

STATIS et valeurs typologiques des groupes faunistiques

Résumé

La fiche donne une première présentation de la méthode ACT-STATIS (Lavit, Ch., 1988, *Analyse conjointe de tableaux quantitatifs*. Masson, Paris. 1-240). On aborde le cas d'un multi-tableau formé d'un tableau faunistique et de la répartition des taxons en plusieurs groupes faunistiques. Une mesure de la valeur typologique de chaque groupe est proposée.

Plan

1 — Introduction.....	2
2 — Données traitées.....	7
3 — La norme des opérateurs.....	7
4 — Cartes factorielles des analyses séparées	10
5 — L'analyse du compromis	11
6 — Axes des analyses séparées et plan du compromis.....	13
7 — Reconstruction du plan du compromis par tableau	16
8 — Trajectoires	17
9 — Valeur typologique d'un groupe faunistique.....	19
Annexe : Pour refaire les calculs.....	20
Références	29

D. Chessel, J. Thioulouse et M. Hanafi

1 — Introduction

La méthode ACT-STATIS (“Analyse Conjointe de tableaux - Structuration des Tableaux à Trois Indices de la Statistique”) est référencée à la thèse de L'Hermier des Plantes ¹. Elle fait l'objet de l'ouvrage de référence de Ch. Lavit ² et d'une synthèse récente ³. L'essentiel du support théorique de base est contenu dans les articles de Y. Escoufier ⁴.

Les utilisations de la méthode en écologie sont très rares. On doit citer l'étude d'Amanieu & Coll. ⁵ Le résumé de cet article contient des éléments précis sur le champ d'utilisation et la problématique sous-jacente. Il dit :

Au cours de 13 mois d'échantillonnage, neuf stations ont été systématiquement inventoriées dans une lagune littorale méditerranéenne, l'étang de Prévost à Palavas. Sur chaque échantillon (par date et par station) a été calculée la diversité de Shannon rapportée à la biomasse de la macrofaune benthique. Une méthode d'analyse factorielle (méthode STATIS) a été appliquée aux matrices de données en vue d'estimer les distances, du point de vue de la diversité des diverses stations entre elles et d'en déduire les regroupements de stations les plus cohérents. Les résultats font ressortir qu'il y a, selon le type d'analyse, deux (stations marines, stations lagunaires) ou trois (les stations marines étant en outre dissociées en deux ensembles) groupes mis en évidence; ces groupes correspondent aux ensembles biocénotiques fondamentaux déjà décrits dans l'étang. La diversité apparaît donc comme un bon indice des regroupements biocénotiques; les estimations qu'elle permet à ce point de vue sont en outre d'autant meilleures que l'on prend en compte un plus grand nombre de groupes systématiques; en revanche les groupes systématiques isolés sont peu discriminants, les Mollusques étant toutefois plus utiles à cet égard que les Polychètes, qui sont eux-mêmes préférables aux Crustacés.

Les données sont tridimensionnelles (stations-dates-variables). Les variables sont les indices de diversité calculés par groupe faunistique et la discussion porte sur la valeur typologique de ces groupes faunistiques (valeurs vues ici par le biais des indices de diversité, ce qui est un choix particulier). On notera d'entrée que la situation abordée (dates-stations-espèces) est très générale et que la question de la valeur typologique des groupes faunistiques est souvent introduite dans la discussion écologique sans qu'on connaisse de méthodes clairement établies pour la mesurer. En fait les données initiales croisent des relevés (formant un plan d'observations spatio-temporel) et des espèces, lesquelles forment des groupes faunistiques qui ont une plus ou moins grande valeur pour identifier le fonctionnement spatio-temporel des communautés. On trouve alors trois situations caractéristiques représentées dans la figure 1.

Dans la première (A), on mesure l'abondance des taxons dans un ensemble de stations donné. Les taxons sont répartis par groupes faunistiques. Chaque groupe définit une typologie des sites. Peut-on définir une typologie moyenne ou une typologie somme? Intuitivement, une typologie moyenne prendrait en compte la structure qui se retrouverait régulièrement d'un tableau à l'autre, tandis qu'une typologie somme additionnerait les éléments typologiques provenant de l'un ou l'autre des tableaux. La typologie moyenne peut-elle ne contenir que la partie commune aux typologies induites par chaque groupe ? La typologie somme peut-elle au contraire réunir tous les éléments de structures originaux induits par chaque groupe ? Est-il alors possible de repérer les apports respectifs à cette typologie somme ? Peut-on enfin caractériser la valeur typologique de chaque groupe ? La présente fiche aborde ces questions.

Dans la seconde situation (B), on mesure l'abondance des taxons dans des sites, lesquels sont maintenant répartis en groupe. Ils peuvent faire partie d'un ensemble régional donné et être l'objet d'une visite partielle ou totale à plusieurs reprises (dates). Ils peuvent au contraire appartenir à des ensembles géographique distincts (régions). Dans chaque ensemble on a une typologie réciproque espèces-sites classique. Peut-on comparer les typologies d'espèces ainsi définies ? Y a t'il pérennité ou variabilité (temporelle ou régionale) de la séparation inter-spécifique des niches des taxons. Certaines dates ou certaines régions sont-elles incapables de reproduire la typologie inter-spécifique somme ou moyenne ? La seconde fiche abordera ces questions.

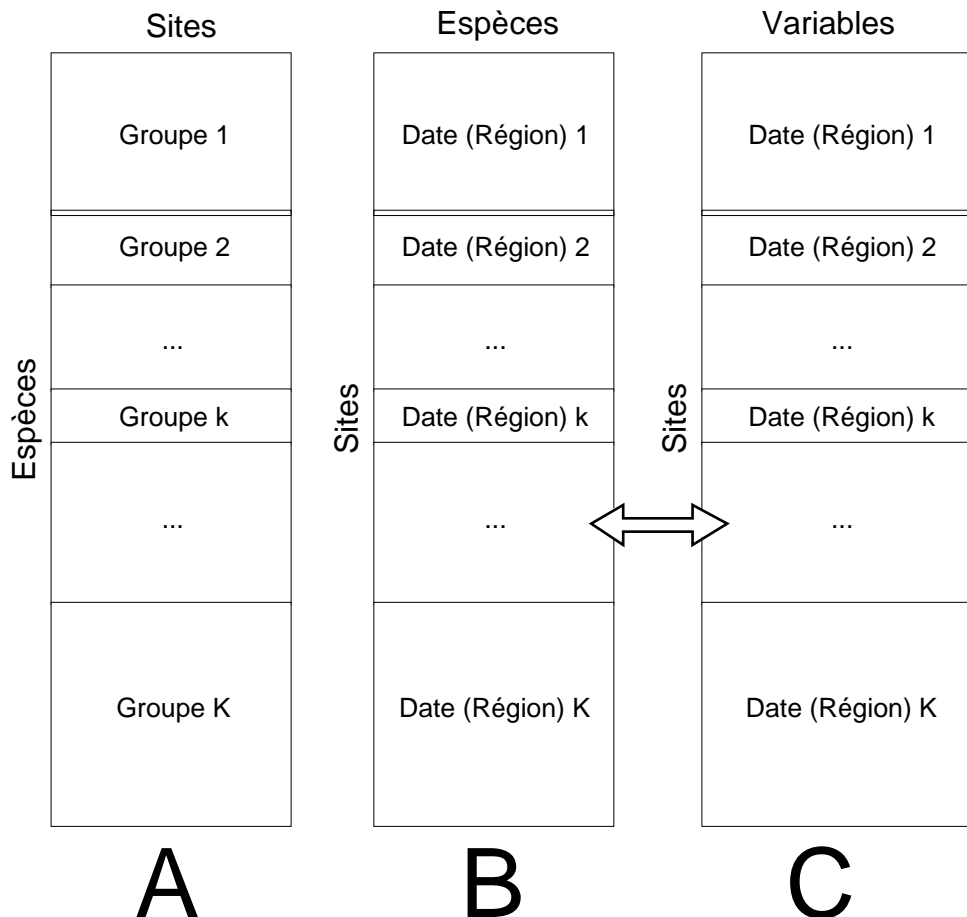


Figure 1 : trois situations écologiques donnant un K-tableau.

Dans la troisième situation (C), on retrouve la seconde en utilisant des variables mésologiques. On peut se poser les mêmes questions. La problématique définie sur la structure s'étend à la co-structure espèce-milieu.

Peut-on définir une co-typologie somme ou moyenne ? Peut-on en mesurer la réalisation partielle ou totale dans certains groupes ? On abordera cet aspect dans la troisième fiche. D'autres situations seront abordées ensuite, les situations expérimentales qui justifient une approche "du type STATIS" étant extrêmement diversifiées.

Ces questions relèvent d'une situation unique (figure 2). Dans les situations A et B de la figure 1, on trouve les espèces en lignes, puis en colonnes. Il suffirait d'une transposition pour obtenir deux tableaux sites-espèces classiques, mais le premier serait partitionné sur les lignes et le second le serait sur les colonnes.

La théorie classique distingue les deux cas par souci de respecter l'habitude très ancrée qu'une analyse en composantes principales place les individus en lignes et les variables en colonnes, le centrage se faisant colonne par colonne.

L'essentiel est de reconnaître que chaque tableau relève d'une analyse qui lui est propre (centrée par ligne ou par colonne, suivant les cas, voire non centrée ou doublement centrée, si cela lui convient, ou normée dans un sens ou dans l'autre). Chaque tableau relève d'une analyse pertinente : c'est le seul critère à prendre en compte par l'utilisateur.

Il faut, en effet, se méfier. Formellement les données se présentent *très exactement* comme dans un problème d'analyse discriminante. L'analyse discriminante fera (sensiblement) l'analyse des tableaux pour en faire disparaître tous les effets. On ne s'y intéressera qu'à la différences des moyennes entre groupes. L'analyse discriminante est voisine de l'analyse inter-classes laquelle est antinomique avec l'intra-classe qui calculera les moyennes des groupes pour en éliminer les différences. L'analyse intra-classe est donc proche de STATIS qui se consacre à l'étude de la structure interne alors que discriminante et inter-classe se consacrent aux différences moyennes entre groupe. Chaque groupe de méthodes fait, pour atteindre ses objectifs, tout ce qui est nécessaire pour faire disparaître les effets mis en évidence par l'autre. Autant savoir ce qu'on cherche avant d'appuyer sur le bouton. L'analyse intra-classe et STATIS ayant des points communs, il faudra établir ce qui en fait la différence.

