

L'analyse triadique partielle

Résumé

La fiche décrit le mode d'emploi de l'analyse triadique partielle. La méthode, qu'on peut appeler STATIS sur les tableaux, par opposition à STATIS sur les opérateurs, permet de faire une analyse moyenne de plusieurs ACP portant sur les mêmes individus et les mêmes variables. Pour plusieurs AFC, dans les mêmes conditions, le problème des pondérations est résolu par l'AFC moyenne de Foucart (1978, Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des données* : 2, 67-84). Les données de Blondel et Farré (1988, The convergent trajectories of bird communities along ecological successions in european forests. *Œcologia (Berlin)* : 75, 83-93.) illustrent avec précision les questions qu'on peut ainsi aborder dans un cube de données.

Plan

1 — STATIS sur les X : moyenne de structures.....	2
1.1 — La représentation des données : le module Curves.....	4
1.2 — L'option STATIS : Table averaging	7
2 — Moyennes d'analyse des correspondances.....	12
2.1 — Deux questions posées Blondel et Farré.....	13
2.2 — Les propositions de Foucart (1978).....	16
2.3 — La variabilité de l'expression d'une structure.....	22
Références	27

D. Chessel, J. Thioulouse & M. Simier

1 — STATIS sur les X : moyenne de structures

On considère un plan d'observations spatio-temporel complet décrit sur la carte Méaudret :

The screenshot shows a software window titled "Data 001 Méaudret (24 relevés/10 variables)". On the left, there is a list of 24 rows of data, each containing 10 numerical values. On the right, there is a menu for selecting variables. The menu lists the following variables with their corresponding codes:

- Temp
- Débit
- pH
- Condu
- Oxyg
- DbO5
- Oxyd
- Ammo
- Nitra
- Phos

Below the menu, a legend titled "Code des variables" lists the full names for each code:

- 1- Température
- 2- Débit (l/s)
- 3- pH
- 4- Conductivité (mmho/cm)
- 5- Oxygène (% saturation)
- 6- DBO5 (mg/l oxygène)
- 7- Oxydabilité (idem)
- 8- Ammoniaque (mg/l)
- 9- Nitrates (mg/l)
- 10- Orthophosphates (mg/l)

Récupérer les fichiers Mil (24 lignes-relevés et 10 colonnes-variables) et Code_Var. Enlever la colonne 5 et les lignes 21 à 24 de Mil (fichier M 20-9), ainsi que les lignes 21 à 24 de CodeVar. Vérifier qu'on obtient les données consignées dans le tableau 1 de l'article en annexe. La variable éliminée prend une valeur constante dans un bloc de lignes et on ne conserve que les cinq stations sur la même rivière. La présentation est alors de la forme 5 tableaux (stations) portant sur les mêmes individus (dates) et les mêmes variables (9 variables).

On considère que les données sont formées de 4 tableaux portant sur les mêmes individus (5 stations) et les mêmes descripteurs (9 variables). Tous les multi-tableaux ayant la même forme dans ADE-4, il convient de s'y ramener. Récupérer sur la carte Méaudret+1 le fichier Plan, enlever les 4 dernières lignes (fichier P 20 lignes-2 colonnes) et trier le fichier M avec la variable 2 (Numéro de la date) de P (après passage dans l'option ReadCateg) avec l'option CateRowSort de FilesUtil :

The screenshot shows the "CateRowSort" dialog box. It contains the following fields and options:

- Input file:** M 20 9
- .cat file:** P.cat
- Column number for sorting:** 2
- Output file:** A
- Buttons:** Quit, Ok

Vérifier que le contenu du fichier A est celui du tableau 1. Préparer un fichier binaire provisoire Blo qui contient 4 fois la valeur 5 sur une seule colonne. Utiliser l'option InitKTab de KTabUtil :

FilesUtil	CurveClass	
	CurveModels	RLQ
CategVar		
MatAlg	Scatters	KTabUtil
	ScatterClass	KTA
PCA HTA	ADEScatters	STATIS

Les fichiers créés A!.ktp, A!TLI et ses associés, A!TCc et ses associés, A!T4a et ses associés sont décrits dans la fiche Statis 1 (page 21).

1	10.00	41.00	8.50	295.00	2.30	1.40	0.12	3.40	0.11
2	11.00	158.00	8.30	315.00	7.60	3.30	2.85	2.70	1.50
3	11.00	198.00	8.50	290.00	3.30	1.50	0.40	4.00	0.10
4	12.00	280.00	8.60	290.00	3.50	1.50	0.45	4.00	0.73
5	13.00	322.00	8.50	285.00	3.60	1.60	0.48	4.60	0.84
6	13.00	62.00	8.30	325.00	2.30	1.80	0.11	3.00	0.13
7	13.00	80.00	7.60	380.00	21.00	5.70	9.80	0.80	3.65
8	15.00	100.00	7.80	385.00	15.00	2.50	7.90	7.70	4.50
9	16.00	140.00	8.00	360.00	12.00	2.60	4.90	8.40	3.45
10	15.00	160.00	8.40	345.00	1.70	1.90	0.22	10.00	1.74
11	1.00	25.00	8.40	315.00	1.60	0.50	0.07	6.40	0.03
12	3.00	63.00	8.00	425.00	36.00	8.00	12.50	2.20	6.50
13	2.00	79.00	8.10	350.00	7.10	1.90	2.70	13.20	3.70
14	3.00	85.00	8.30	330.00	2.00	1.40	0.42	12.00	1.60
15	2.00	72.00	8.60	305.00	1.60	0.90	0.10	9.50	1.25
16	3.00	118.00	8.00	325.00	1.60	1.20	0.17	1.80	0.19
17	3.00	252.00	8.30	360.00	9.50	2.90	2.52	4.60	1.60
18	3.00	315.00	8.30	370.00	8.70	2.80	2.80	4.80	2.85
19	3.00	498.00	8.30	330.00	4.80	1.60	1.04	4.40	0.82
20	2.00	390.00	8.20	330.00	1.70	1.20	0.56	5.00	0.60

Tableau 1 : Tableau de données formé de 4 blocs (dates) de 5 lignes (stations) et 9 colonnes (variables).

Normaliser les données par colonne et par sous-tableau :

Les fichiers A.ktta, A.ktpl et A.ktpc forment un triplet statistique standard. Représenter les données ainsi préparées avec le modules Curves.

1.1 — La représentation des données : le module Curves

Les abscisses sont les dates, les variables sont les colonnes de A.ktta, les étiquettes des variables sont dans Code_Var :

Lines	
X file (default = 1, 2, 3, ..., n)	<input type="text" value="A!TLI"/> 20 3
X file column number (default = 1)	<input type="text" value="3"/>
Y file (no default)	<input type="text" value="A.ktta"/> 20 9
Cumulated data (1=yes, 2=no)	<input type="text"/>
Variable label file (or #)	<input type="text" value="Code_Var"/>
Draw curves (1=yes, 2=no)	<input type="text"/>
Draw points (1=yes, 2=no)	<input type="text"/>
Row label file (or #)	<input type="text"/>
Number of curves by window	<input type="text"/>

Windows	Options
Close	<input type="checkbox"/>
Graphics	<input type="checkbox"/>
File Selection	<input type="checkbox"/>
Min. & Max.	<input type="checkbox"/>
Row & Col. sel	<input type="checkbox"/>

Les lignes sont triées par dates pour une représentation multifenêtrée :

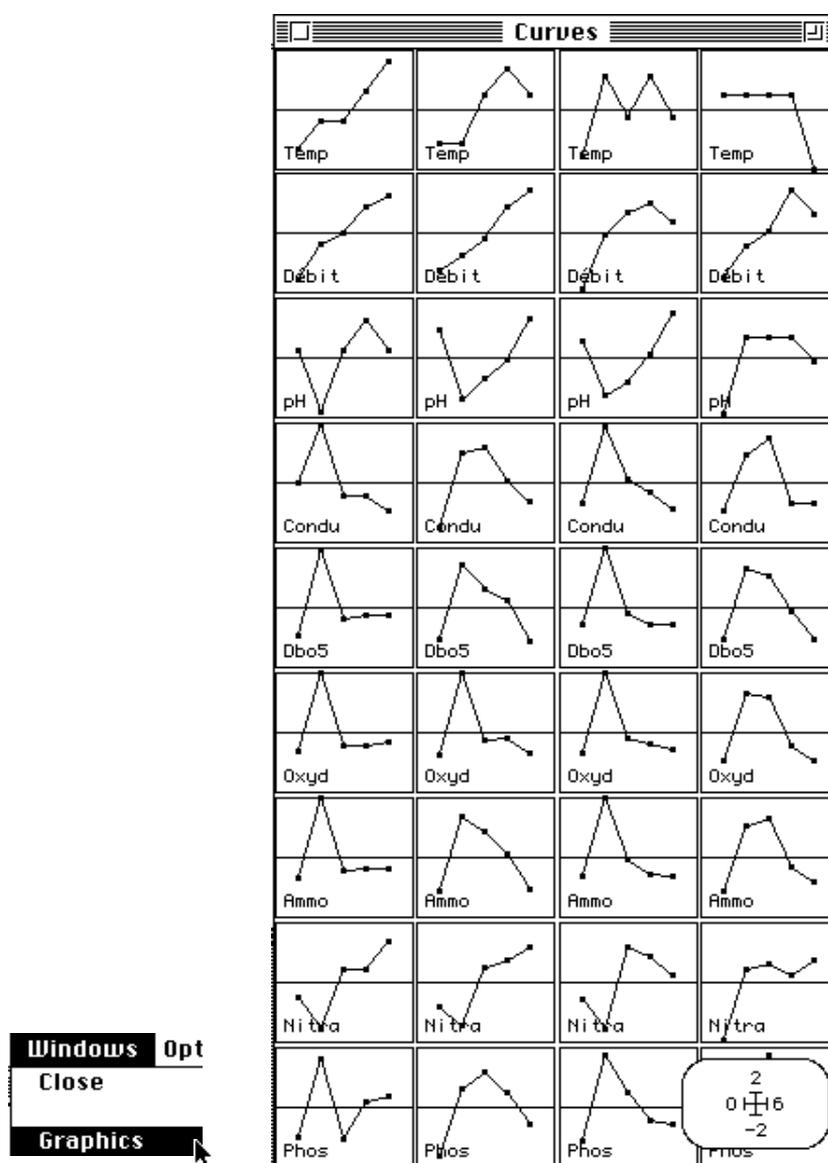
Row & col. selection	
Col. selection:	<input type="text"/>
Row selection method:	<input checked="" type="radio"/> File <input type="radio"/> Keyboard
Row selection file (.cat):	<input type="text" value="A!TLI.cat"/>
Selection col. number:	<input type="text" value="1"/>

Les bornes et le multifenêtrage sont ajustés à la nature de la structure des données :

Min/Max					
Min. abscissa:	<input type="text" value="0"/>	<input type="checkbox"/>	Horiz. graphs:	<input type="text" value="4"/>	<input type="checkbox"/>
Max. abscissa:	<input type="text" value="6"/>	<input type="checkbox"/>	Vert. graphs:	<input type="text" value="9"/>	<input type="checkbox"/>
Min. ordinate:	<input type="text" value="-2"/>	<input type="checkbox"/>	Nb. grad. X:	<input type="text" value="1"/>	
Max. ordinate:	<input type="text" value="2"/>	<input type="checkbox"/>	Nb. grad. Y:	<input type="text" value="1"/>	
Window height:	<input type="text" value="600"/>		G factor:	<input type="text"/>	
Window width:	<input type="text" value="300"/>		<input type="checkbox"/> Square drawing		
<input type="button" value="Draw"/>		<input checked="" type="checkbox"/> Draw frame			
		<input checked="" type="checkbox"/> Scale box			

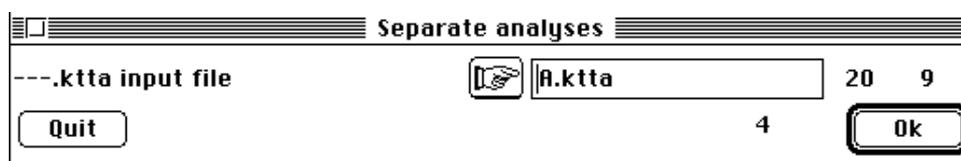
Windows	Options
Close	<input type="checkbox"/> %W
Graphics	<input type="checkbox"/> %G
File Selection	<input type="checkbox"/> %F
Min. & Max.	<input type="checkbox"/> %M
Row & Col. selection	<input type="checkbox"/> %R

L'affichage de la fenêtre de graphique est exactement celui de la figure 1 de l'annexe :

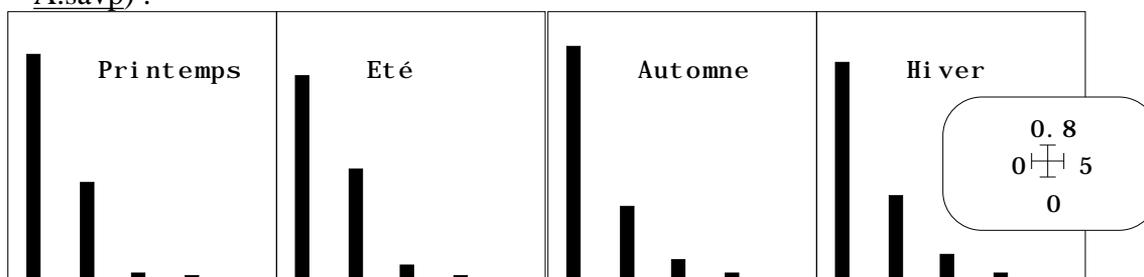


Le rôle du graphique dépasse le simple service rendu. Il exprime le point de vue utilisé. En abscisse le gradient amont-aval (5 stations) vu lors de quatre visite (les blocs de colonnes de fenêtres Juin, Août, Novembre, Février) à l'aide de 9 descripteurs (les blocs de lignes de fenêtres). L'importance de ce rôle a récemment été souligné par F. Lescourret¹.

