient négatif sur le code 1 et un coefficient positif sur le code 2 et présente des corrélations de signes inverses. Des variables qui participent avec des coefficients non négligeables ont des corrélations nulles (Pente) et inversement (Température). C'est surtout les coefficients des variables Module, Pente et Température qui interdisent pratiquement d'interpréter le descripteur. Du côté variables explicatives, cette analyse a le principal défaut de l'analyse canonique. L'article de ter Braak (1990)³³ donne une synthèse des conflits d'appréciation nés de cette contradiction.

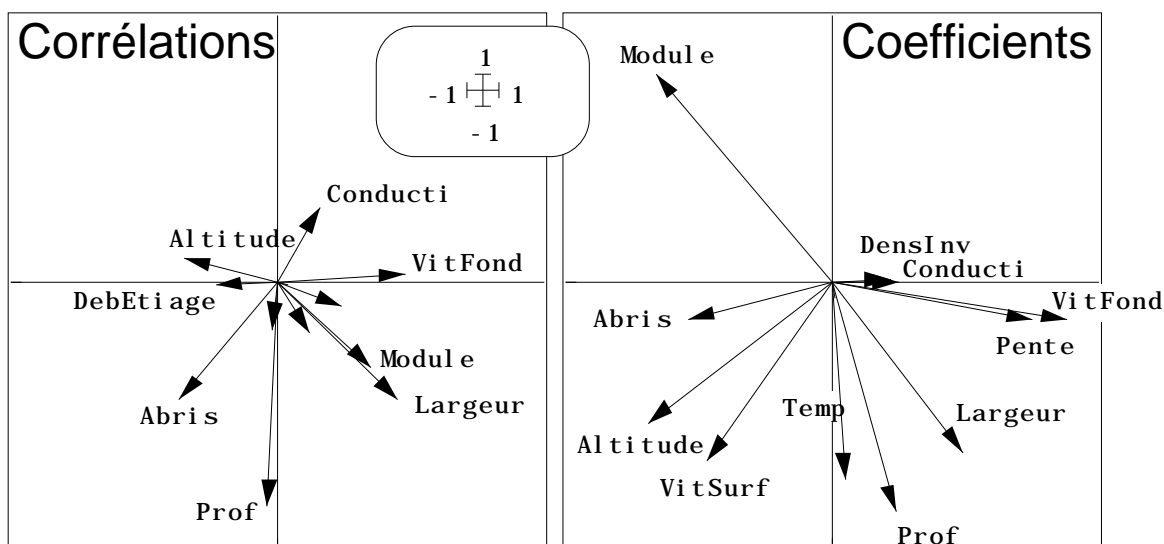


Figure 21

3-4 — Bilan

L'essentiel de l'échec vient du rapport entre le nombre d'explicatives (10) et le nombre de stations (33). Il est facile de trouver des combinaison de variables biologiques prédictibles à 60 % par des variables de milieu, mais on ne peut donner au prédicteur une signification. Pourtant la liaison globale est significative : sa description est des plus délicate. On notera que la fiche décrit le fonctionnement de la méthode et ses limites.

On retiendra que l'ACPVI traite de deux paquets de variables quantitatives, explicatives et à expliquer. Elle fonctionne sur l'analyse du tableau des valeurs prédites par les régressions multiples variable par variable. Elle permet d'obtenir une combinaison des variables à prédire de R^2 maximum ou une combinaison d'explicatives maximisant la somme des corrélations avec les expliquées. Elle s'interprète du côté des explicatives comme une ACP (on dit ACP sur variables instrumentales pour préciser que cette ACP est contraintes par les explicatives). Du côté des explicatives elle s'interprète comme une analyse discriminante (combinaison de variables de variance unité optimisant le critère de Stewart et Love). Les tests de permutations³⁴ permettent de contrôler sa signification. Tous ces approches sont organiquement liées par le schéma de dualité³⁵. On peut l'appeler analyse en composantes explicatives³⁶ ou analyse des redondances³⁷. Ses limites sont la non-linéarité des relations (synthèse dans³⁸ qui justifie l'extension à l'analyse des correspondances), le nombre trop important d'explicatives qui justifie le passage à la régression PLS (synthèse dans³⁹).

Références

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

² 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.