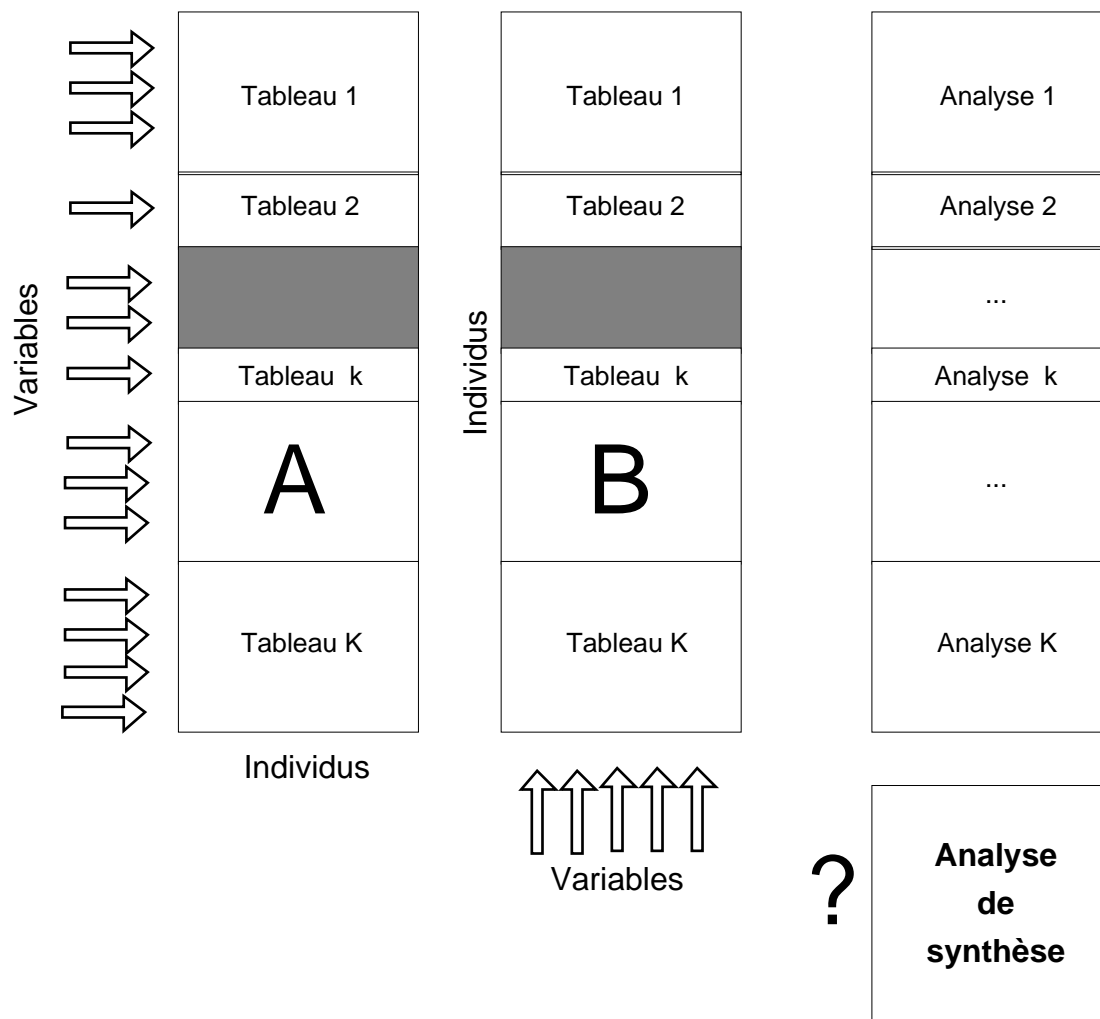


Figure 2 : la situation abordée par STATIS est celle de K analyses liées par une des marges.

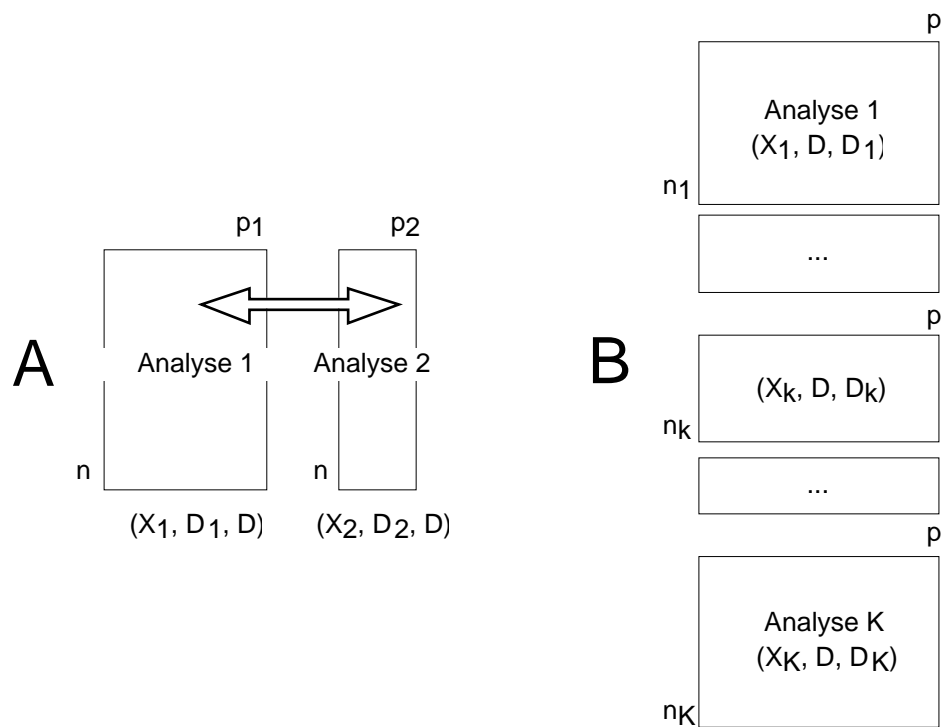


Figure 3 : STATIS aborde une situation qui généralise la notion de co-structure de 2 à K tableaux.

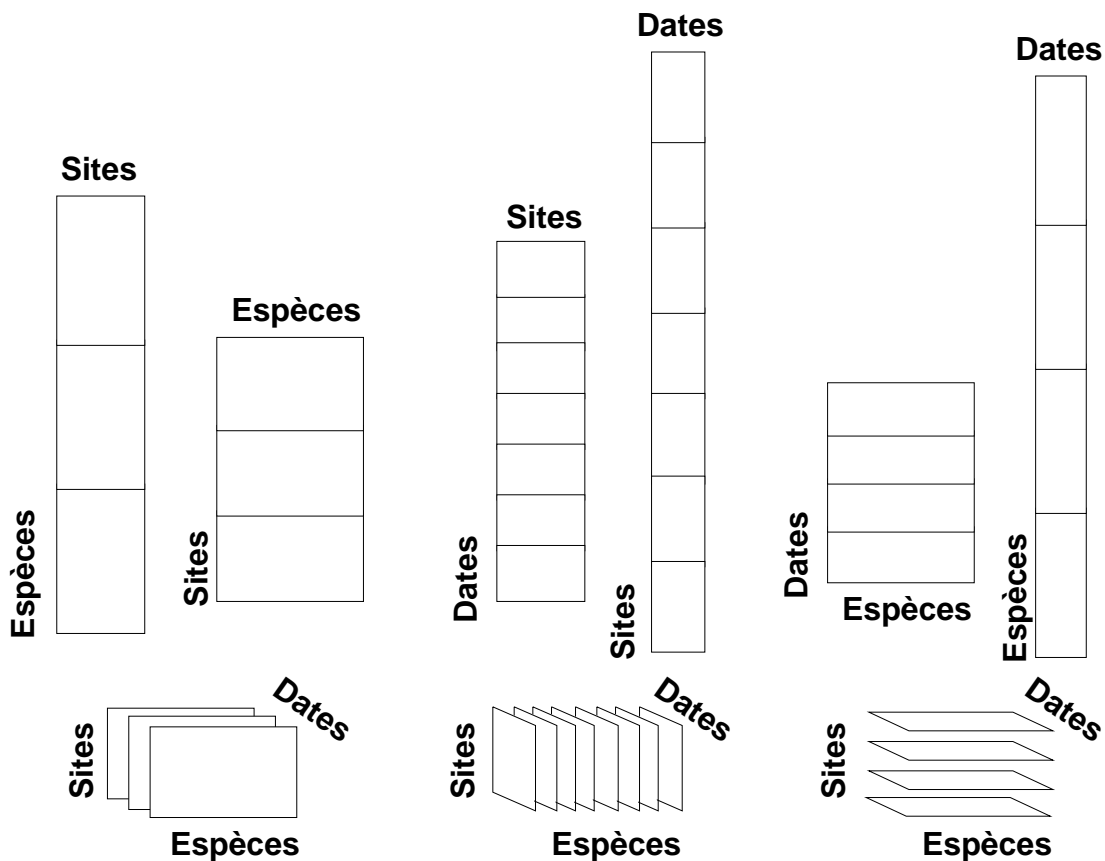


Figure 4 : Six configurations pour STATIS sont générées par un cube de données

	Q	P	R	J	E	C	D	K	B	A	G	M	L	F	H	N		Q	P	R	J	E	C	D	K	B	A	G	M	L	F	H	N						
1 - Hemiptera																	6 - Diptera																						
A1	3						2										F1	3	4	4	4	3	3	3	3	4	3	3	3	1	3	3	4						
A2	2															1	F2	1	2		2		2		3	2	2	2			3	1							
A3	2	3	2	2	2	4	3	1	3	3	1	1	1				F3	2	2		2		2	2	2	2													
A4	1		2		1					1						1	F4	3	4		3	3																	
A5	1		2		1					1					1	1	F5	1	3	2	1	2	3	2	2	2	2			1	2	3							
A6		1		1		2	1			1	1	1	2	2	3		F6	3		3	2	3	3		2		2	2			2								
A7			2			1			1	2	1	1				1	F7	3	4	3	4	4	4		3	3	2	3		2	3	3							
A8			2								1					1	F8				3	2			2		2			2									
A9			1		4	2		3								3	F9		3			2										2							
Aa					2	1		2		1	2					1	Fa				3							3											
2 - Odonata																	Fb																						
Ab			1				1			1						2	Fc	2	4	2	3	3	4	3	3	4	3	2	3		3	3	3						
B1		2	2		1					1							Fd				2	4	4	2	4	4	3	2	3	3	3	3							
B2		2								3	1	2	1				Fe				3	3									2								
B3		3	2		2	2				1	1	1				1	Ff	2			3	3	3	4	3	3	3	3		3	4	4							
B4	2	2	2	1	2			1	2			2		1	1	3	Fg			1				3								3							
B5		2	3	2	2	3	2	2	3	3	2	3				1	Fh			2					3														
B6			1	1	3	2	2	3	3	2	3			3	2	3	Fi				3		3	2	2	3	3	3	4	4	4								
B7							2				1	2				3	Fj					2	3		2	3		3	2		3								
3 - Trichoptera																	Fk																						
C1	2	3	2	2	2	2	1	1	2	3		2		1			Fl								3		2		3	3	4								
C2	1		1	2	2	3	1	3	3	3	2	2	1		2	3	Fm														3	2							
C3		1		2		1						2		1	1		7 - Hydracarina																						
C4			2	3	1	1	1		1	1					2	G1			2						2														
C5					3	2											G2			1					2	1	2				2	2							
C6							3		1	1	1		2	3	2		G3				2	2			1						2								
C7							2	2		2	2			2	2		G4													2	2								
C8							3			1			2	1		8 - Malacostraca																							
C9							2				2		3	1	2	H1					3	3	3	2	3	3				2	3								
Ca							2	2		2		2	2	3	3	H2													2	2									
Cb		1								1				2	3	H3														2	3	3							
Cc											2		2	2		9 - Mollusca																							
Cd												2	3			I1						4	4	2	3	3	3	3		3	4								
4 - Ephemeroptera																	I2																						
D1					4	3	2	4	4	1	3		2	1			I3					1	1			2	1	1											
D2					3	3	3	3	3	3	3	2	3	3	4		I4								2		2	2		1	3								
D3										1			2	3	4		I5									3					3								
D4							2		1	1		2	3	3		I6									2					1									
5 - Coleoptera																	I7																						
E1	1	2	1	1		1	2			1		1					I8												3		2	3							
E2	1	2			2				2		2	1	2		1		10 - Oligochaeta																						
E3		1			2				2		3	2	1		1	J1	4	3	4	4	2	2	2	3	2	2	3			1	2	3							
E4		2		2	2			1	2							1	J2	2		2	3	2	3	3	2	3	3		2	3		2	3						
E5			1	2					1								J3	2		2					1		2	1	4	1	3	2							
E6			2								1						J4						3	2	1	2	2	3	4										
E7			1						2								J5													3	2	3							
E8			1				2			2	3				2		J6										1		3	3	3								
E9					2	1																																	
Ea					2	1																																	
Eb									1	2	1					1																							
Ec										1	2					2																							
Ed															2	2																							

Tableau 1 : données faunistiques de L.E. Friday (1987).