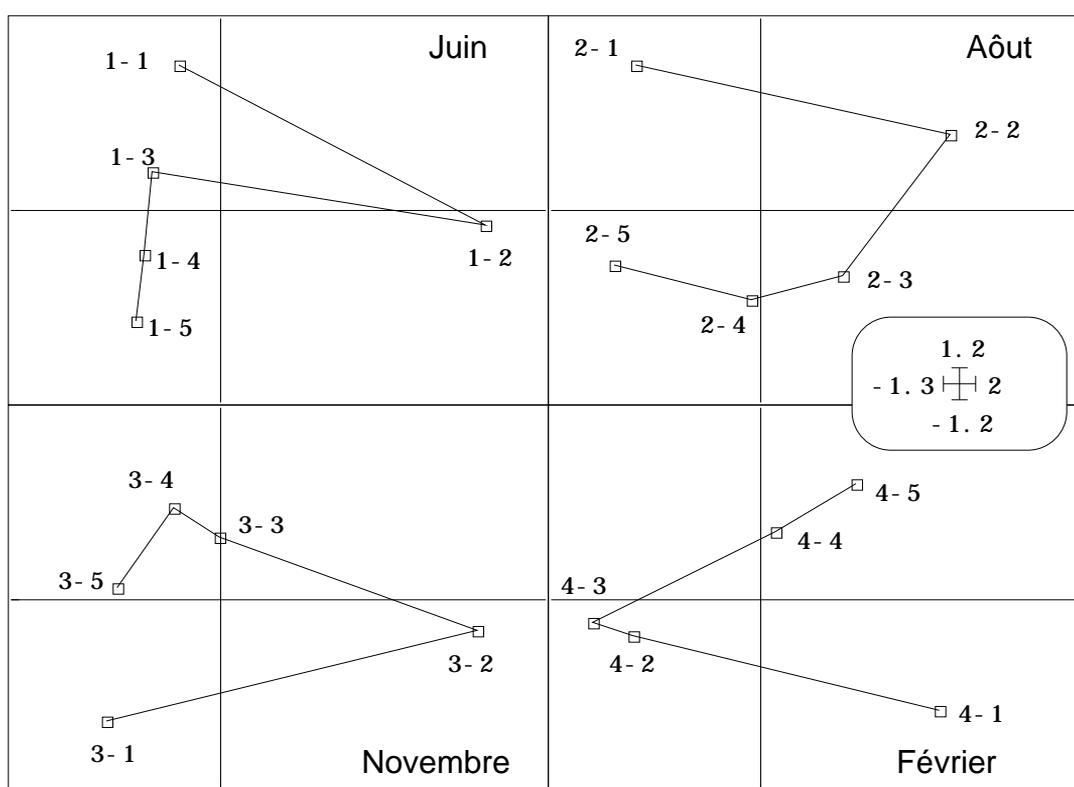
L'annexe décrit la procédure de calcul de l'analyse triadique partielle accessible avec un simple programme d'ACP et plusieurs manipulations relativement lourdes. Dans ADE-4, le tout est exécuté simplement. On exécutera au préalable les quatre ACP normées séparées avec KTA :



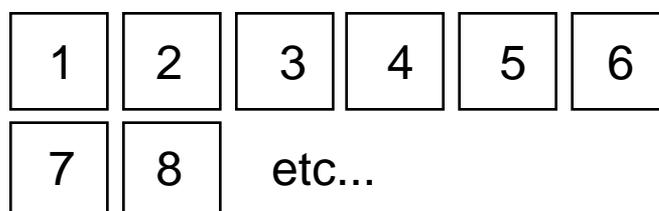
Chaque analyse séparée fait une description multivariée du gradient amont-aval. Les valeurs propres indiquent la nécessité d'une représentation plane (Curves sur le fichier A.savp) :



Les cartes factorielles des lignes (Trajectoires dans Scatters sur A.saTLLi, triées par dates sur la colonne 1 de A!TLLi, étiquetées par A!saTLLi.label) ne sont pas coordonnées (figure 5 de l'annexe) :



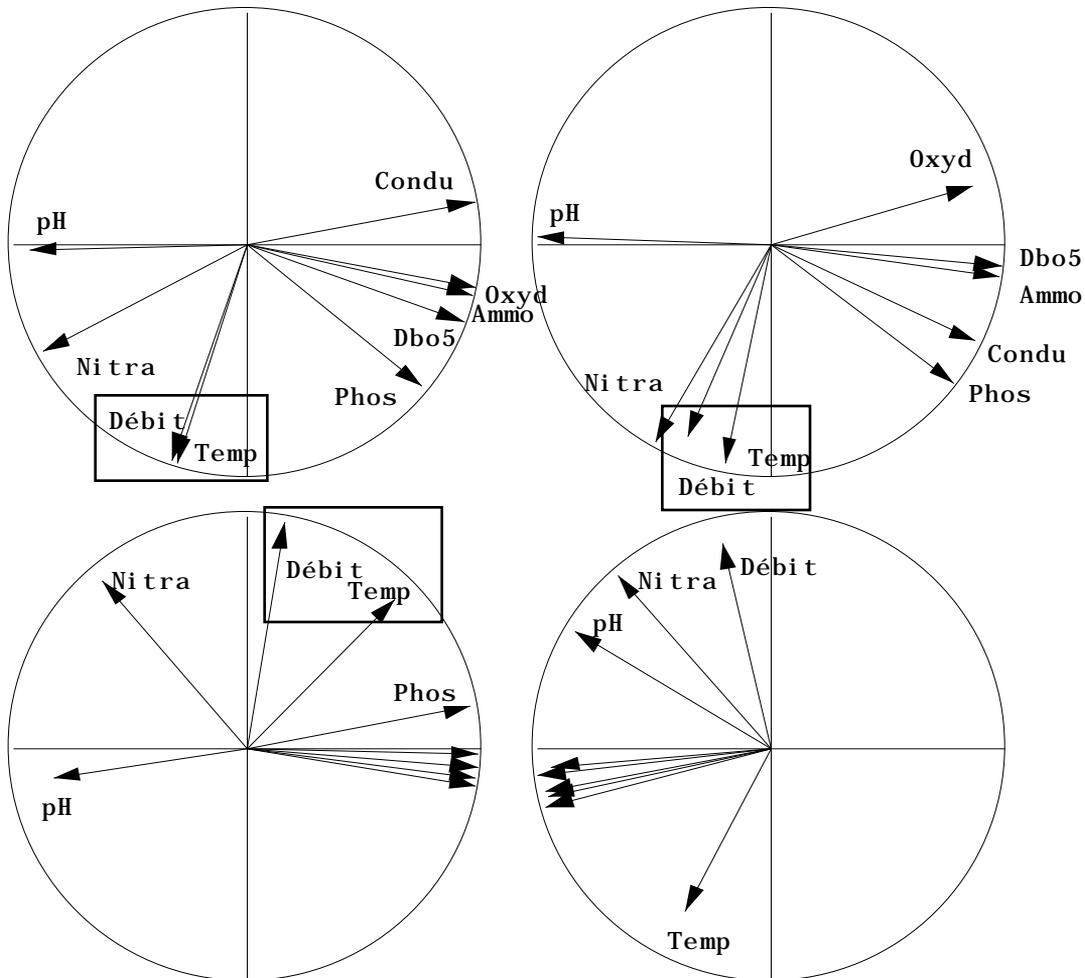
Noter que lorsque les fenêtres ne sont pas étiquetées, tous les modules d'ADE-4 utilise le multifenêtrage ligne par ligne :



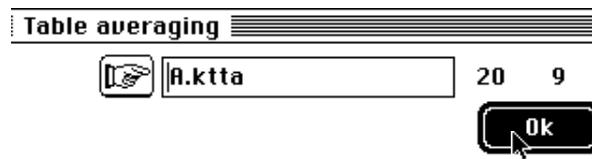
Dupliquer quatre fois le contenu du fichier d'étiquettes Code_Var dans un fichier d'étiquettes Code_Var.TCc pour simplifier la lecture des cartes des colonnes. Utiliser Scatters sur A.saTCco, triées par dates sur la colonne 1 de A!TCc étiquetées par Code_Var.TCc. On obtient le reste du contenu de la figure 5 de l'annexe et une illustration précise de la nécessité de coordonner les quatre ACP, ce qui est la fonction

de l'analyse triadique partielle (ATP) appelée aussi dans la terminologie de l'école de Montpellier STATIS sur les X.

L'annexe 5 parle d'analyse triadique mais la réponse à cet article de P. Kroonenberg² démontre que l'appellation est impropre et doit être réservée à l'ACP 3-modes.



1.2 — L'option STATIS : Table averaging



Input file A
Number of rows: 20, columns: 9

L'option Table averaging est conçu comme le module Operator averaging. Il s'agit d'abord de typologie moyenne ou compromis. La différence essentielle est que, dans le cas présent, deux tableaux sont directement comparable, puisqu'ils portent sur les mêmes individus (stations) et les mêmes variables (descripteurs). Il n'y a pas lieu de passer par la comparaison d'opérateurs si une seule dimension est en commun. Notons n le nombre de lignes et p le nombre de colonnes de chacune des analyses séparées, \mathbf{D}_n et \mathbf{D}_p les normes associées. On peut donc calculer un produit scalaire entre tableaux :

$$\text{Covv}(\mathbf{X}_k, \mathbf{X}_j) = \text{Trace}(\mathbf{X}_k^t \mathbf{D}_n \mathbf{X}_j \mathbf{D}_p) = \text{Trace}(\mathbf{X}_j^t \mathbf{D}_n \mathbf{X}_k \mathbf{D}_p)$$

D'où le coefficient de corrélation entre deux tableaux :

$$Rv(\mathbf{X}_k, \mathbf{X}_j) = \frac{\text{Covv}(\mathbf{X}_k, \mathbf{X}_j)}{\sqrt{Vav(\mathbf{X}_k)} \sqrt{Vav(\mathbf{X}_j)}}$$

```
----- Corrélation matrix -----
[ 1] 1000
[ 2] 693 1000
[ 3] 789 767 1000
[ 4] 283 534 479 1000
-----
```

Les RV sont élevés, mais la structure du tableau 4 est manifestement la plus éloignée du groupe des 3 autres.

```
File A.xa+RV contains cosinus between tables
It has 4 rows and 4 columns
File A.xa+CV contains inner products between tables
It has 4 rows and 4 columns
```

La matrice diagonalisée est la matrice des Covv, car on suppose que pour des tableaux de mêmes individus et mêmes variables présentent des inerties de même ordre de grandeur (sinon, c'est un fait qui rentre en ligne de compte). Ici, c'est d'ailleurs la même chose, car les variances vectorielles sont les inerties des nuages, donc les moyennes des variances des variables (chaque colonne a un poids uniforme) et valent l'unité, car les variables sont normalisées.