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

- ⁴ Obadia, J. (1978) L'analyse en composantes explicatives. *Revue de Statistique Appliquée* : 24, 4, 5-28.
- ⁵ Wollenberg, A.L. (1977) Redundancy analysis, an alternative for canonical analysis. *Psychometrika* : 42, 2, 207-219.
- ⁶ 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.
- ⁷ Lebreton, J.D., Roux, M., Bacou, A.M. & Banco, G. (1990) Bioméco (Biométrie-Ecologie), version 3.9, software of staistical Ecology for PC.
- ⁸ Benzecri, J.P. & Coll. (1973) *L'analyse des données. II L'analyse des correspondances*. Bordas, Paris. 1-620.
- ⁹ Escoufier, Y. (1982) L'analyse des tableaux de contingence simples et multiples. *Metron* : 40, 53-77.
- ¹⁰ 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.
- ¹¹ 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.
- ¹² Blondel, J. (1986) *Biogéographie évolutive*. Masson, Paris, p. 73-78.
- ¹³ Blondel, J. & Farre, H. (1988) The convergent trajectories of bird communities along ecological successions in european forests. *Æcologia (Berlin)* : 75, 83-93.
- ¹⁴ Benzecri, J.P. (1983) Analyse de l'inertie intra-classe par l'analyse d'un tableau de correspondances. *Les Cahiers de l'Analyse des données* : 8, 3, 351-358.
- ¹⁵ Escoufier, B. & Drouet, D. (1983) Analyse des différences entre plusieurs tableaux de fréquence. *Les Cahiers de l'Analyse des Données* : 8, 4, 491-499.
- ¹⁶ Escoufier, Y. (1987) The duality diagramm : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute , Serie G .Springer Verlag, Berlin. 139-156.
- ¹⁷ 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.
- ¹⁸ Sabatier, R. (1987b) *Méthodes factorielles en analyse des données : approximations et prise en compte de variables concomitantes*. Thèse de doctorat d'état. Université de Montpellier. 1-224.
- ¹⁹ Ter_Braak, C.J.F. (1987b) *Unimodal models to relate species to environment*. Agricultural Mathematics Group, Box 100, NL-6700, AC Wageningen, The Netherlands. 1-152.

- ²⁰ Cazes, P., Chessel, D. & Doledec, S. (1988) L'analyse des correspondances internes d'un tableau partitionné : son usage en hydrobiologie. *Revue de Statistique Appliquée* : 36, 39-54.
- ²¹ Daudin, J.J. (1980) Partial association measures and an application to qualitative regression. *Biometrika* : 67, 581-590.
- ²² Ter_Braak, C.J.F. (1988) Partial Canonical Analysis. In : *Classification an related methods of data analysis*. Bock, H.H. (Ed.) North Holland. 551-558.
- ²³ Daudin, J.J. (1981) Analyse factorielle des dépendances partielles. *Revue de Statistique Appliquée* : 29, 2, 15-29.
- ²⁴ Yanai, H. (1986) Some generalizations of cerrespondence analysis in terms of projection operators. In : *Data Analysis and Informatics IV*. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 193-207.
- ²⁵ 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.
- ²⁶ Pontier, J., Jolicœur, P. & Pernin, M.O. (1987) Analyse canonique complète. *Statistique et Analyse des données* : 12, 1 et 2, 124-148.
- 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.
- ²⁷ Lebreton, J.D. & Yoccoz, N. (1987) Multivariate analysis of bird count data. *Acta Œcologica, Œcologia Generalis* : 8, 2, 125-144.
- ²⁸ Ter Braak, C.J.F. (1985) Correspondence analysis of incidence and abundance data : properties in terms of a unimodal reponse model. *Biometrics* : 41, 859-873.
- ²⁹ Kenkel, N.C. & Orloci, L. (1986) Applying metric and nonmetric multidimensionnal scaling to ecological studies : some new results. *Ecology* : 67, 919-928.
- ³⁰ Yanai, H. (1986) Some generalizations of cerrespondence analysis in terms of projection operators. In : *Data Analysis and Informatics IV*. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 193-207.
- ³¹ Ter Braak, C.J.F. (1987-1992) CANOCO - a FORTRAN program for Canonical Commnity Ordination. Microcomputer Power, Ithaca, New-York, USA.
- ³² Stewart, D.K. & Love, W.A. (1968) A General Canonical Correlation Index. *Psychological Bulletin* : 70, 160-163.
- ³³ Ter_Braak, C.J.F. (1990) Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika* : 55, 519-531.
- ³⁴ Kazi-Aoual, F., Hitier, S., Sabatier, R. & Lebreton, J.D. (1994) Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* : (In press).

- ³⁵ Sabatier, R. (1983) *Approximations d'un tableau de données. Application à la réconstitution des paléoclimats*. Thèse de 3^e cycle, Université de Montpellier. 1-184.
- ³⁶ Obadia, J. (1978) L'analyse en composantes explicatives. *Revue de Statistique Appliquée* : 24, 4, 5-28.
- ³⁷ Wollenberg, A.L. (1977) Redundancy analysis, an alternative for canonical analysis. *Psychometrika* : 42, 2, 207-219.
- ³⁸ Ter_Braak, C.J.F. (1987b) *Unimodal models to relate species to environment*. Agricultural Mathematics Group, Box 100, NL-6700, AC Wageningen, The Netherlands. 1-152.
- ³⁹ Höskuldsson, A. (1988) PLS regression methods. *Journal of Chemometrics* : 2, 211-228.