La situation décrite dans la figure 2 donne un sens à l'analyse de la structure de chacun des tableaux, lesquels ont les colonnes en commun. On cherchera à comparer des typologies et mettre en évidence ce qui se retrouve d'un tableau à l'autre. STATIS est donc liée à l'analyse de co-inertie qui aborde la question à deux tableaux seulement

(figure 3). L'une comme l'autre demande que les objets en commun aient dans chaque tableau les mêmes poids.

STATIS étend donc l'analyse de co-inertie dans deux directions. D'une part on passe de deux à K tableaux (K quelconque). D'autre part l'appariement, qui normalement portait sur les lignes-individus (variables en colonnes) dans la co-inertie est aussi bien autorisé pour les individus que pour les variables. Les K tableaux ont donc en commun la pondération \mathbf{D} de leur p colonnes communes. Il faut alors noter que dans un cube de données, il y a six manières d'aborder la méthode STATIS (figure 4).

2 — Données traitées

Les données qui illustreront ce premier emploi de la méthode STATIS sont publiées par Friday⁶. Elles sont formées de 10 tableaux faunistiques regroupant un total de 91 taxons et 16 étangs. Il y a respectivement 11, 7, 13, 4, 13, 22, 4, 3, 8 et 6 taxons par groupe. Chaque tableau est identifié par le groupe de taxons utilisés (1 - Hemiptera, 2 - Odonata, 3 - Trichoptera, 4 - Ephemeroptera, 5 - Coleoptera, 6 -- Diptera, 7 - Hydracarina, 8 - Malacostraca, 9 - Mollusca, 10 - Oligochaeta. Les 16 étangs sont étiquetés simplement par les lettres Q P R J E C D K B A G M L F H N. Les données sont dans le tableau 1.

S'agissant de relevés de macrofaune benthique standardisés, chaque tableau relève d'une ACP centrée par lignes (les espèces sont en lignes et les relevés en colonnes). La figure 5 donne les valeurs propres de ces 10 analyses élémentaires.

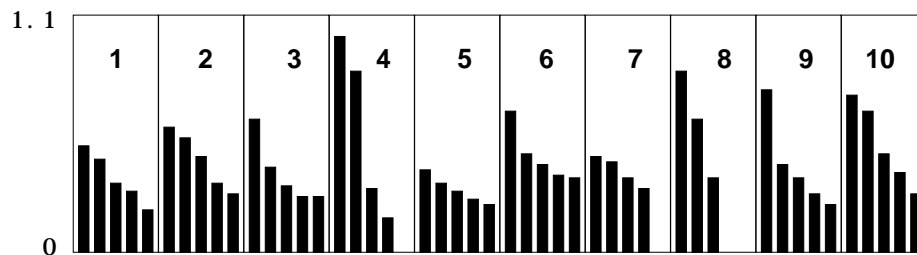


Figure 5 : Valeurs propres des analyses séparées.

Chaque tableau définit une structure. Selon les critères habituels, la mesure de la structure d'un tableau est son inertie totale. Ces inerties sont consignées dans le tableau 2. On s'attend à ce qu'un groupe faunistique comportant de nombreuses espèces fournisse une inertie totale (somme des variances des abondances des taxons) plus grande. Il n'en est rien et la corrélation inertie-richesse est nulle. La figure 5 souligne le fait que l'inertie totale n'indique rien sur la présence d'une typologie : on peut dire qu'utilisés seuls les groupes 1, 2, 5, 7, 8 et 10 n'auraient pas grand chose à dire sur une éventuelle typologie de relevés. On garderait pour les groupes 3, 6 et 9 le premier axe et pour le seul groupe 4 les deux premiers. Curieusement il n'y a que 4 espèces éphéméroptères et c'est le tableau qui semble le plus structuré.

3 — La norme des opérateurs

STATIS propose de mesurer la valeur d'une analyse non par la somme des valeurs propres d'une analyse élémentaire (qui vaut l'inertie totale) mais par la somme des carrés de ces valeurs propres (carré de la norme d'Hilbert-Schmidt de l'opérateur d'inertie dit opérateur d'Escoufier associé à cette analyse). Il s'agissait d'abord dans cette modification d'une conséquence de la logique algébrique sous-jacente. En fait, la

signification écologique de cette innovation ne saurait échapper à un ... écologue. En effet, on sait que plus les valeurs propres d'une analyse sont différentes (c'est-à-dire, plus les premières valeurs propres sont grandes et les suivantes petites), meilleure est l'analyse (au sens que l'expression issue de la projection ne peut être un artefact). En passant de la somme (ou la moyenne) à la somme des carrés, on passe de la moyenne à la variance, de l'abondance à la diversité. Plus l'inertie exprimée sur les premiers axes est grande, pour une valeur d'inertie donnée, plus la norme de l'opérateur est grande. Les normes sont dans le tableau 2. Le groupe des éphéméroptères l'emporte sans conteste et c'est un autre groupe pauvre qui prend la seconde valeur.

	Inertie	Richesse	Norme
Tab 1	0.683	11	0.327
Tab 2	0.999	7	0.481
Tab 3	0.87	13	0.423
Tab 4	1.807	4	1.227
Tab 5	0.506	13	0.207
Tab 6	1.381	22	0.544
Tab 7	0.557	4	0.294
Tab 8	1.182	3	0.802
Tab 9	1.007	8	0.602
Tab 10	1.404	6	0.732

Tableau 2 : indices de valeurs typologiques par groupe faunistique.

On retiendra que pour le k^{ième} triplet statistique $(\mathbf{X}_k, \mathbf{D}, \mathbf{D}_k)$, on substitue à la notion habituelle d'inertie :

$$I_t = \text{Trace}(\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D}) = \lambda_i$$

celle de variance vectorielle ou norme d'opérateur (Escoufier 1973, op. cit.) :

$$Vav(\mathbf{X}_k) = \text{Trace}(\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D} \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D}) = \lambda_i^2$$

A la notion de mesure de structure, on ajoute le mesure de co-structure par le produit scalaire utilisé dans la norme précédente :

$$\begin{aligned} Covv(\mathbf{X}_k, \mathbf{X}_j) &= \text{Trace}(\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D} \mathbf{X}_j^t \mathbf{D}_j \mathbf{X}_j \mathbf{D}) \\ &= \text{Trace}(\mathbf{X}_j \mathbf{D} \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D} \mathbf{X}_j^t \mathbf{D}_j) = \mu_i \end{aligned}$$

Il s'agit exactement de la co-inertie totale associée au couple de tableaux qui se décompose dans la somme des valeurs propres de l'analyse de co-inertie.

Il s'en suit qu'on peut mesurer la corrélation entre deux triplets par le coefficient de corrélation vectoriel :

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

Nous ne ferons pas la distinction habituellement pratiquée entre les opérateurs \mathbf{VQ} et les opérateurs \mathbf{WD} de la théorie des opérateurs d'Escoufier. Le programme STATIS d'ADE-4 recouvre les deux cas de manière identique. La figure 6 permet de comprendre de quoi il s'agit. Dans la situation A de la figure 6 on considère deux tableaux d'ACP

centrée appariés par les lignes-individus. On possède alors deux matrices comparables de produits scalaires \mathbf{W}_1 et \mathbf{W}_2 et les éléments qui précèdent s'écrivent :

$$Vav(\mathbf{X}_1) = Trace(\mathbf{X}_1 \mathbf{D}_1 \mathbf{X}_1^t \mathbf{D} \mathbf{X}_1 \mathbf{D}_1 \mathbf{X}_1^t \mathbf{D}) = Trace(\mathbf{W}_1 \mathbf{D} \mathbf{W}_1 \mathbf{D})$$

$$Vav(\mathbf{X}_2) = Trace(\mathbf{X}_2 \mathbf{D}_2 \mathbf{X}_2^t \mathbf{D} \mathbf{X}_2 \mathbf{D}_2 \mathbf{X}_2^t \mathbf{D}) = Trace(\mathbf{W}_2 \mathbf{D} \mathbf{W}_2 \mathbf{D})$$

$$Covv(\mathbf{X}_1, \mathbf{X}_2) = Trace(\mathbf{X}_1 \mathbf{D}_1 \mathbf{X}_1^t \mathbf{D} \mathbf{X}_2 \mathbf{D}_2 \mathbf{X}_2^t \mathbf{D}) = Trace(\mathbf{W}_1 \mathbf{D} \mathbf{W}_2 \mathbf{D})$$

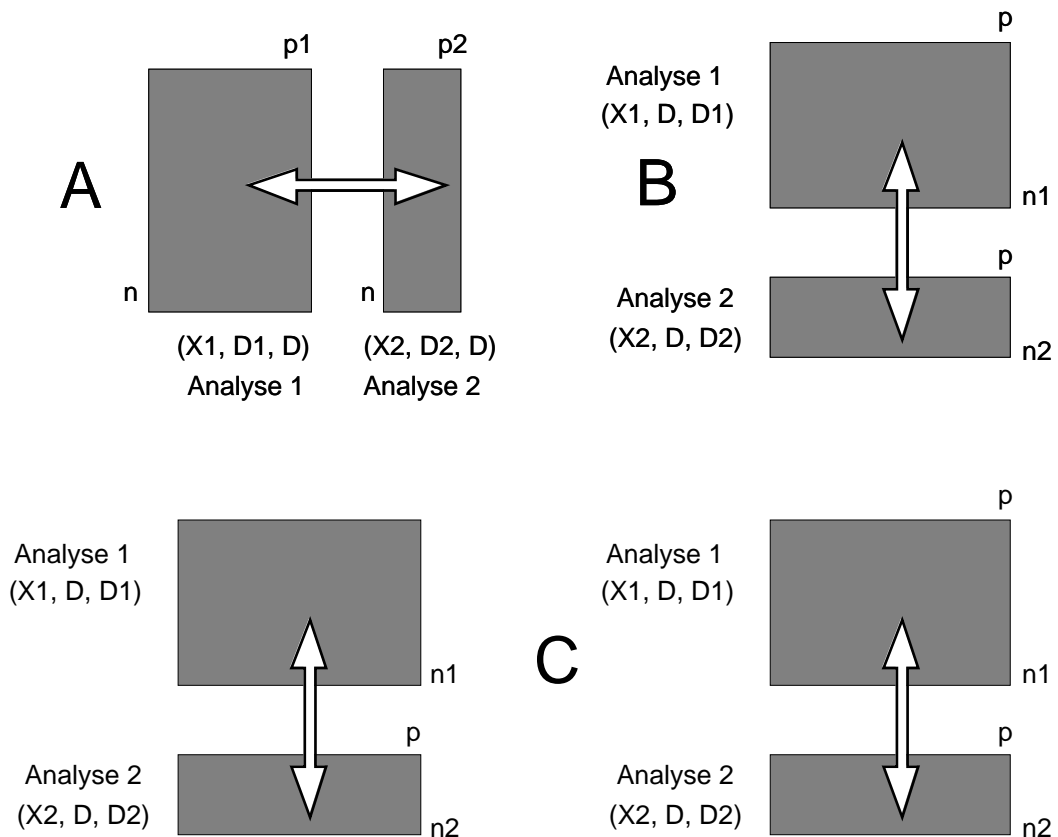


Figure 6 : STATIS aborde de la même manière des tableaux appariés par les individus ou par les variables. A - Deux tableaux d'ACP centrés dans la présentation traditionnelle ont les mêmes lignes-individus ; on compare des matrices de produits scalaires entre individus (matrices \mathbf{W}). B - Deux tableaux d'ACP centrés dans la présentation traditionnelle ont les mêmes colonnes-variables ; on compare des matrices de covariances (matrices \mathbf{V}). C - Dans ADE-4, un seul mode de représentation est utilisé (association par les colonnes) ; on compare des matrices \mathbf{U} qui sont soit du type \mathbf{V} soit du type \mathbf{W} , suivant le type de centrage prédéfini.