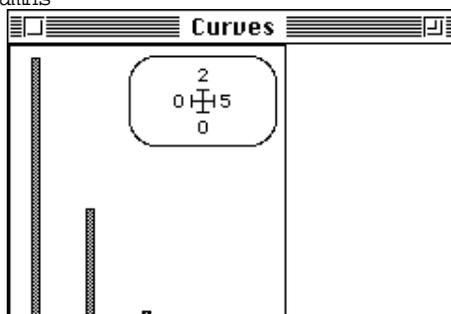
Comme pour STATIS sur opérateurs, nous avons décidé d'éliminer de la discussion les vecteurs propres de la matrice des covariances vectorielles (Covv) de rang supérieur à 1, pour souligner clairement qu'on cherche ici une structure moyenne, et qu'on cherchera ailleurs à décrire l'évolution autour de cette moyenne. Les figures 2 et 3 de l'annexe ne sont donc pas reproductibles. C'est un choix qui devrait simplifier et éclaircir l'usage de STATIS pour des utilisateurs occasionnels. On s'est contenté de reproduire les valeurs propres :

```
Compromise with inner products between tables
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +2.8121E+00 +0.7030 +0.7030 | 02  +7.5415E-01 +0.1885 +0.8916 |
03  +2.5368E-01 +0.0634 +0.9550 | 04  +1.8003E-01 +0.0450 +1.0000 |
```

Cette diagonalisation a pour fonction d'attribuer à chaque tableau un poids (Cf. weights ci-dessous). Le poids attribués au tableau 4 est moindre que celui des trois autres. La combinaison des tableaux utilisant ces poids est un nouveau tableau de synthèse combinant les tableaux initiaux à proportion de leurs apports à la description de la structure commune dite compromis. Ce nouveau tableau, dont le contenu importe peu (ce sont des combinaisons des valeurs des tableaux initiaux avec des coefficients tous positifs), a pour fonction de définir des axes et des composantes, donc des vecteurs de \mathbb{R}^n et de \mathbb{R}^p , qui exprime la structure compromis. Le programme est donc consacré essentiellement à une recherche d'un compromis inter-tableaux et à l'étude de la structure de ce compromis. Les valeurs propres de ce compromis sont :

```
Num. Eigenval.  R.Iner.  R.Sum  | Num. Eigenval.  R.Iner.  R.Sum  |
01  +1.9112E+00 +0.6796 +0.6796 | 02  +8.1084E-01 +0.2883 +0.9680 |
03  +6.7772E-02 +0.0241 +0.9921 | 04  +2.2308E-02 +0.0079 +1.0000 |
05  +0.0000E+00 +0.0000 +1.0000
```

File A.xa+vp contains the eigenvalues of compromise analysis
It has 9 rows and 1 columns



Les coordonnées des colonnes et des lignes du compromis sont conservées :

File A.xa+co contains column scores (Norm = 1 for column weights)
in the analysis of the compromise
It has 9 rows and 2 columns

File :A.xa+co
-----Minimum/Maximum:
Col.: 1 Mini = -1.3409 Maxi = 0.86289
Col.: 2 Mini = -2.037 Maxi = -0.01506

File A.xa+li contains row scores (Norm = 1 for row weights)
in the analysis of the compromise
It has 5 rows and 2 columns

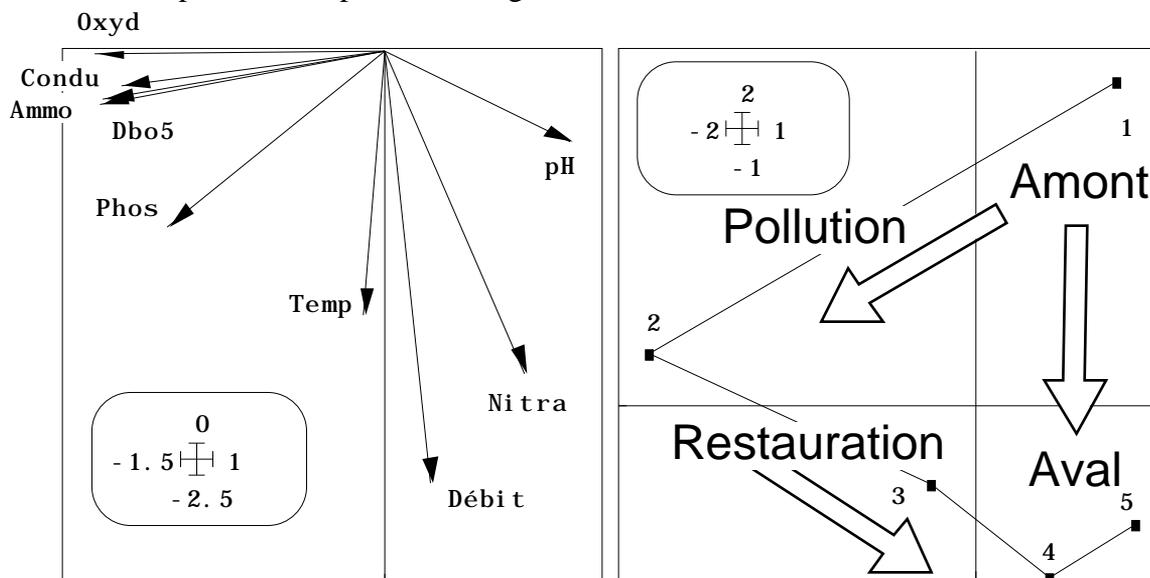
File :A.xa+li
-----Minimum/Maximum:
Col.: 1 Mini = -1.834 Maxi = 0.89433
Col.: 2 Mini = -0.9731 Maxi = 1.8175

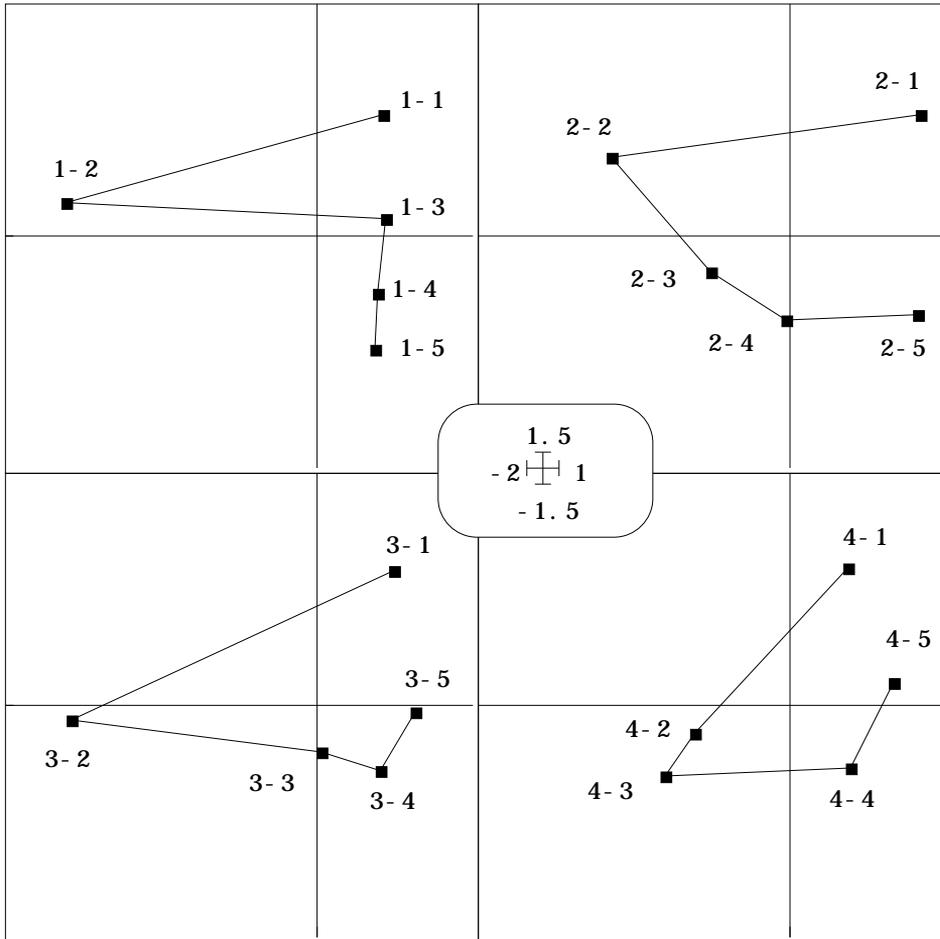
On notera qu'il s'agit de scores de norme unité pour les pondérations communes. Utiliser Scatters pour reproduire la figure 4 de l'annexe (ci-dessus). Ces vecteurs (dont on a représenté les composantes) définissent des plans sur lesquels peuvent être projeté lignes et colonnes de chacun des tableaux : il s'agit alors de véritables projections.

File A.xaTl1 contains standard row scores with lambda norm
It has 20 rows and 2 columns
It is to be used with --Tl1.label and --Tl1.cat files
File :A.xaTl1

-----Minimum/Maximum:
Col.: 1 Mini = -1.6104 Maxi = 0.83894
Col.: 2 Mini = -0.7287 Maxi = 0.88372

Scatters permet de reproduire la figure 6 de l'annexe :

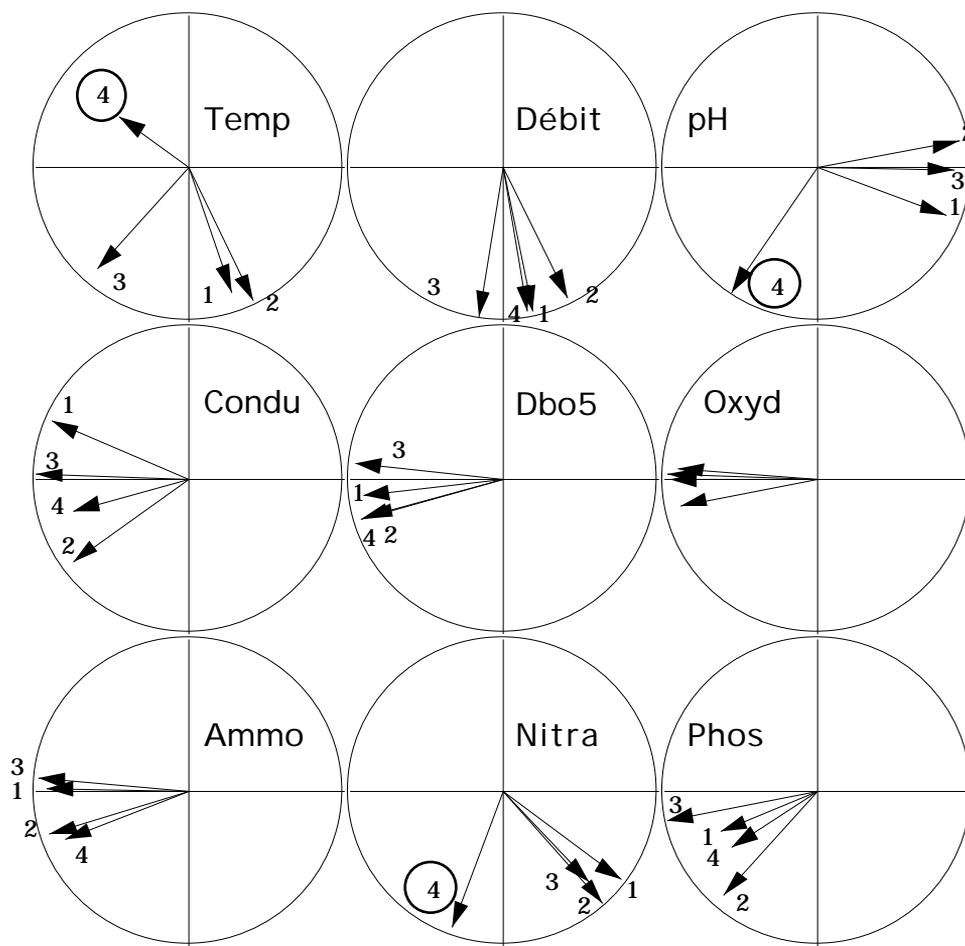




La figure 7 utilise CurveClass :

		Min. abscissa: 0 Max. abscissa: 6 Min. ordinate: -1.7 Max. ordinate: 1 Window height: 300 Window width: 300
Lines		
X file (default = 1, 2, 3, ..., n)		A!TLI 20
X file column number (default = 1)		3
Y file (no default)		A.xaTLI 20
Categories file (.cat)		A!TLI.cat
Categories file column number		1

Pour les variables :



File A.xaTCO contains columns scores issued from each table by
 It has 36 rows and 2 columns
 It is to be used with --Tcc.label and --Tcc.cat files
 File :A.xaTCO
 -----Minimum/Maximum:
 Col.: 1 Mini = -0.99704 Maxi = 0.91879
 Col.: 2 Mini = -0.97939 Maxi = 0.37549

La figure 8 est alors aisément reproductible (ci-dessus). Le programme s'achève par l'édition des trois paramètres résumant la position d'un tableau par rapport au compromis. Les normes (inertie totale) sont égales à l'unité, les poids et les Cosinus carrés soulignent la ressemblance de trois des quatre tableaux et l'originalité relative des mesures d'hiver (Février : tableau 4).