Dans la situation B de la figure 6, le centrage est opéré sur les lignes-variables et ce qui est comparable entre les deux tableaux concerne les matrices de covariances \mathbf{V}_1 et \mathbf{V}_2 :

$$Vav(\mathbf{X}_1) = Trace(\mathbf{X}_1^t \mathbf{D}_1 \mathbf{X}_1 \mathbf{D} \mathbf{X}_1^t \mathbf{D}_1 \mathbf{X}_1 \mathbf{D}) = Trace(\mathbf{V}_1 \mathbf{D} \mathbf{V}_1 \mathbf{D})$$

$$Vav(\mathbf{X}_2) = Trace(\mathbf{X}_2^t \mathbf{D}_2 \mathbf{X}_2 \mathbf{D} \mathbf{X}_2^t \mathbf{D}_2 \mathbf{X}_2 \mathbf{D}) = Trace(\mathbf{V}_2 \mathbf{D} \mathbf{V}_2 \mathbf{D})$$

$$Covv(\mathbf{X}_1, \mathbf{X}_2) = Trace(\mathbf{X}_1^t \mathbf{D}_1 \mathbf{X}_1 \mathbf{D} \mathbf{X}_2^t \mathbf{D}_2 \mathbf{X}_2 \mathbf{D}) = Trace(\mathbf{V}_1 \mathbf{D} \mathbf{V}_2 \mathbf{D})$$

En utilisant systématiquement la forme B avec des tableaux centrés par lignes ou par colonnes (ou de tout autre manière), on recouvre les deux situations classiques (et on en génère nombre de nouvelles) en manipulant des matrices d'opérateurs qui sont du type **V** ou **W** suivant les cas. Pour éviter toute confusion, on notera dans la présentation commune :

$$U_{kk} = \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \quad \text{Var}(\mathbf{X}_k) = \text{Trace}(\mathbf{U}_{kk} \mathbf{D} \mathbf{U}_{kk} \mathbf{D})$$

$$U_{jj} = \mathbf{X}_j^t \mathbf{D}_j \mathbf{X}_j \quad \text{Covv}(\mathbf{X}_k, \mathbf{X}_j) = \text{Trace}(\mathbf{U}_{kk} \mathbf{D} \mathbf{U}_{jj} \mathbf{D})$$

La matrice des coefficients Rv, pour les 10 tableaux considérés, vaut (valeurs en 1/1000) :

[1]	1000									
[2]	442	1000								
[3]	527	509	1000							
[4]	439	571	543	1000						
[5]	498	406	463	305	1000					
[6]	428	640	614	624	451	1000				
[7]	502	307	433	316	473	338	1000			
[8]	347	432	510	596	310	494	418	1000		
[9]	407	491	424	610	496	528	594	638	1000	
[10]	389	410	442	407	254	514	284	335	242	1000

4 — Cartes factorielles des analyses séparées

Quand on exécute les analyses séparées on obtient la figure 7.

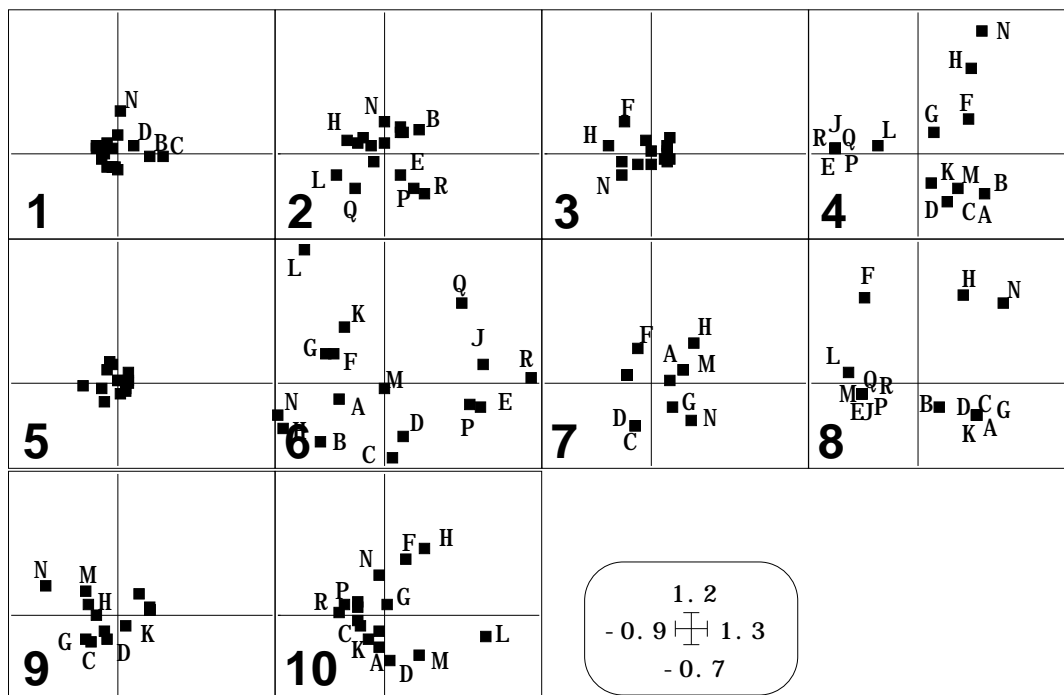


Figure 7 : Cartes 1-2 des 10 ACP séparées.

Le hasard impose qu'elles soient incohérentes. En effet, même si une structure identique s'exprimait d'un groupe à l'autre, il y a une chance sur deux de trouver un axe ou son opposé (deux vecteurs propres normés du même sous-espace) et donc très peu de chances (1 sur 512) de trouver tous les axes dans le même sens. Cette situation laisse

croire à une incohérence globale des ordinations exécutées par chaque groupe faunistique qu'on ne sait en aucune manière comparer deux à deux. Cela se ferait par 45 analyses de co-inertie entre les paires de deux tableaux distincts, ce qui n'est pas raisonnable. Les valeurs propres de la figure 5 s'expriment, en outre, par la variance des coordonnées. Sur les plans des tableaux 4, 8 et 10, on note cependant des points communs (groupe H, F et N) qui seront mis en évidence par STATIS.

5 — L'analyse du compromis

Le tableau proposé à la méthode STATIS est formé de 10 sous-tableaux appariés par leurs colonnes. Chacune de ces colonnes porte le poids 1/10. Les lignes forment des blocs par sous-tableau qui portent un poids uniforme (dans le tableau 1 à 11 lignes chaque ligne porte le poids 1/11, ...). Chacun des tableaux est simplement centré par ligne. Dans l'analyse du tableau k ($1 \leq k \leq 10$), la matrice :

$$U_{kk} = \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k$$

est la matrice des produits scalaires de l'ACP classique du tableau faunistique centré par taxon.

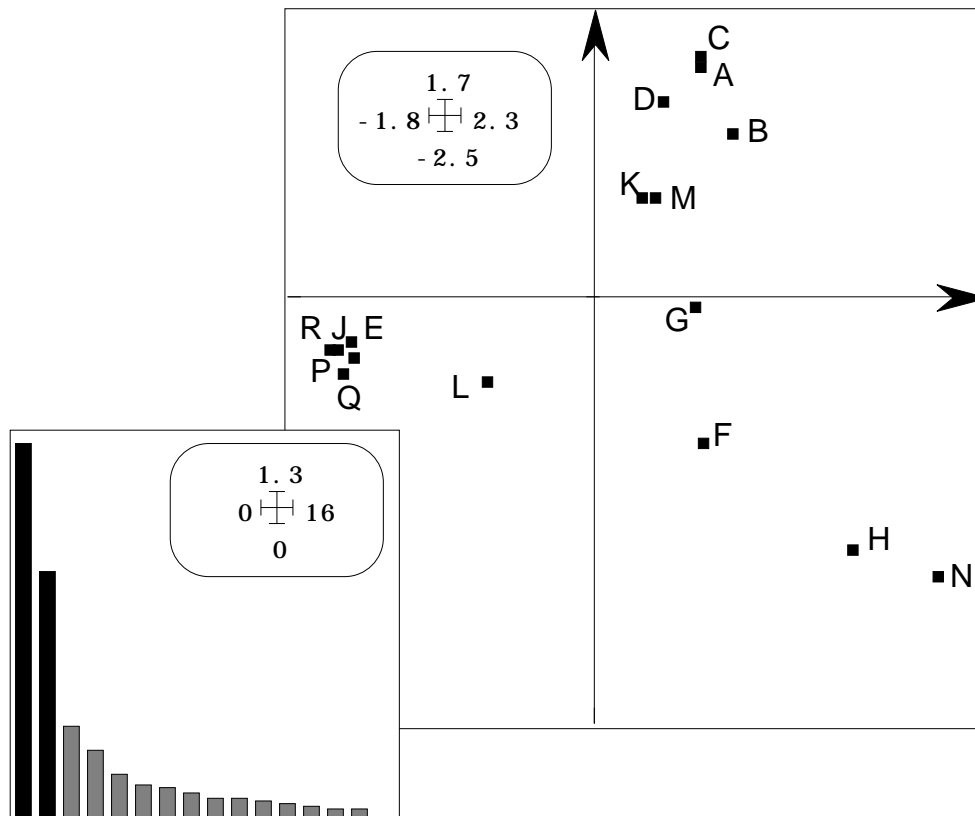


Figure 8 : carte des relevés dans l'analyse du compromis.

La seule modification introduite dans les analyses simples est une pondération uniforme des espèces. L'inertie de chaque tableau n'est pas la somme des variances des taxons mais la moyenne des variances des taxons. C'est pourquoi, bien qu'on compare des tableaux avec des nombres sensiblement différents d'espèces, les valeurs propres (figure 5) et les cartes des relevés (figure 7) sont comparables. La matrice diagonalisée dans l'analyse séparée du tableau k est :

$$U_{kk} \mathbf{D} = \mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{D} \quad \mathbf{D} = \frac{1}{16} \mathbf{I}_{16}$$

Chacune de ces matrices a 16 lignes et 16 colonnes. L'idée fondamentale de STATIS est d'en faire une combinaison, qu'on appelle matrice compromis pour obtenir par diagonalisation une carte de référence. Cette combinaison s'écrit :

$$\mathbf{UD} = \sum_{k=1}^K \alpha_k \mathbf{U}_{kk} \mathbf{D}$$

Les coefficients vérifient la contrainte habituelle en ACP :

$$\sum_{k=1}^K \alpha_k^2 = 1$$

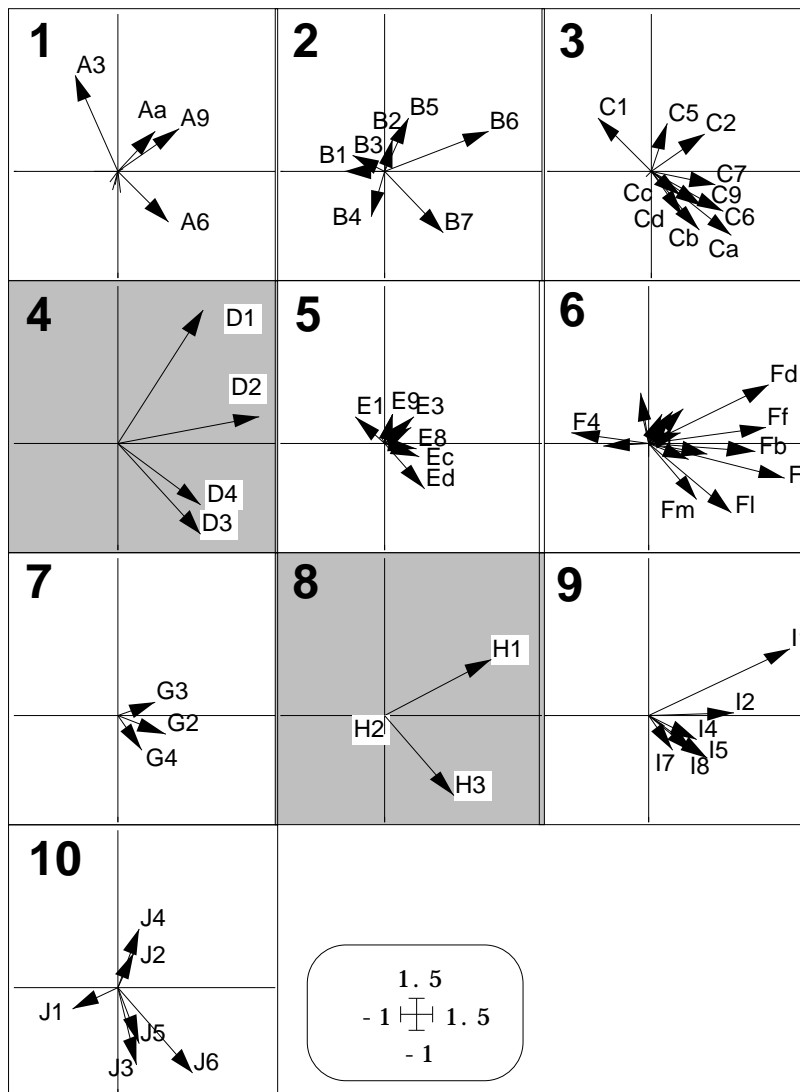


Figure 9 : carte des espèces dans l'analyse du compromis.

Ils sont calculés pour optimiser $Trace(\mathbf{UDUD})$, c'est-à-dire la somme des carrés de ses valeurs propres. Le résultat (dérivé de l'ACP du nuage des opérateurs avec la métrique d'Hilbert-Schmidt) est obtenu par le premier vecteur propre normé de la matrice à K lignes et K colonnes qui contient à la ligne k et la colonne j la quantité :

$$\text{Covv}(\mathbf{X}_k, \mathbf{X}_j) = \text{Trace}(\mathbf{U}_{kk} \mathbf{D} \mathbf{U}_{jj} \mathbf{D})$$

Les coefficients de compromis, pour l'exemple étudié valent :

1	2	3	4	5	6	7	8	9	10
0.112	0.207	0.180	0.733	0.057	0.260	0.086	0.388	0.280	0.258

A nouveau, le rôle des groupes 3 et 8 est mis en évidence. Le groupe des coléoptères (5) est délibérément ignoré. La diagonalisation du compromis donne les valeurs propres et la carte factorielle de la figure 8.

Cette carte des relevés de l'analyse du compromis s'accompagne d'une carte espèce liée à la précédente par les règles de calcul habituelles dans l'analyse d'un triplet statistique. Si on note \mathbf{z}_k le $k^{\text{ième}}$ vecteur propre \mathbf{D} -normé de la matrice \mathbf{UD} et λ_k la $k^{\text{ième}}$ valeur propre associée, les colonnes du tableau de départ (relevés) sont positionnées par les composantes de $\sqrt{\lambda_k} \mathbf{z}_k$ et les lignes (espèces) par \mathbf{XDz}_k . La carte des espèces est constituée ici de covariances entre abondances des taxons et coordonnées normalisées des lignes (figure 9).

6 — Axes des analyses séparées et plan du compromis

Les aides à l'interprétation graphique des relations entre les analyses séparées et l'analyse du compromis sont simples d'emploi dans ADE-4. La première approche concerne la projection des axes des analyses séparées⁷. Le tableau k définit le triplet statistique $(\mathbf{X}_k, \mathbf{D}, \mathbf{D}_k)$ dont la diagonalisation donnent des coordonnées lignes et colonnes. On sait que les axes principaux sont des vecteurs \mathbf{D} -normés dont les composantes sont les coordonnées normalisées dites de variance unité. Les axes du compromis sont dans le même espace et on peut projeter systématiquement les 4 premiers axes de chaque analyse (par convention, si le rang du tableau est inférieur à 4 on complète par des vecteurs nuls) sur les axes du compromis. La représentation graphique peut se faire par numéros d'ordre des axes (figure 10) ou par tableau (figure 11).

Dans la première représentation on souligne la notion de compromis. ADE-4 propose de définir plus strictement cette fonction essentielle. Dans les travaux cités, en effet, STATIS commence par la diagonalisation de la matrice des covariances vectorielles et conserve plusieurs vecteurs propres. Cette opération s'appelle l'analyse de l'interstructure. Nous n'avons conservé que le premier et éliminé sans examen les suivants. Le premier vecteur propre a un statut radicalement différent des suivants, exactement comme dans l'ACP non centrée d'un tableau homogène. Toutes ses composantes sont de même signe (on les prend automatiquement positives) et définissent les poids des opérateurs dans le compromis. Il a donc pour fonction de calculer une moyenne, comme l'ACP non centrée associée au facteur 1 la définition d'un modèle multiplicatif. Les suivants caractériseront l'écart au modèle moyen.

En mélangeant ces deux fonctions dans l'analyse de l'interstructure, on crée une confusion regrettable. Doit-on s'interroger sur l'existence du compromis ? Pas plus, à notre avis, qu'on doit s'interroger sur l'existence d'une moyenne. Il existe un opérateur moyen de K opérateurs d'inertie comme il existe une valeur moyenne de K mesures.

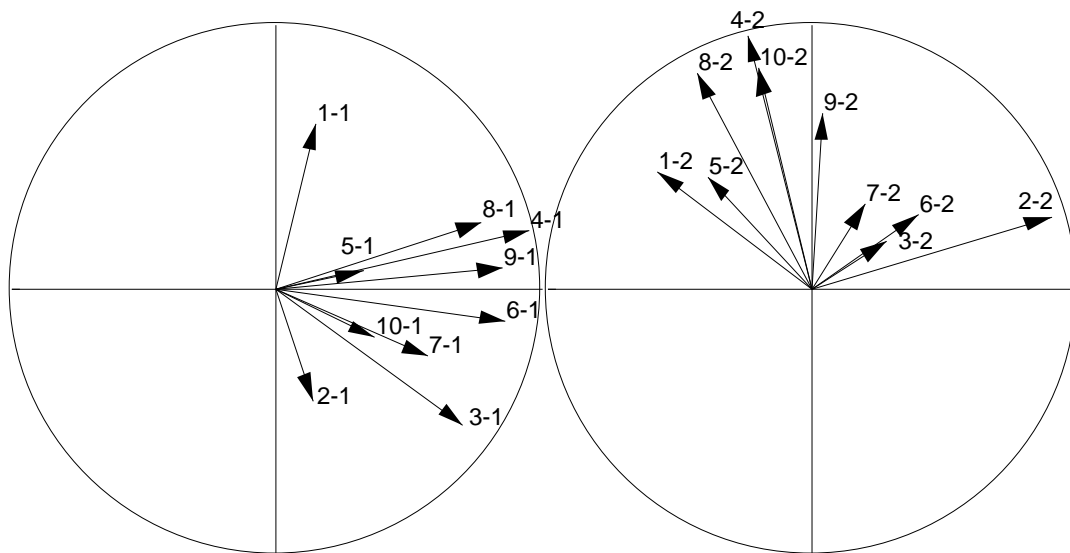


Figure 10 : projection des axes des analyses séparées dans le plan de l'analyse du compromis. A gauche : projections des axes 1 (10 tableaux). A droite : projections des axes 2.

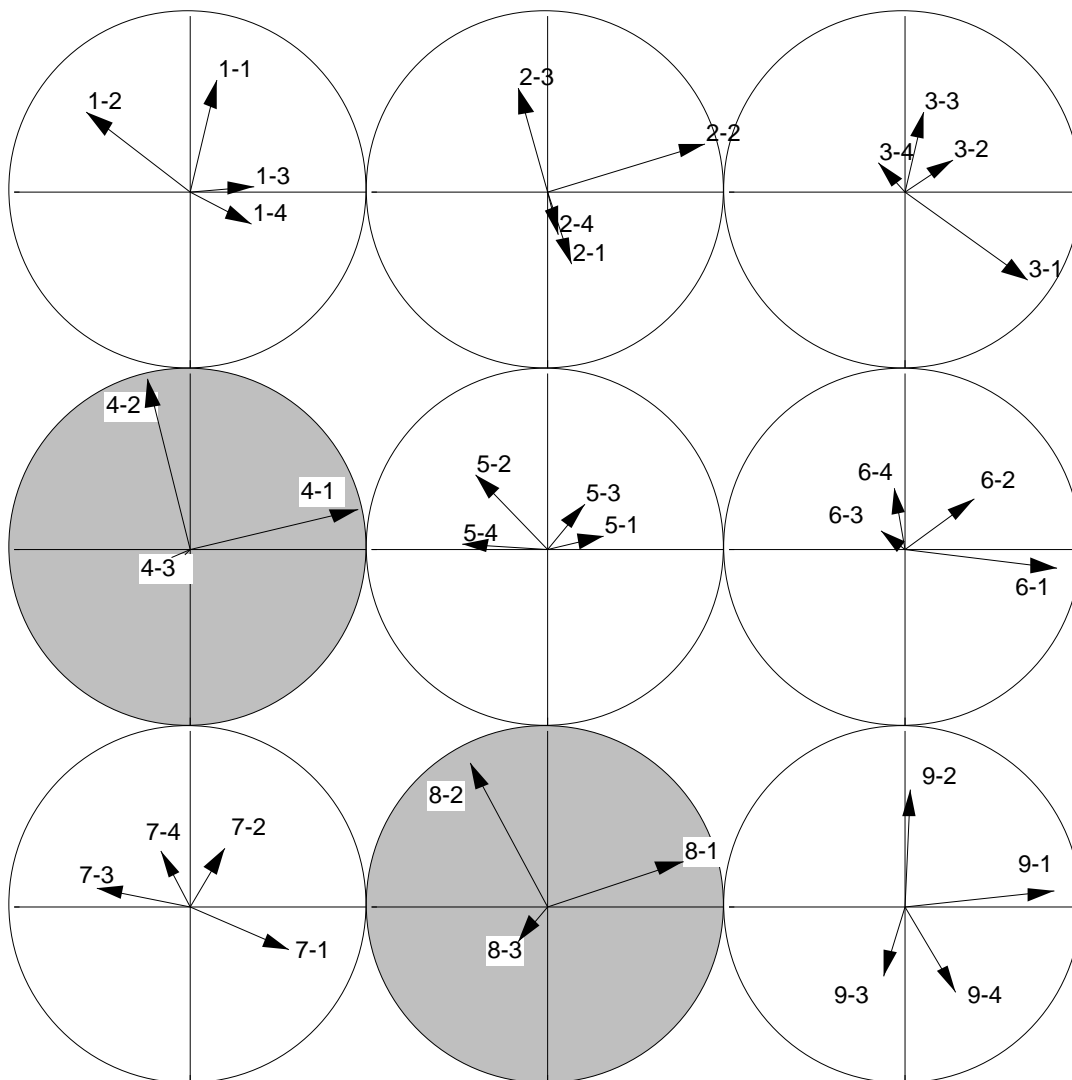


Figure 11 : projection des 3 ou 4 premiers axes des analyses séparées dans le plan de l'analyse du compromis.. Représentation par tableau..