Typological value indices
 Rows = row number of each table
 Weights = Weights of tables in the compromise
 NS norm2 = Squared norm (Hilbert-Schmidt) of tables
 Cos2 = Squared cosinus between table and approximed compromise

Number	Rows	Weights	NS norm2	Cos2
1	5	5.067e-01	1.000e+00	0.743
2	5	5.404e-01	1.000e+00	0.779
3	5	5.510e-01	1.000e+00	0.815
4	5	3.843e-01	1.000e+00	0.386

L'analyse triadique partielle est une manière simple de faire plusieurs ACP simultanées. C'est encore vraie pour plusieurs analyses des correspondances. La richesse problématique associée à l'acquisition de données cubiques mérite à cette

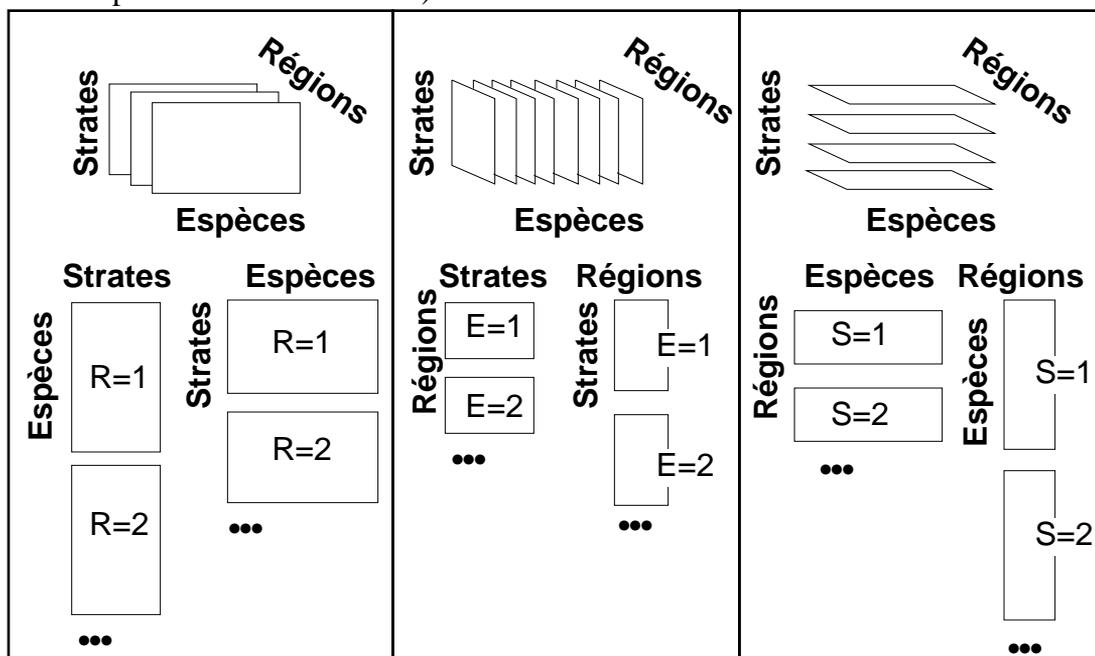
occasion d'être soulignée. Il est question ici de montrer que la neutralité sous-entendue aux pratiques de l'analyse des données est une idée reçue.

2 — Moyennes d'analyse des correspondances

Le jeu de données utilisées est proposé par J. Blondel et H. Farré³. Il illustre un des problèmes fondamentaux de l'écologie factorielle. En confrontant un cortège faunistique à un paramètre de structure de l'habitat, on définit la notion de profil écologique ou de niche écologique. Quand on recommence la même opération à une autre date ou dans une autre région la relation binaire faune-milieu devient une relation ternaire faune-milieu-région. Les données sont dans la carte Successions de la pile ADE•Data :

Data 0851 Successions (79-24) - J. Blondel & H. Farré 1988	
0/0/0/0/0/0/36/16/6/0/0/0/5/6/0/0/6/0/8/16/8/0/0/0/	
0/	PHPH Rougequeue à front blanc
0/	PAPA Mésange nonette
0/	Po1 0/0/0/0/0/
0/	Po2 0/0/33/39/
0/	Po3 0/0/19/4/
0/	Po4 33/3/0/0/
0/	Po5 0/0/24/0/
0/	Po6 0/0/1/0/0/
0/	Bu1 0/0/1/0/0/0/0/0/0/
0/	Bu2 0/0/2/0/0/0/0/0/0/
0/	Bu3 2/0/3/18/36/2/1/0/
0/	Bu4 0/
0/	Bu5 0/
0/	Bu6 1/
0/	Pr1 2/
0/	Pr2 0/
0/	Pr3 0/
0/	Pr4 0/
0/	Pr5 0/
0/	Pr6 0/0/0/0/0/0/0/0/0/

Il s'agit de mesurer la variabilité du cortège avifaunistique entre 4 régions (Pologne, Bourgogne, Provence et Corse), le long du gradient de fermeture de la végétation vu par six strates d'échantillonnage (1- végétation buissonnante basse (hauteur < 1 m) à 6- forêts de plus de 20 m de hauteur).



2.1 — Deux questions posées Blondel et Farré

Récupérer les fichiers PoBuPrCo.txt (79 espèces - 24 séries de relevés), Code Esp_BF88 (79 chaînes de 4 caractères codant les espèces), Code Rel_BF88 (24 chaînes de 3 caractères codant les séries), Noms Esp_BF88 (79 chaînes de caractères). L'abondance d'une espèce dépend de la région et généralement de l'architecture de la végétation : ce n'est plus à prouver ou à décrire. Le profil d'une espèce dépend-il de la région ? C'est déjà moins simple. L'hétérogénéité de l'avifaune entre régions dépend-il de la nature de la végétation ? Voilà une autre question.

La relation entre cortège faunistique et architecture de la végétation définit dans chaque région une structure de tableau donc une analyse. Cette analyse est une analyse des correspondances, sans contestation (typologie d'espèces par leur courbe de réponse inter-strates, typologie de strates par leur profil spécifique). Dans une région, le tableau est de type espèces-strates. Il y a quatre tableaux de ce type, donc une structure moyenne et des divergences régionales autour de cette structure.

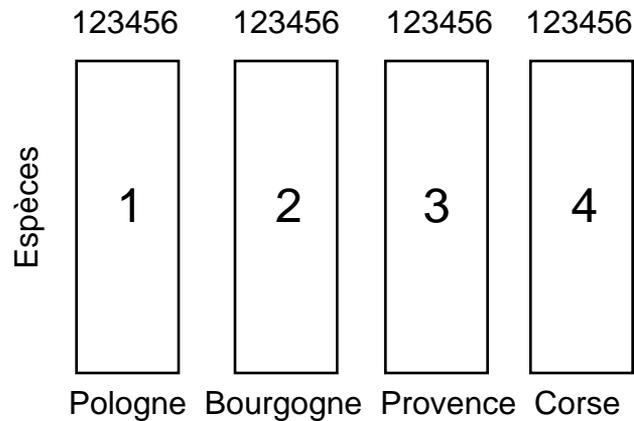
La relation entre cortège faunistique et zones biogéographiques définit dans pour chaque strate de végétation une structure de tableau donc une analyse. Cette analyse est une analyse des correspondances (typologie d'espèces par leur distribution géographique, typologie de régions par leur contenu spécifique). Dans une strate, le tableau est de type espèces-régions. Il y a six tableaux de ce type, donc une structure moyenne et des divergences, fonction de la végétation, autour de cette structure.

L'abondance d'une espèce dans chaque région et chaque strate définit un modèle de répartition, demandant l'analyse d'un tableau homogène par une analyse simplement centrée. Il y a 79 tableaux de ce type. Que signifierait la notion de modèle moyen ? Le plus simple est de se référer à l'espèce sans signification écologique, uniformément présente dans chaque strate et chaque région. On peut penser à une typologie de modèles (courbes de réponse bivariées).

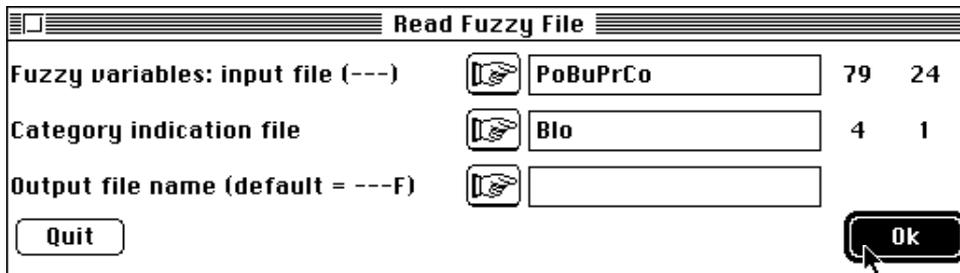
Ces indications sont incitatives à une réflexion préliminaire dans l'étude des cubes de données. Il convient, en effet, de garder son calme, tant un cube de données peut supporter potentiellement d'approches statistiques. Comme nous allons le voir, l'intention peut conduire à des résultats radicalement différents, sans que la validité des opérations soit mise en cause. La première chose à faire est de distinguer ce qui relève de l'observation de ce qui relève de l'organisation de l'information. Ici, nous avons deux effets fixes, à savoir la végétation et la région. On aurait pu étudier un autre facteur écologique, par exemple l'altitude, et un autre corpus biogéographique, par exemple plusieurs massifs montagneux. Deux des arêtes du cube de données sont l'expression de l'intention de l'observation. La troisième, au contraire n'est pas maîtrisée. C'est la liste des espèces observables ou observées. Son contenu est fourni par les écosystèmes étudiés.

Lorsque les trois marges sont des effets fixes, par exemple mesure d'un paramètre x dans 4 types de végétation, dans 3 classes d'altitude et dans 5 régions, les données forment un cube vrai. On peut désirer modéliser la variable x en fonction des 3 facteurs contrôlés, voire étudier les interactions ternaires : c'est le domaine des analyses à trois modes et plus. Consulter les ouvrages de référence⁴. ADE-4 ne contient aucune proposition dans ce domaine. Lorsque deux marges sont des effets fixes, il y a deux grands types d'objectifs. Le premier est celui des variables explicatives : construire un modèle de l'effet strate-régions pour chacune des espèces (effet simple A ou B, effet additif A+B, effet partiel A sachant B ou B sachant A, ...). Le second est celui de la comparaison de structures, c'est-à-dire de l'effet d'un facteur sur la structure engendré par l'autre.

Nous avons ici deux questions de ce type.



Le fichier binaire PoBoPrCo contient 79 espèces-lignes et 4 blocs (régions) de 6 colonnes (strates). Les valeurs sont des densités de couples nicheurs. Constituer un fichier binaire Blo contenant sur une colonne la valeur 6 quatre fois. Préparer l'ACM floue de ce fichier :



```
Input file: PoBuPrCo
Row number: 79, column number: 24
Number of modalities for each variable: file Blo
Row number: 4
Missing data: 154
```

Les données manquantes sont définies par les espèces absentes dans une région toute entière.

```
Variable number 1 has 6 categories
-----
[  1] Category:   1   Freq.:   0.0302
[  2] Category:   2   Freq.:   0.0646
[  3] Category:   3   Freq.:   0.234
[  4] Category:   4   Freq.:   0.116
[  5] Category:   5   Freq.:    0.21
[  6] Category:   6   Freq.:   0.345
Missing data Num:  40 Freq.:   0.506
```

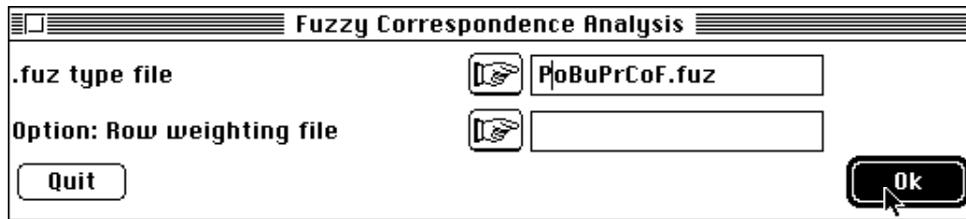
40 espèces sont absentes en Pologne. La région compte 39 espèces (Blondel & Farré, op. cit., tableau 2 page 85).

```
Variable number 2 has 6 categories
-----
Missing data Num:   34 Freq.:   0.43
```