Que l'on considère ensuite la dispersion autour de la moyenne est un problème évidemment pertinent, mais qui ne remet pas en cause la notion même de moyenne.

On notera qu'il existe deux niveaux d'écart entre les structures. Le premier concerne l'écart entre la structure d'un tableau et la structure du tableau fictif qui a pour opérateur d'inertie le compromis. Le second concerne l'écart entre l'opérateur d'inertie d'un tableau et l'opérateur d'inertie du compromis. C'est loin d'être la même chose. En effet, un tableau peut ne pas reproduire la typologie moyenne : cela ne signifie pas qu'il ne produit aucune typologie. En abordant les deux questions simultanément, la présentation classique de STATIS rend son usage délicat pour un non statisticien.

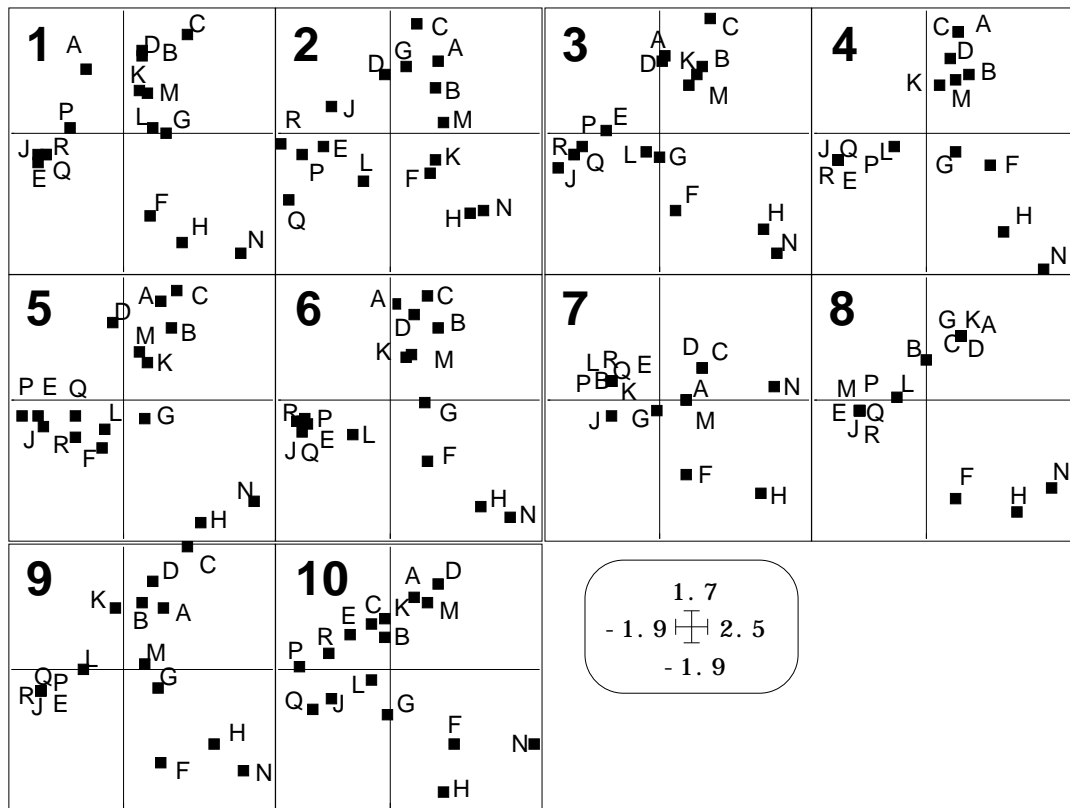


Figure 12 : Reconstitution de la carte du compromis par chacun des 10 tableaux.

La question écologique ainsi posée est d'importance. Des groupes faunistiques différents pourraient-ils générer des typologies de relevés pertinentes mais différentes ? Ici, la réponse est non. En effet, quand un tableau reproduit peu la structure du compromis, il ne produit aucune structure. On le voit, par exemple pour le groupe des coléoptères (5). L'angle des axes principaux est très important (figure 10). Son système d'axes n'a rien à voir avec le plan du compromis (figure 11). La covariance entre l'abondance des espèces de ce groupe et les coordonnées des relevés est très faible (figure 9). Son poids dans le compromis est très faible (0.057, Cf. § 2.4). Il n'est pas concerné par la typologie compromis. En même temps, son inertie et sa norme sont les plus faibles (en dépit de la présence de 13 taxa). Ses valeurs propres ne distinguent aucune direction privilégiée (figure 4). Il est faiblement structurant.

Pour simplifier et préciser l'objectif, nous dirons que STATIS définit une structure moyenne (calcul du compromis), décrit cette structure (analyse du compromis) et discute de la possibilité pour chacun des tableaux de produire cette structure. Pour l'écologie, la structure du compromis est naturellement celle qui est induite par la variabilité du milieu, laquelle est identique pour chaque groupe faunistique et s'exprime ou ne s'exprime pas dans chaque groupe. Cela renvoie à une version de STATIS

décentrée sur l'analyse d'un même tableau de référence extérieur au K-tableau proprement dit.

7 — Reconstruction du plan du compromis par tableau

La seconde aide à l'interprétation concerne ce qui s'appelle dans l'ouvrage de C. Lavit (op. cit.) les trajectoires. Le terme dérive du fait que si les différents tableaux sont associés à des dates distinctes, représenter un objet commun pour chaque tableau revient à définir une évolution. Les coordonnées issues du compromis sont des codes numériques (ici centrés) de norme au carré égales aux valeurs propres.

Ces coordonnées expriment la structure du compromis. On a noté \mathbf{z}_j le $j^{\text{ème}}$ vecteur propre \mathbf{D} -normé de la matrice \mathbf{UD} et λ_j la $j^{\text{ème}}$ valeur propre associée, les coordonnées du compromis (relevés) valent les composantes de $\sqrt{\lambda_j} \mathbf{z}_j$ (figure 8). Les lignes de chaque tableau (espèces) sont alors représentées par les composantes de $\mathbf{X}_k \mathbf{D} \mathbf{z}_j$ (figure 9), soit pour le tableau k par $\mathbf{X}_k \mathbf{D} \mathbf{z}_j$.

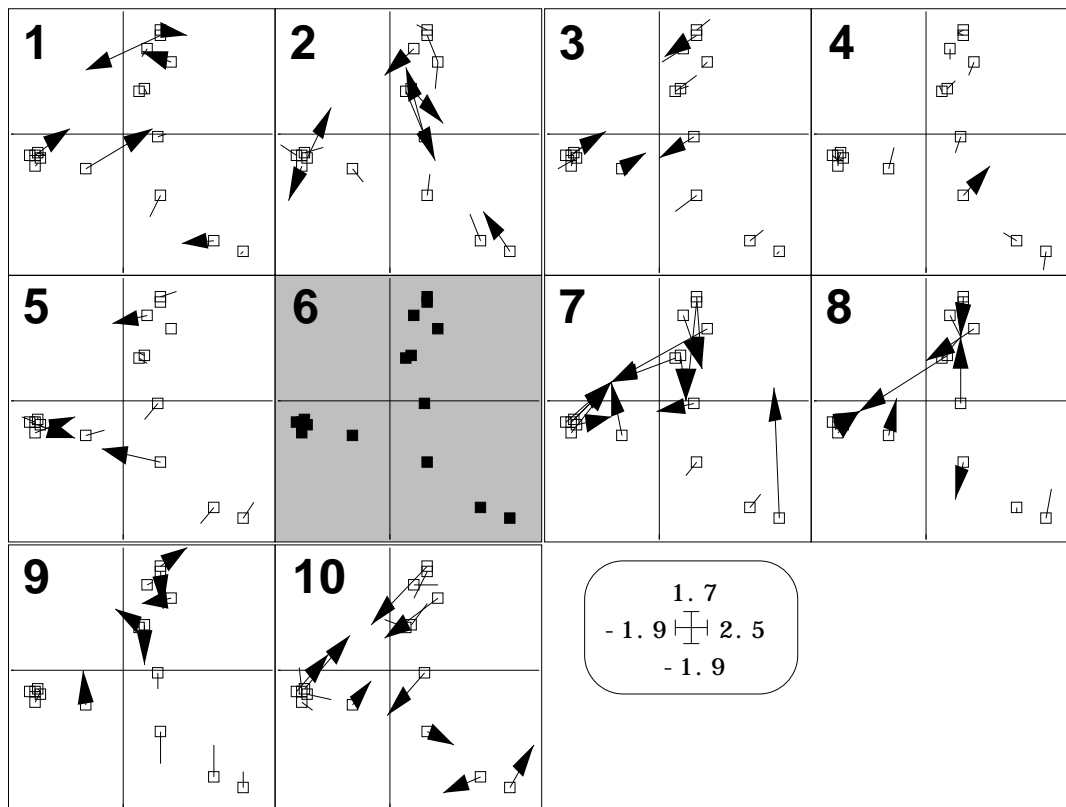


Figure 13 : Reconstitution de la carte du compromis par chacun des 10 tableaux.

On aimerait retourner aux colonnes (relevés) communes dans une vision propre à chaque tableau. Nous avons deux options.

Un groupe riche (Coléoptères) peut reconstituer l'ordination initiée par un groupe pauvre (Éphéméroptères) sans pour autant la définir. A l'autre extrêmes le groupe le

plus petit (3 Malacostracés) participe fortement à la définition du compromis, sans pouvoir en reconstituer le résultat.

Pour C. Lavit, les trajectoires sur l'axe j sont définies par $\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{Dz}_j = \mathbf{U}_{kk} \mathbf{Dz}_j$, sans que cette représentation ait une valeur optimale. Nous proposons de modifier légèrement cette option en usant de :

$$\frac{\mathbf{X}_k^t \mathbf{D}_k \mathbf{X}_k \mathbf{Dz}_j}{\|\mathbf{X}_k \mathbf{Dz}_j\|} = \frac{\mathbf{U}_{kk} \mathbf{Dz}_j}{\|\mathbf{X}_k \mathbf{Dz}_j\|}$$

D'autre part, nous proposons d'utiliser :

$$\mathbf{X}_k^t (\mathbf{X} \mathbf{D}_k \mathbf{X}^t)^{-1} \mathbf{X}_k \mathbf{Dz}_j = \mathbf{P}_k \mathbf{z}_j$$

soit le modèle des coordonnées du compromis obtenu par régression multiple sur les lignes du tableau \mathbf{X}_k .

On cherche simplement à reconstituer les coordonnées du compromis à l'aide de chaque tableau (figure 12). On peut représenter l'erreur de reconstitution en plaçant sur chaque carte la carte de référence du compromis (figure 13). On note que globalement chaque tableau reconstitue assez correctement la carte du compromis. En particulier, le groupe 5, dont on a déjà discuté, montre une bonne reproduction de cette carte. Cette aptitude est liée au nombre de taxons. Pour le groupe 6 (Diptères), la reconstitution est parfaite puisqu'il y a plus d'espèces (22) que de relevés (16). Avec ce groupe, on peut reconstituer parfaitement n'importe quelle ordination externe. Le tableau 6 est d'inertie élevée (1.381) et de norme relativement faible 0.544). Son poids dans le compromis est limité (0.260). Il peut reproduire la structure du compromis mais contribue faiblement à la définir.

8 — Trajectoires

Notons que quatre type de conditions numériques se présentent dans la figure 14.

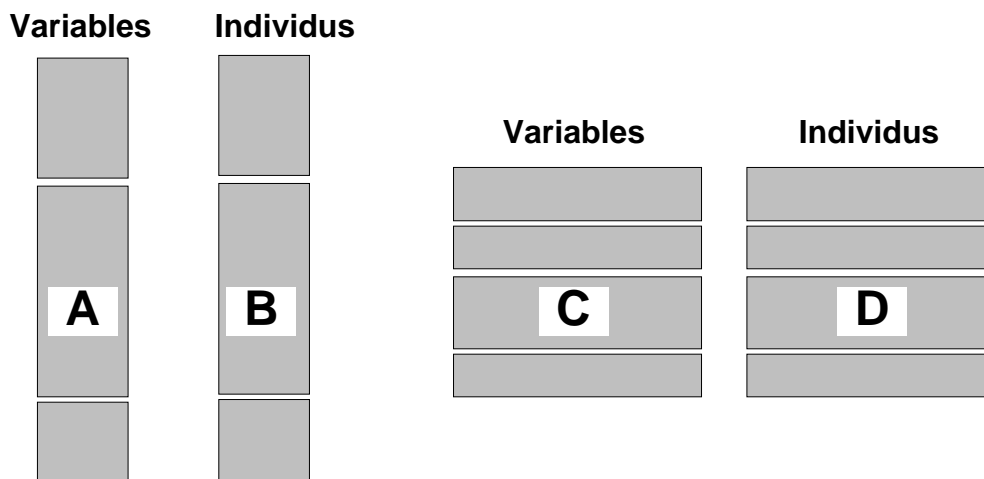


Figure 14 : Situations numériques typiques de K-tableaux. Toutes les situations mixtes sont possibles.

Dans les situations A et B les colonnes du K-tableaux (individus ou variables) sont nettement moins nombreuses que les lignes de chaque tableau. La proposition de reconstitution des coordonnées du compromis par chacun des tableaux est sans intérêt. Cette technique trouve tout son sens dans les situations C et D.

La situation présente est un mélange des cas B et D. Les individus sont en lignes : certains tableaux ont peu de variables (espèces) et d'autres en ont beaucoup. Voir que les 4 Éphéméroptères reproduisent le compromis aussi bien que les 13 coléoptères (pour 16 stations) n'est pas sans intérêt. L'approche par les trajectoires renforce l'idée d'une valeur très inégale de la valeur typologique des groupes. \mathbf{z}_j est le $j^{\text{ème}}$ vecteur \mathbf{D} normé du compromis. On peut s'en servir comme d'un axe (ici une composante principale dans la terminologie classique) de l'ACP du tableau k . $\mathbf{X}_k \mathbf{D} \mathbf{z}_j$ contient les coordonnées des variables du tableau k sur cette composante principale (figure 9). Si \mathbf{z}_j était vraiment une composante principale du tableau (et non un axe du compromis) le vecteur normé porté par $\mathbf{X}_k \mathbf{D} \mathbf{z}_j$ serait un axe principal de l'ACP du tableau k . Notons alors :

$$\mathbf{a}_j^k = \frac{\mathbf{X}_k \mathbf{D} \mathbf{z}_j}{|\mathbf{X}_k \mathbf{D} \mathbf{z}_j|}$$

$\mathbf{X}_k^t \mathbf{D}_k \mathbf{a}_j^k$ contient alors les coordonnées des colonnes du tableau k sur cet axe. S'il était un axe principal $|\mathbf{X}_k^t \mathbf{D}_k \mathbf{a}_j^k|^2$ vaudrait dans ce cas la valeur propre correspondante et l'inertie projetée sur cet axe. Appelons la pseudo-valeur propre associée à \mathbf{z}_j dans le tableau k . Elle exprime la capacité d'un axe du compromis à servir d'axe d'inertie du tableau k . Les normes sont calculées au sens des pondérations de chacun des tableaux.

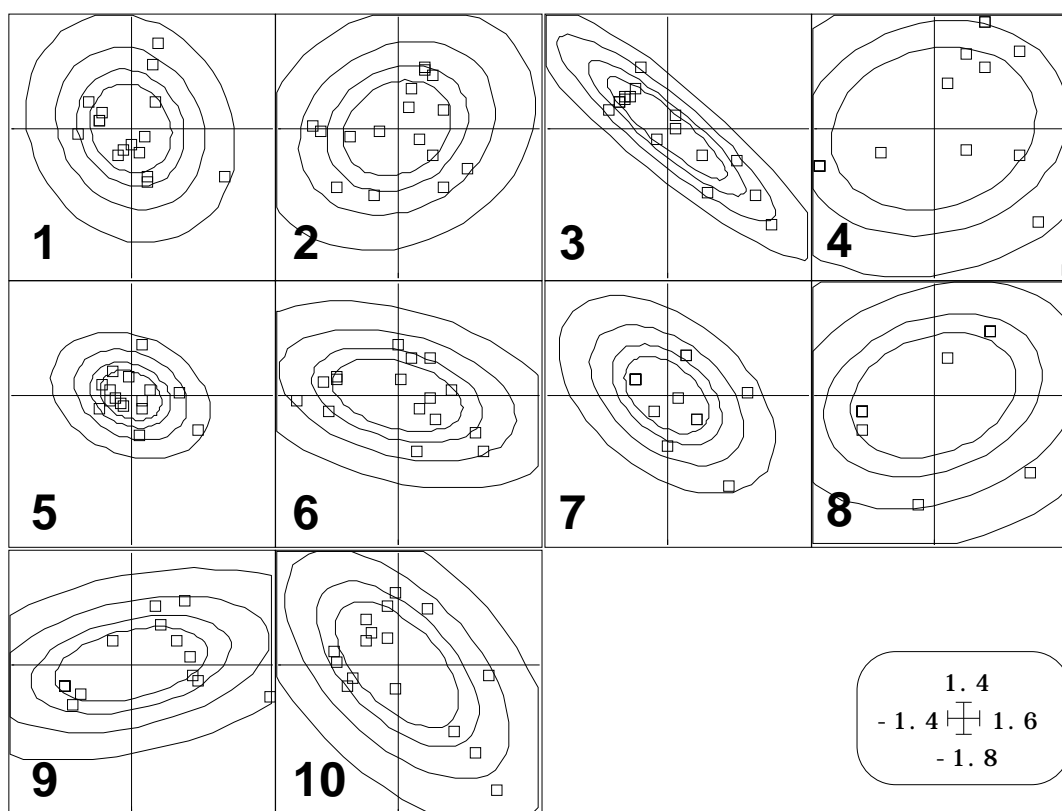


Figure 15 : covariances des coordonnées factorielles dans un plan obtenu avec la technique des trajectoires. Ellipses d'inertie des sous-nuages à 50, 70, 90 et 99%.

Le terme trajectoire est alors discutable. Il invite à considérer la coordonnée que prend la même colonne dans les différents tableaux pour en suivre éventuellement l'évolution. De ce point de vue, aucune propriété d'optimalité, ni en terme de stabilité ni

en terme d'évolution n'est en jeu. Représenter toutes les colonnes d'un même tableau sur un axe a un sens, puisqu'on y voit la typologie des colonnes induites par chaque tableau par son opérateur d'inertie en partant d'un vecteur commun du compromis. Mais utiliser cette représentation à deux dimensions impose de faire l'impasse sur un fait d'importance : les vecteurs \mathbf{a}_j^k pour deux valeurs de j ne sont pas orthogonaux et la représentation cartésienne du nuage n'est pas correcte. On le voit clairement dans la figure 15.

La covariance des coordonnées est induite par la non orthogonalité des \mathbf{a}_j^k . STATIS ne peut donc clairement proposer une représentation canonique des colonnes du K-tableau ajustée à chacun des sous-tableaux. La variance représentée (ou pseudo-valeur propre cumulée) ne peut servir d'indice de valeur typologique.

9 — Valeur typologique d'un groupe faunistique

Il apparaît dans cette première étude que STATIS ne possède pas une manière convaincante de figurer les colonnes d'un sous-tableau. Elle n'est pas construite pour cela. L'essentiel est dans le calcul du compromis, les valeurs propres et les coordonnées colonnes issues du compromis (carte commune à tous les sous-tableaux) et les coordonnées des lignes des sous-tableaux. Manque, à défaut d'une représentation, une indication numérique apte à quantifier la prise en compte du compromis par chacun des tableaux. En revenant à l'essentiel, on peut proposer le calcul suivant.

\mathbf{U}_{kk} est l'opérateur de covariance du tableau k et $\mathbf{U}_{kk}\mathbf{D}$ l'opérateur d'inertie associé. On fait une combinaison de ces opérateurs pour obtenir un opérateur d'inertie \mathbf{UD} optimal. On diagonalise cet opérateur et on en extrait, par le choix des facteurs conservés une partie structurée, qui en est une approximation de rang f (nombre de facteurs conservés) optimale et s'écrit :

$$\hat{\mathbf{U}}\mathbf{D} = \sum_{j=1}^f \sigma_j \mathbf{z}_j \mathbf{z}_j^t \mathbf{D}$$

On peut mesurer l'adéquation de la structure de l'opérateur $\mathbf{U}_{kk}\mathbf{D}$ à celle de $\hat{\mathbf{U}}\mathbf{D}$ en mesurant l'angle que font ces deux opérateurs au sens de la métrique d'Hilbert-Schmidt. Leur produit scalaire s'écrit :

$$\begin{aligned} \text{Trace}(\mathbf{U}_{kk}\mathbf{D}\hat{\mathbf{U}}\mathbf{D}) &= \text{Trace} \mathbf{U}_{kk}\mathbf{D} \sum_{j=1}^f \sigma_j \mathbf{z}_j \mathbf{z}_j^t \mathbf{D} = \sum_{j=1}^f \text{Trace}(\mathbf{U}_{kk}\mathbf{D}\sigma_j \mathbf{z}_j \mathbf{z}_j^t \mathbf{D}) \\ &= \sum_{j=1}^f \sigma_j \text{Trace}(\mathbf{z}_j^t \mathbf{D} \mathbf{U}_{kk} \mathbf{D} \mathbf{z}_j) = \sum_{j=1}^f \sigma_j \|\mathbf{X}_k \mathbf{D} \mathbf{z}_j\|^2 \end{aligned}$$

D'autre part :

$$\text{Trace}(\hat{\mathbf{U}}\mathbf{D}\hat{\mathbf{U}}\mathbf{D}) = \text{Trace} \sum_{j=1}^f \sigma_j \mathbf{z}_j \mathbf{z}_j^t \mathbf{D} \sum_{k=1}^f \sigma_k \mathbf{z}_k \mathbf{z}_k^t \mathbf{D} = \sum_{j=1}^f \sigma_j^2$$

D'où, en notant λ_j^k la $j^{\text{ième}}$ valeur propre du $k^{\text{ième}}$ sous-tableau :

$$\text{Cos}^2(\alpha_k) = \frac{\sum_{j=1}^f \sigma_j \|\mathbf{X}_k \mathbf{Dz}_j\|^2}{\sum_{j=1}^f \sigma_j^2 \sum_{j=1}^p (\lambda_j^k)^2}$$

Tout se passe comme si les valeurs propres de l'analyse du compromis et les valeurs propres de l'analyse des tableaux étaient des composantes de vecteurs, la substitution au numérateur de $\|\mathbf{X}_k \mathbf{Dz}_j\|^2$ à λ_j^k permettant de mesurer exactement ce que l'axe j du compromis est pour le tableau k .