La région Bourgogne compte 45 espèces. Les régions Provence et Corse compte 39 espèces.

```
Variable number 3 has 6 categories
Missing data Num:   40 Freq.:   0.506
Missing data Num:   40 Freq.:   0.506
-----
Output file: PoBuPrCoF
```

Exécuter l'analyse et garder trois facteurs :



Passer du tableau flou au multi-tableau :

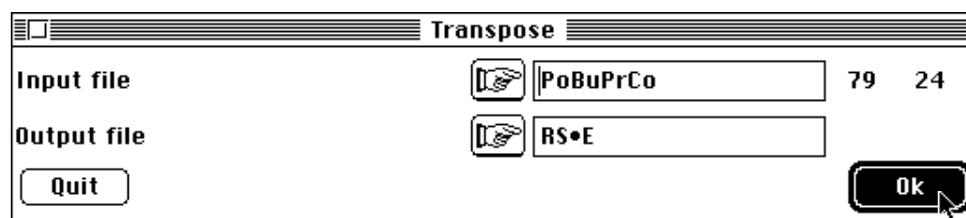


Exécuter STATIS sur les X :

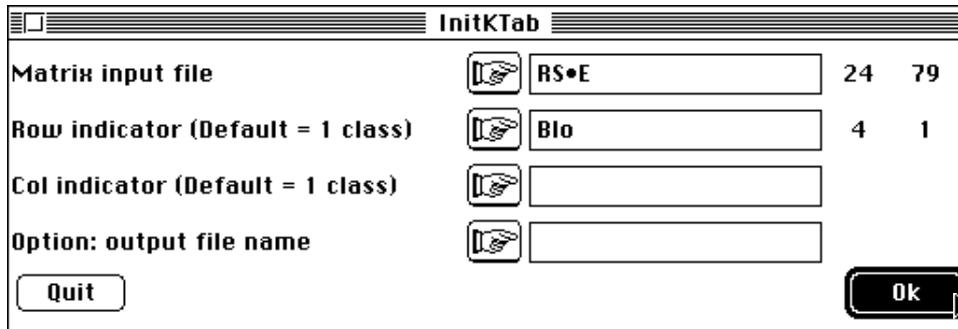


On n'ira pas plus loin. En effet le codage flou permet de ramener les pondérations des tableaux à une valeur commune pour ce qui est des colonnes, mais chaque tableau a conservé une pondération ligne qui lui est propre. C'est un problème très général. L'AFC est très particulière dans la famille des méthodes à un tableau, en ce sens qu'elle calcule les poids avec les données. On peut lui imposer une des marges : au prix d'une modification du tableau on garde une AFC en recalculant l'autre. Là s'arrête le possible. Si K tableaux ont les mêmes lignes et les mêmes colonnes, on ne sait pas faire K AFC simultanées ayant les mêmes pondérations. On sait encore moins définir une AFC moyenne optimale qui aurait des pondérations identiques à celles des K tableaux de départ. Détruire tous les fichiers dont le nom commence par PoBuPrCoF avant de continuer.

Transposer le fichier PoBuPrCo :



Utiliser l'option InitKTab de KTabutil :



```

Option: InitKTab
Input file: RS•E
Row indicator file: Blo
Col indicator file:
Output ASCII file: RS•E!.ktp
-----
Qualitative variables file: RS•E!TLl
Number of rows: 24, variables: 3, categories: 34
Auxiliary ASCII output file RS•E!TLl.label: labelling file
-----
Qualitative variables file: RS•E!Tcc
Number of rows: 316, variables: 3, categories: 162
Auxiliary ASCII output file RS•E!Tcc.label: labelling file
-----
Qualitative variables file: RS•E!T4a
Number of rows: 16, variables: 2, categories: 8
Auxiliary ASCII output file RS•E!T4a.label: labelling file
-----

```

On possède quatre tableaux d’AFC 6 lignes -79 colonnes superposés. Le nom RS•E désigne Régions-Stations en lignes et Espèces en colonnes.

2.2 — Les propositions de Foucart (1978)

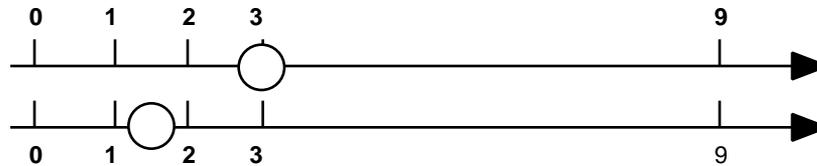
T. Foucart, qui a travaillé sur la méthode STATIS⁵, s’est posé très tôt la question qui est apparue au paragraphe précédent⁶. Il part de la constatation qu’on peut aussi bien concevoir une table de contingence comme un tableau d’ACP particulière que comme une matrice de covariance particulière. Mais dans un cas comme dans l’autre la question des pondérations interdit de généraliser STATIS (aussi bien sur les tableaux que sur les opérateurs). Il propose une opération qui n’a pas l’esthétique mathématique de STATIS mais qui est efficace. Il note :

Dans cet article, nous avons proposé des définitions de tendance et de structure susceptibles d’être utilisées dans l’étude des suites de tableaux de probabilités indexées par le temps. Si la technique simple d’analyse des évolutions des tendances repose effectivement sur la définition que nous en avons donnée, il n’en est malheureusement pas de même en ce qui concerne les techniques d’étude des évolutions de la structure : nous ne sommes pas partis des équivalences entre structures pour mettre au point les méthodes qui ont été décrites. Si ce manque de cohérence nuit à la qualité de notre exposé, il ne diminue en rien l’intérêt de ces équivalences et l’efficacité de ces méthodes.

La difficulté vient de la présentation classique de STATIS : interstructure, compromis, intrastructure. En termes d’un seul tableau, tout se passe comme si on disait matrice de corrélation, moyenne, variance. L’interstructure définit une typologie de structure, le compromis définit une structure moyenne et l’intrastructure représente la variabilité autour de la moyenne (centrage). C’est comme si on voulait définir l’analyse en composantes principales avant la moyenne. Nous avons souligné que STATIS définit

essentiellement la moyenne, et ce de façon élaborée. En effet, on calcule généralement une moyenne en définissant au préalable les poids utilisés. Dans STATIS on définit une moyenne en calculant les poids pour que cette moyenne soit la meilleur possible.

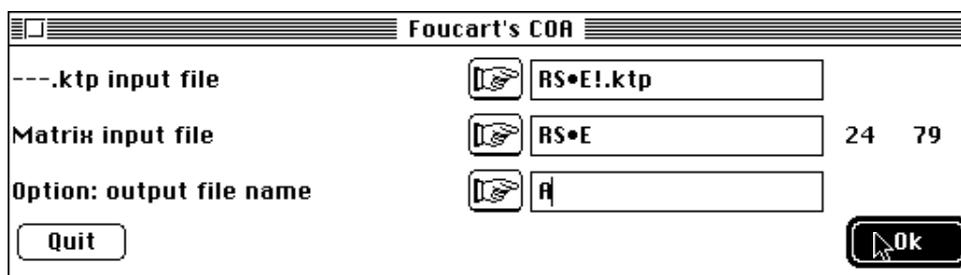
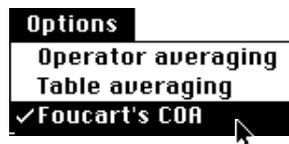
Cela conduit à éliminer le rôle des points douteux. Par exemple, la moyenne à pondération uniforme de 5 valeurs vaut 3, mais si on considère que la cinquième valeur est bizarre, on dit que cette moyenne vaut 1.5. On fait évidemment cette opération, dans STATIS, sur les structures et non sur les valeurs. On fait effectivement cette opération sur des valeurs en ACP non centrée.



Foucart propose simplement de faire le compromis en prenant une moyenne uniformément pondérée des tableaux. Soit K tableaux d'AFC. Le $k^{\text{ème}}$ tableau a , comme les autres, I lignes et J colonnes. Sont terme général est x_{ij}^k et la somme de toute les valeurs est $x_{..}^k$.

Le tableau de fréquence associé est $\mathbf{P}_k = [x_{ij}^k / x_{..}^k]$. La moyenne est $\mathbf{P} = (1/K) \sum_k \mathbf{P}_k$.

On fait l'analyse des correspondances de \mathbf{P} , structure compromis utilisant une pondération uniforme et l'intrastructure consiste à projeter en individus supplémentaires les lignes et les colonnes des K tableaux de départ. Les pondérations de l'AFC du compromis servent de référence générale. Si on n'a aucune raison de pondérer inégalement les tableaux, cette analyse ne pose aucun problème de signification pour l'utilisateur. Dans STATIS, utiliser :



```
Option Foucart's COA
Input file RS•E
Number of rows: 24 columns: 79 blocs: 4
-----
```

Le tableau compromis est calculé et son analyse des correspondances est exécutée :

```
fc/COA: Correspondance analysis
Input file: A_M
Number of rows: 6, columns: 79
File A_M.fcpl contains the edge distribution of rows
It has 6 rows and 1 column
File A_M.fcpc contains the edge distribution of columns
```

It has 79 rows and 1 column
 File A_M.fcta contains the doubly centred table DI-1*P*DJ-1 -1I*1J'
 It has 6 rows and 79 columns
 Total inertia: 0.991344

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+5.7435E-01	+0.5794	+0.5794	02	+2.2225E-01	+0.2242	+0.8036
03	+9.1626E-02	+0.0924	+0.8960	04	+7.2898E-02	+0.0735	+0.9695
05	+3.0223E-02	+0.0305	+1.0000	06	+0.0000E+00	+0.0000	+1.0000

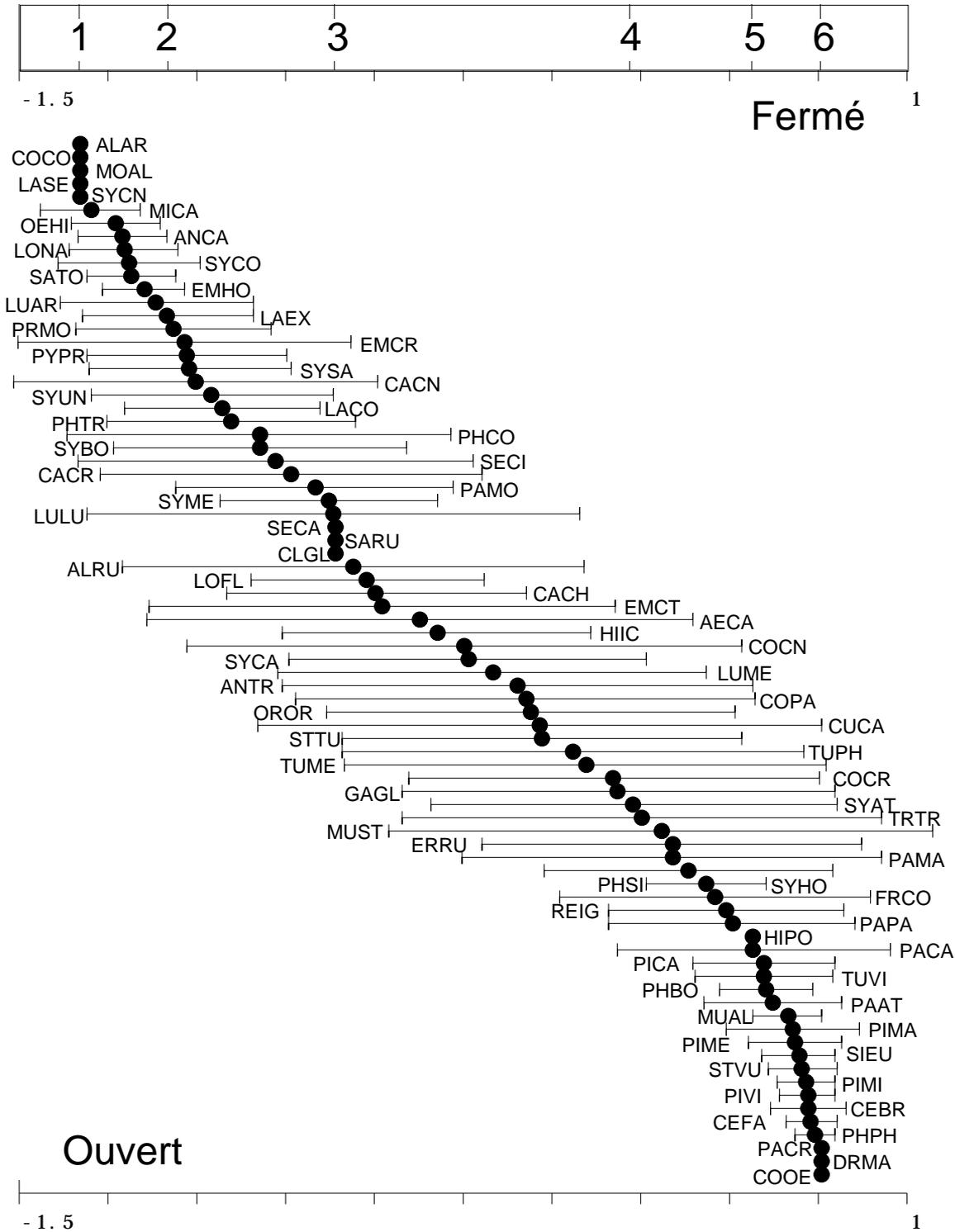


Figure 1 : Compromis de la structure Strates-Espèces.

File A_M.fcvp contains the eigenvalues and relative inertia for each axis
 --- It has 79 rows and 2 columns

File A_M.fcco contains the column scores
 --- It has 79 rows and 2 columns

File :A_M.fcco ----- Minimum/Maximum -----
 Col.: 1 Mini = -1.75692 Maxi = 0.999686
 Col.: 2 Mini = -1.62795 Maxi = 1.78529

File A_M.fcli contains the row scores
 --- It has 6 rows and 2 columns

File :A_M.fcli ----- Minimum/Maximum -----
 Col.: 1 Mini = -1.33149 Maxi = 0.75762
 Col.: 2 Mini = -0.767466 Maxi = 0.841644

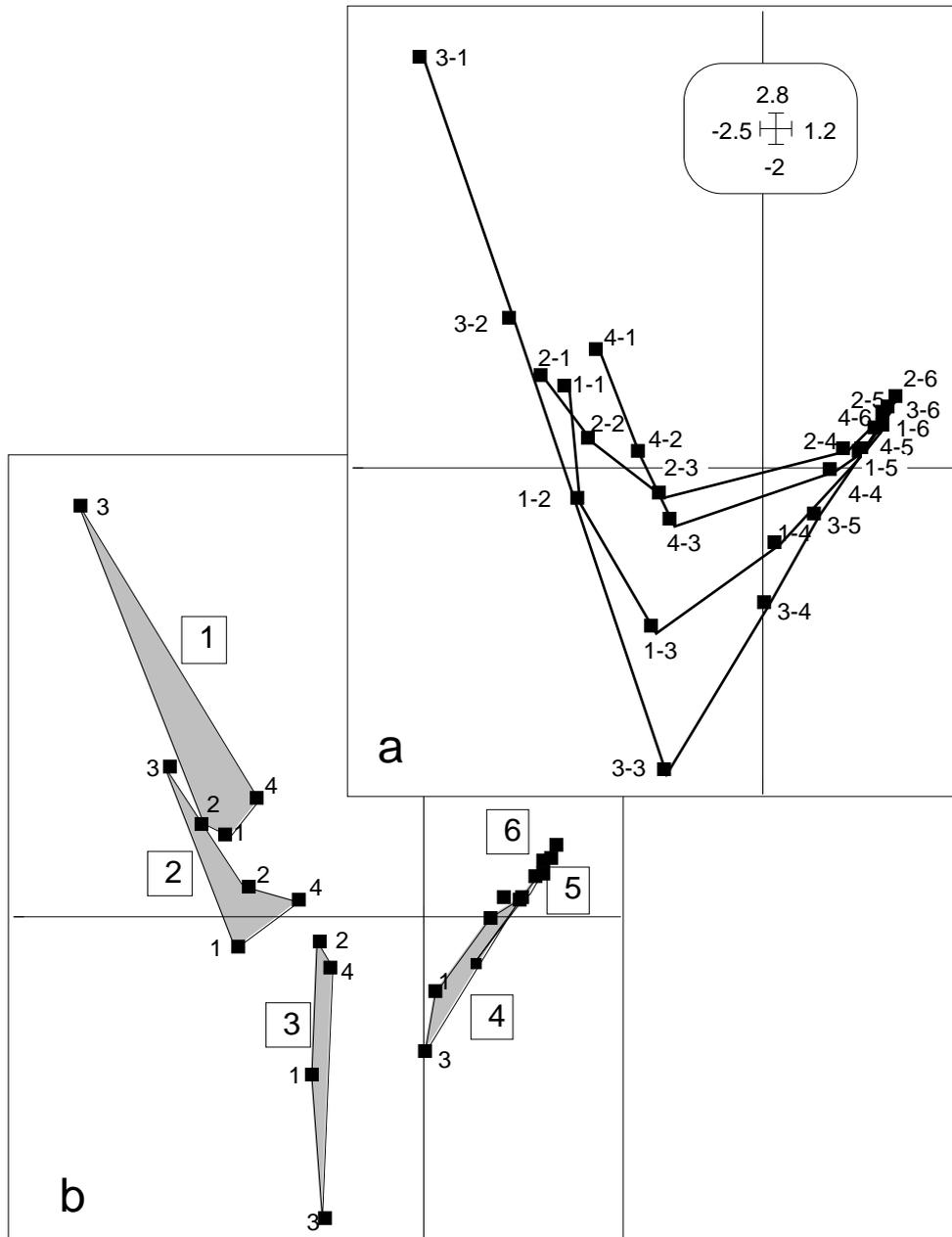


Figure 2 : Compromis des structures espèces-strates. Les discordances inter-régions proviennent surtout des difficultés de coordination d'expériences dans des séries de végétation très différentes. La structure du compromis, malgré ces difficultés, est très forte.

Les lignes du multi-tableau sont projetées (6x4 = 24) :

File A_M_TLl contains the coordinates of the projected rows
 It has 24 rows and 2 columns
 File :A_M_TLl

-----Minimum/Maximum:
 Col.: 1 Mini = -2.0774 Maxi = 0.82481
 Col.: 2 Mini = -1.8677 Maxi = 2.4982

Les colonnes du multi-tableau sont projetées (79x4 = 316) :

File A_M_TCc contains the coordinates of the projected columns
 It has 316 rows and 2 columns
 File :A_M_TCc

-----Minimum/Maximum:
 Col.: 1 Mini = -1.7569 Maxi = 0.99969
 Col.: 2 Mini = -1.6279 Maxi = 1.7853

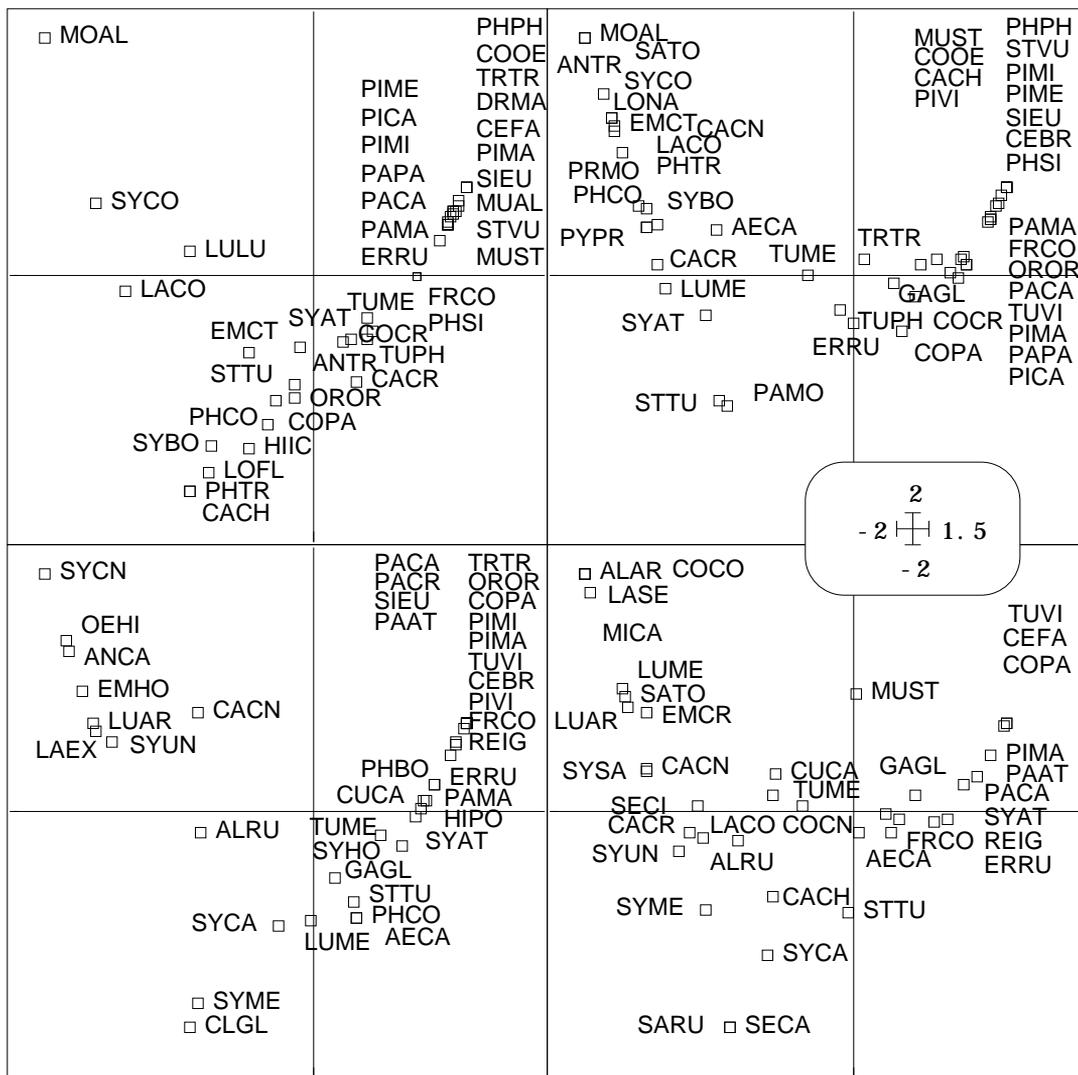


Figure 3 : Projection des colonnes de chaque tableau. La représentation privilégie la lisibilité des étiquettes mais cache l'architecture des sous-nuages reprise dans la figure 4.

Nous avons d'abord fait une analyse des correspondances. Toutes régions confondues, il s'agit d'ordonner les espèces sur le gradient défini par l'architecture de la végétation. C'est un cas typique d'utilisation de TabMeanVar (figure 1) ou Tables : TabMeanVar.

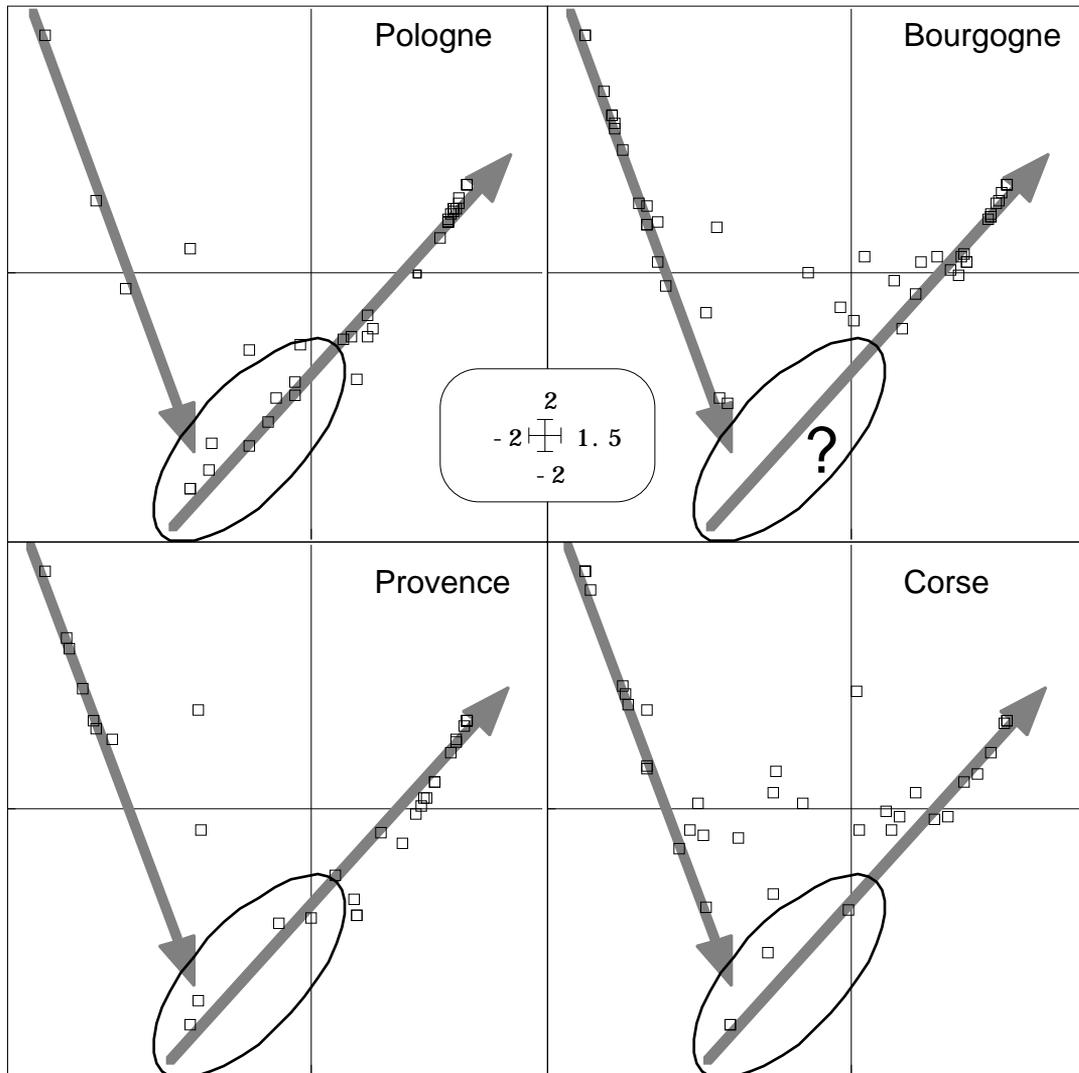


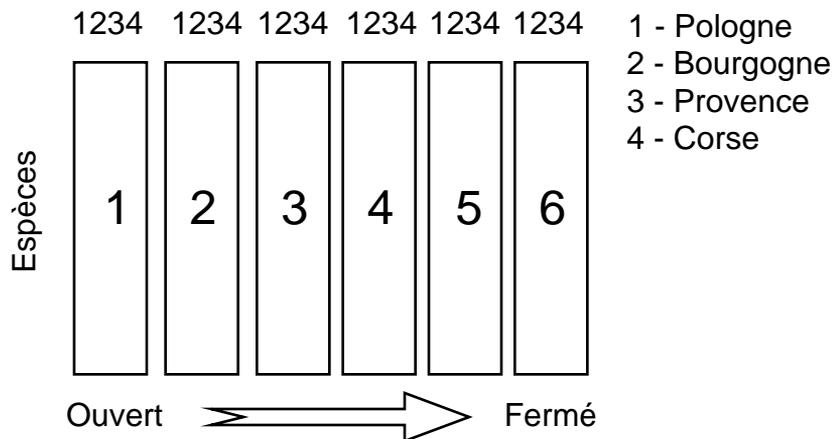
Figure 4 : Mise en évidence de la forme commune des nuages des espèces par régions. Une même disposition dans un plan des 6 strates (analyse du tableau moyen) positionne les espèces de chaque région par leur position moyenne (à une constante près). En dépit des difficultés expérimentales inhérentes au choix de stratifications parallèles dans des situations géographiques très différentes, on reproduit remarquablement bien la même structure. Le point d'interrogation souligne une différence importante.

Le fichier A_M_TL1 donne la figure 2. Chaque point de ces cartes se réfère à la typologie compromis des espèces par un averaging à une dilatation près. On y apprend que le compromis est plus précis dans la strate 6 que dans la strate 1. La position du point Provence de la strate 1 est assez original. La convergence des listes d'espèces en milieu forestier, au centre des préoccupations de Blondel & Farré (op. cit.) s'accompagne d'une convergence des structures, le calage des gradients entre eux étant d'autant plus précis qu'est grand le nombre d'espèces qui y participent. Les figures 1 et 2 forment, en tout cas, un excellent outil d'exploration de la structure d'une face (ici espèces-strates) du cube de données.

Pour manipuler les étiquettes des espèces il faut coller quatre fois le contenu du fichier Code_Esp_BF88 dans un fichier Code_Esp_BF88_TCc et utiliser encore Scatters. On obtient la figure 3, reprise dans la figure 4. Dans les deux cas, on a enlevé le paquet de points à l'origine qui correspond à l'ensemble des taxons n'apparaissant pas dans une région donnée. Sur la figure 4, on a simplement enlevé les étiquettes et

souligner la forme commune des cinq nuages de points. La structure compromise fait complètement abstraction des différences entre cortèges faunistiques qui sera discutée par ailleurs. Une structure moyenne (compromis entre régions de la relation avifaune-végétation) s'exprime avec une reproductibilité assez remarquable. Remarquable aussi est la notion de permanence de structures et d'hétérogénéité de contenu. La liste des espèces évolue considérablement d'une région à l'autre, mais la structure se conserve. Alors que l'inverse est possible. Dans les analyses d'évolution temporelle d'une structure spatiale on peut s'attendre soit à conserver une carte (modèle spatial) en changeant de descripteurs, soit à changer la carte en conservant les descripteurs. Dans un cas, on cherchera à cartographier l'évolution, dans l'autre on décrira l'évolution d'une carte. On peut s'attendre enfin, en changeant de descripteurs, à changer de structure.

2.3 — La variabilité de l'expression d'une structure



Il y a 6 analyses des correspondances de la structure Espèces - Régions, donc une analyse moyenne.

InitKTab

Matrix input file 24 79

Row indicator (Default = 1 class) 6 1

Col indicator (Default = 1 class)

Option: output file name

Foucart's COA

---.ktp input file

Matrix input file 24 79

Option: output file name

```
Option Foucart's COA
Input file S•R/E
Number of rows: 24 columns: 79 blocs: 6
-----
```

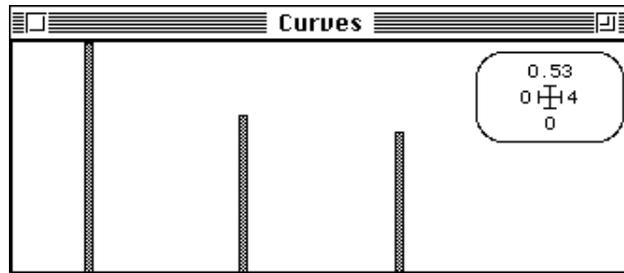
Le tableau de fréquence moyen est calculé et analysé par une AFC standard :

```

fc/COA: Correspondance analysis
Input file: S•R/E_M
Number of rows: 4, columns: 79
File S•R/E_M.fcpl contains the edge distribution of rows
It has 4 rows and 1 column
File S•R/E_M.fcpc contains the edge distribution of columns
It has 79 rows and 1 column
File S•R/E_M.fcta contains the doubly centred table DI-1*P*DJ-1 -1I*1J'
It has 4 rows and 79 columns
File S•R/E_M.fcma contains:
    the number of rows: 4
    the number of columns: 79
    the total number: 1
  
```

 Total inertia: 1.21032

Num.	Eigenval.	R.Iner.	R.Sum	Num.	Eigenval.	R.Iner.	R.Sum
01	+5.2777E-01	+0.4361	+0.4361	02	+3.5909E-01	+0.2967	+0.7327
03	+3.2346E-01	+0.2673	+1.0000	04	+0.0000E+00	+0.0000	+1.0000



On a ici un exemple rarissime d'analyse de données qui ne conduit pas à une réduction de dimension plus forte que celle qui est implicite dans la méthode choisie.

```

File S•R/E_M.fcvp contains the eigenvalues and relative inertia for each axis
--- It has 79 rows and 2 columns
  
```

```

File S•R/E_M.fcco contains the column scores
--- It has 79 rows and 3 columns
  
```

```

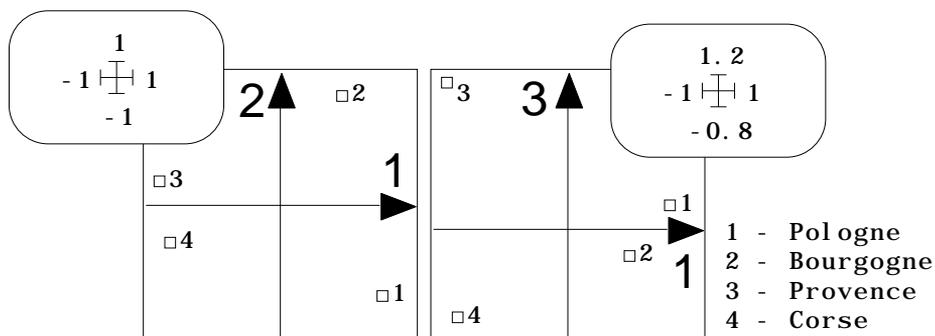
File :S•R/E_M.fcco ----- Minimum/Maximum -----
Col.:  1  Mini =  -1.21479  Maxi =   1.05459
Col.:  2  Mini =  -1.16343  Maxi =   1.35149
Col.:  3  Mini =  -1.14844  Maxi =   1.93755
  
```

```

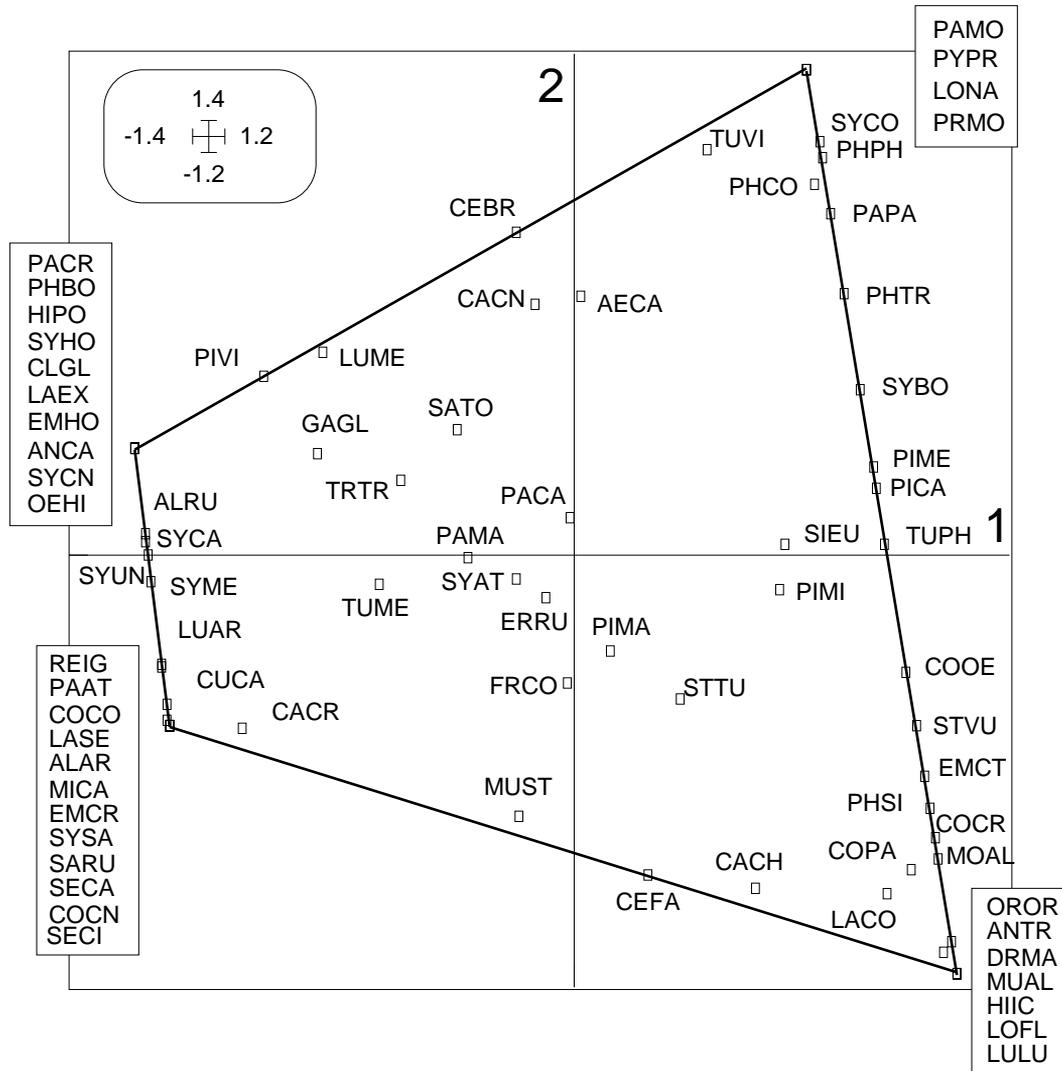
File S•R/E_M.fcli contains the row scores
--- It has 4 rows and 3 columns
  
```

```

File :S•R/E_M.fcli ----- Minimum/Maximum -----
Col.:  1  Mini =  -0.882519  Maxi =   0.766137
Col.:  2  Mini =  -0.697172  Maxi =   0.809868
Col.:  3  Mini =  -0.65316   Maxi =   1.10196
  
```



Il faut trois dimensions pour distinguer les quatre cortèges faunistiques (méditerranéen contre continental, Pologne contre Bourgogne, Provence contre Corse). Plus d'une espèce sur trois n'est présente que dans une région (29/79). La carte 1-2 des espèces dans l'analyse moyenne montre au centre les espèces présentes dans toutes les régions, à la périphérie les espèces présentes dans une ou deux régions. Le jeu des moyennes de moyennes alignent les taxons présents dans deux régions (à droite, Pologne et Bourgogne, à gauche, Provence et Corse). Les espèces à la distribution inter-régionale la plus large (au centre) sont en majorité forestières ce qui définit les gradients convergents de la représentation du type TLI.



Title of the analysis: S•R/E_M.fcta
 Number of rows: 4, columns: 79

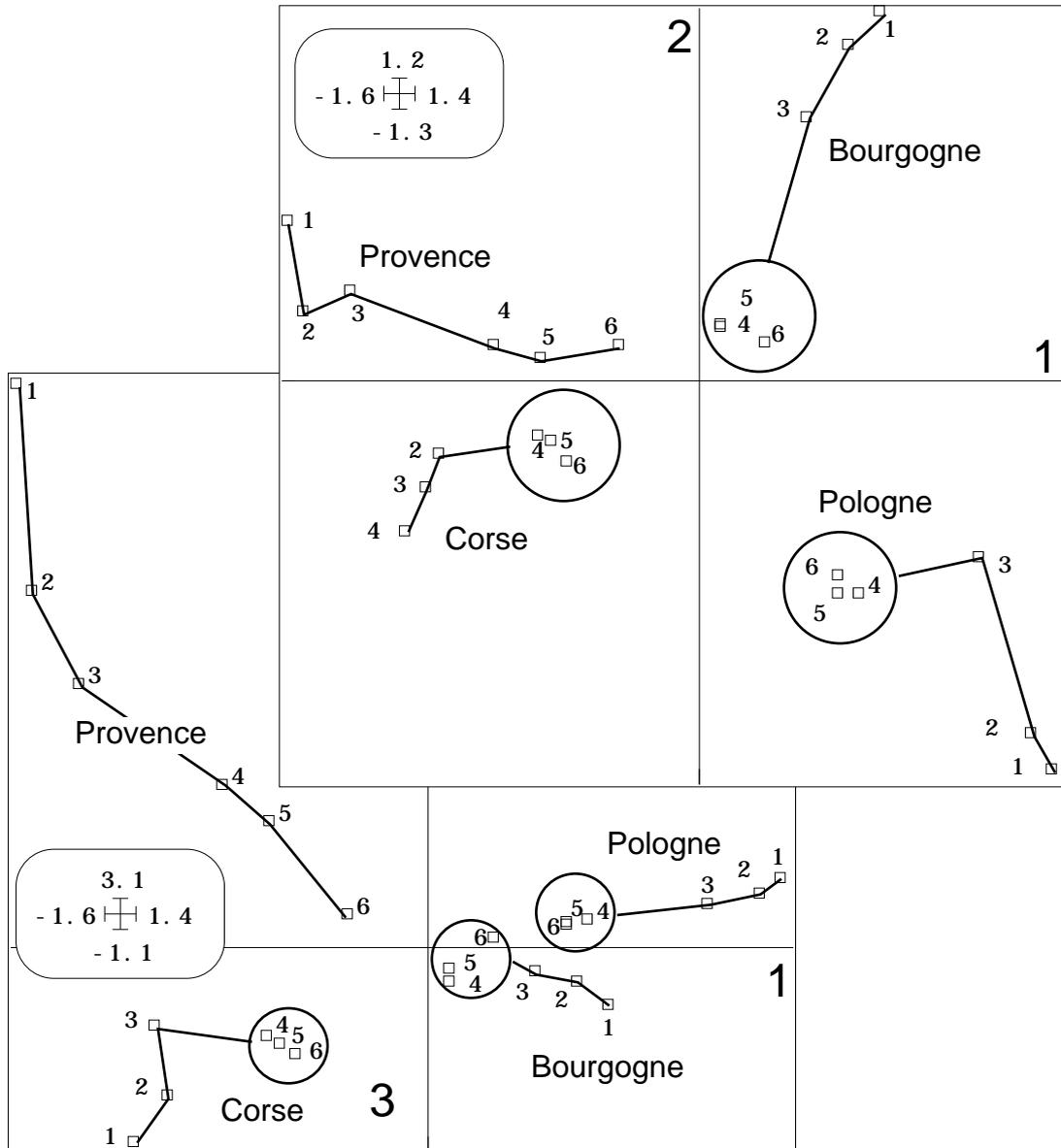
File S•R/E_M_TL1 contains the coordinates of the projected rows
 It has 24 rows and 3 columns

File :S•R/E_M_TL1

-----Minimum/Maximum:
 Col.: 1 Mini = -1.5711 Maxi = 1.3582
 Col.: 2 Mini = -1.2531 Maxi = 1.1911
 Col.: 3 Mini = -1.0461 Maxi = 3.0547

La projection en individus supplémentaires des 24 lignes (6 groupes de 4) contient alors toute l'information recherchée. L'analyse du tableau total fournit de l'information aussi bien inter-régions (variation régionale du cortège faunistique) que inter-strates

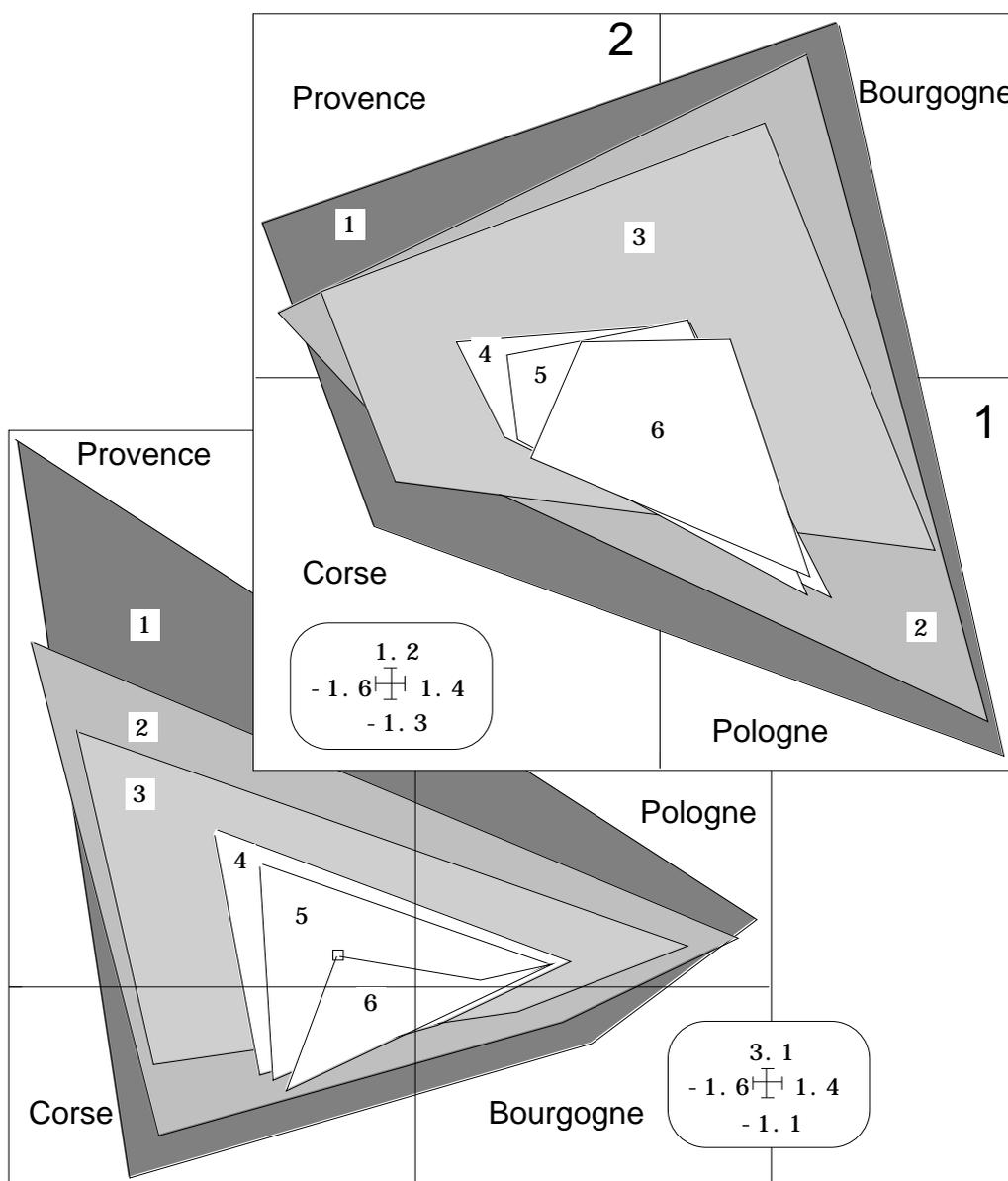
(stratification du cortège faunistique sur le gradient de hauteur de la végétation). Elle peut encore concerner une variation simultanée ou une interaction. L'AFC est apte à enregistrer tous types de structure inter-colonnes qui seront plus ou moins difficiles à identifier, éventuellement à mélanger les genres, ce qui rend l'exercice de dépouillement parfois périlleux. C'est pourquoi Blondel & Farré utilise un plan 1-4 qui contient une partie de l'interaction strates-régions. L'analyse du tableau R•S/E mélange donc tous les éléments que nous venons de voir et elle est donc à déconseiller. Ici, l'analyse ne restitue que de l'information inter-régions, mais indique clairement que celle-ci dépend essentiellement de la strate.



Les trois strates forestières sont quasiment équivalentes. La disparité inter-régionale y est faible, sans être nulle. La typologie se renforce dans le gradient d'ouverture et on peut parler soit de convergence des communautés dans les milieux fermés, soit de divergence des contenus faunistiques dans les milieux ouverts. La deuxième formulation est plus en accord avec la discussion des auteurs.

La même carte est reprise pour souligner combien l'analyse de Foucart a apporté une réponse appropriée à la question posée. On y souligne simplement l'amplitude de la structure avifaune-région en fonction du type de végétation. Il était logique de placer

chacun des 6 tableaux à égalité, pour montrer à quel point ils expriment la structure moyenne de manière différente. On est bien passé, ici aussi, d'une typologie d'états à une typologie de structures, essentiellement due à une variabilité de l'intensité des différences inter-régionales en fonction de la hauteur de la végétation.



Le dernier fichier créé permet de replacer la discussion au niveau de chaque espèce et intéressera les spécialistes.

Title of the analysis: S•R/E_M.fcta
 Number of rows: 4, columns: 79

File S•R/E_M_TCc contains the coordinates of the projected columns
 It has 474 rows and 3 columns

File :S•R/E_M_TCc

-----Minimum/Maximum:

Col.: 1 Mini = -1.2148 Maxi = 1.0546
 Col.: 2 Mini = -1.1634 Maxi = 1.3515
 Col.: 3 Mini = -1.1484 Maxi = 1.9375

Références

- ¹ Lescourret, F. (1994) Modélisation des données temporelles et rôle du graphisme. *Vet Res* : 25, 140-146.
- ² Kroonenberg, P.M. (1989) The analysis of multiple tables in factorial ecology. III Three-mode principal component analysis: "analyse triadique complète". *Acta Œcologica, Œcologia Generalis* : 10, 3, 245-256.
- ³ Blondel, J. & Farre, H. (1988) The convergent trajectories of bird communities along ecological successions in european forests. *Œcologia (Berlin)* : 75, 83-93.
- ⁴ Franc, A. (1992) *Etude algébrique des multitableaux : apports de l'algèbre tensorielle*. Thèse de doctorat, Université de Montpellier II. 1-361.

Kroonenberg, P.M. (1983) *Three-mode principal component analysis*. DSWO Press, Leiden. 1-380.

Coppi, R. & Bolasco, S. Eds. (1989) *Multiway Data Analysis*. Elsevier Science Publishers B.V., North-Holland. 1-552.
- ⁵ Foucart, T. (1983) Une nouvelle approche de la méthode STATIS. *Revue de Statistique Appliquée* : 31. 2. 61-75.
- ⁶ Foucart, T. (1978) Sur les suites de tableaux de contingence indexés par le temps. *Statistique et Analyse des données* : 2, 67-84.