Nous conserverons cette valeur comme mesure synthétique de la valeur typologique du groupe k , c'est-à-dire comme mesure d'adéquation de l'ACP du groupe k à l'ACP de synthèse qui résume toutes les autres.

Le groupe des éphéméroptères est définitivement isolé de cette manière. L'analyse effectuée se résume à la figure 16 qui reprend l'essentiel de la fiche. La valeur typologique d'un groupe faunistique est sans commune mesure avec le nombre de taxa qui le représente.

1- Hemiptera	0. 213	6- Diptera	0. 408
2- Odonata	0. 343	7- Hydracarina	0. 137
3- Trichoptera	0. 355	8- Malacostraca	0. 517
4- Ephemeroptera	0. 929	9- Mollusca	0. 474
5- Coleoptera	0. 106	10- Oligochaeta	0. 166

Annexe : Pour refaire les calculs

Utiliser la carte Friday de la pile ADE-4•Data. Récupérer les fichiers Blo.txt (binaire Blo, 10-1) Fau.Car (binaire Fau 91-16 par Text-Binary, option CharToBin), Label Esp, Label Sta et Label StaTcc. Préparer une structure de fichiers associés à un multi-tableau.

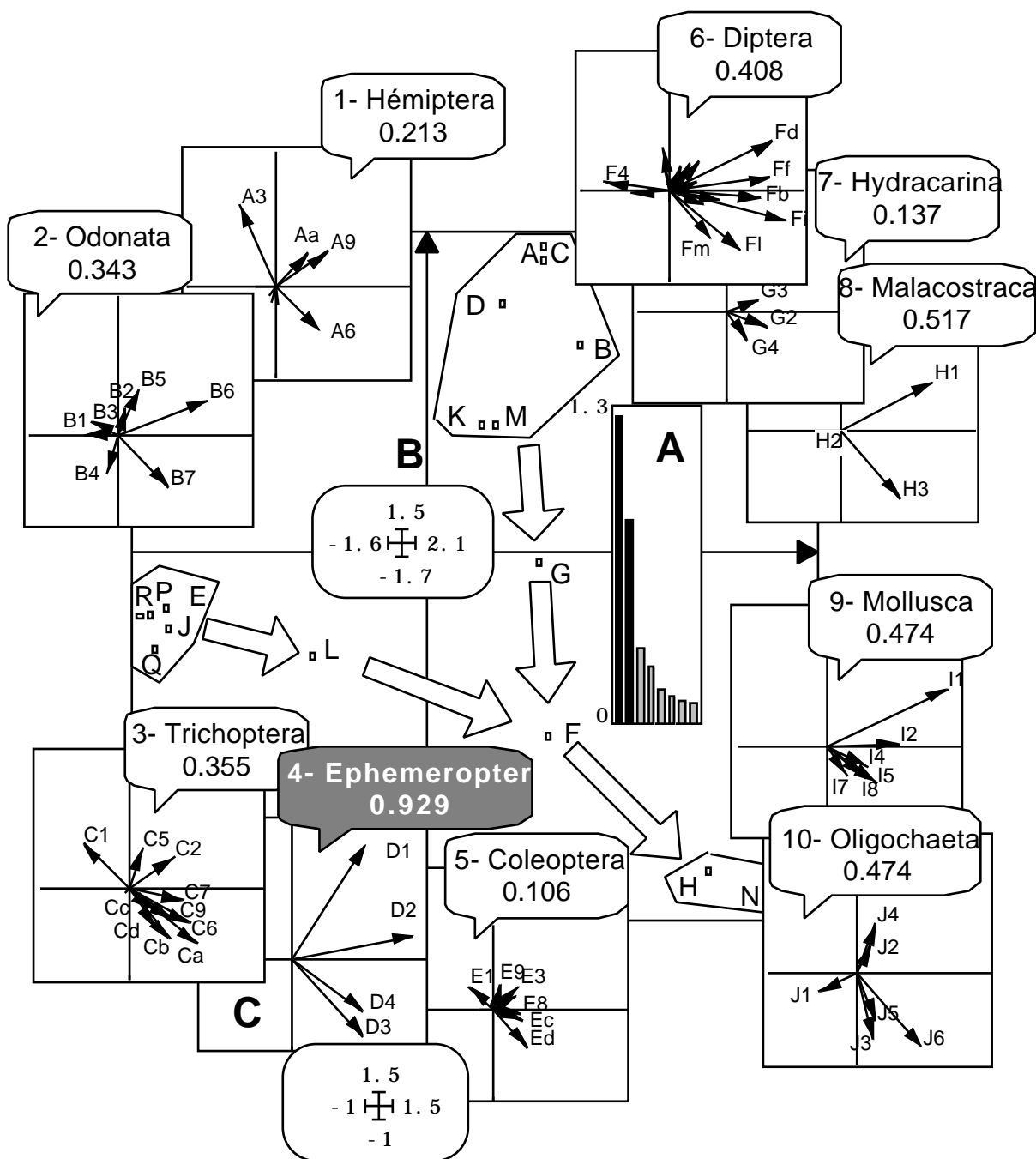


Figure 16 : Résumé du fonctionnement de STATIS. A - valeurs propres du compromis. B - Carte factorielle de synthèse des relevés. C - Carte des espèces par groupes et valeur typologique mesurée par le Cosinus au carré de l'opérateur associé avec le modèle de l'opérateur compromis.

La première question n'est posée que pour récupérer le nombre de lignes de Fau. Le fichier Blo est un fichier binaire qui contient pour chaque sous-tableau le nombre des lignes associées. On obtient 13 fichiers utilitaires qui sont utiles dans toutes les représentations graphiques associées aux méthodes multi-tableaux qui portent sur tous les fichiers d'un type compatible avec Fau. Le nom générique Fau! est choisi par défaut de réponse à la quatrième question.

```

Fau!.ktp
Fau!T4a
Fau!T4a.123
Fau!T4a.cat
Fau!T4a.label
Fau!TCc
Fau!TCc.123
Fau!TCc.cat
Fau!TCc.label
Fau!TLl
Fau!TLl.123
Fau!TLl.cat
Fau!TLl.label

```



```

Option: InitKTab
Input file: Fau
Row indicator file: Blo
Col indicator file:
Output ASCII file: Fau!.ktp

10: 11/7/13/4/13/22/4/3/8/6/
1: 16/

-----
Qualitative variable file: Fau!TLl
Number of rows: 91, variables: 3, categories: 123
Auxiliary ASCII output file Fau!TLl.label: label file
-----
Qualitative variable file: Fau!TCc
Number of rows: 160, variables: 3, categories: 42
Auxiliary ASCII output file Fau!TCc.label: label file
-----
Qualitative variable file: Fau!T4a
Number of rows: 40, variables: 2, categories: 14
Auxiliary ASCII output file Fau!T4a.label: label file
-----

```

Ceci signifie qu'on manipule un tableau avec 10 blocs de lignes contenant respectivement 11, 7, ..., 6 éléments et 1 bloc de colonnes contenant 16 éléments.

Le découpage des lignes en 10 blocs fournit un tableau à 3 variables qualitatives Fau!TLl (91 lignes, 3 variables, 123 modalités. Variable 1 : 10 modalités (numéro du tableau auquel appartient la ligne), variables 2 : 91 modalités (numéro de la ligne dans la numérotation complète, indication qui n'a de pertinence que dans un multi-tableau double avec bloc de lignes et blocs de colonnes, ce qui n'est pas le cas ici), variable 3 : 22 modalités (numéro d'ordre de la ligne dans le sous-tableau auquel elle appartient). Fau!TLl.cat est le descriptif de ce fichier qui est demandé par les programmes graphiques (.cat file). Fau!TLl.label est un fichier d'étiquettes des lignes du multi-tableau sous la forme numéro de tableau - numéro de lignes par tableau. Fau!TLl.123 est un fichier d'étiquettes des modalités parfois utilisé dans les programmes graphiques par un appel automatique. On peut se contenter de mémoriser le procédé d'appellation **TLl** qui signifie numéro de **T**ableau, numéro de **L**igne en général, numéro de ligne par tableau.

Le découpage des colonnes en 1 bloc fournit un tableau à 3 variables qualitatives Fau!TCc (160 lignes, 3 variables, 42 modalités. Variable 1 : 10 modalités (numéro du tableau auquel appartient la colonne), variables 3 : 16 modalités (numéro de la colonne dans la numérotation complète, indication qui n'a de pertinence que dans un multi-tableau doubles avec bloc de lignes et blocs de colonnes, ce qui n'est pas le cas ici), variable 3 : 16 modalités (numéro d'ordre de la colonne dans le sous-tableau auquel elle appartient). Fau!TCc.cat est le descriptif de ce fichier qui est demandé par les programmes graphiques (.cat file). Fau!TCc.label est un fichier d'étiquettes des colonnes du multi-tableau sous la forme numéro de tableau - numéro de colonnes par tableau. Fau!TCc.123 est un fichier d'étiquettes des modalités parfois utilisé dans les programmes graphiques par un appel automatique. On peut se contenter de mémoriser le procédé d'appellation **TCc** qui signifie numéro de **T**ableau, numéro de **C**olonne en général, numéro de colonne par tableau.

La présence de 10 tableaux fournit un tableau à 3 variables qualitatives Fau!T4a (40 lignes, 2 variables, 14 modalités. Variable 1 : 10 modalités (numéro du tableau auquel appartient la ligne), variables 2 : 4 modalités (numéro d'ordre la ligne dans la numérotation de 1 à 4). Fau!T4a.cat est le descriptif de ce fichier qui est demandé par les programmes graphiques (.cat file). Fau!T4a.label est un fichier d'étiquettes des lignes du multi-tableau sous la forme numéro de tableau - numéro de lignes par tableau. Fau!T4a.123 est un fichier d'étiquettes des modalités parfois utilisé dans les programmes graphiques par un appel automatique. On peut se contenter de mémoriser le procédé d'appellation T4a qui signifie numéro de Tableau, numéro d'ordre de 4 axes. La figure 17 explicite cette opération préliminaire un peu compliquée mais fort utile par la suite.

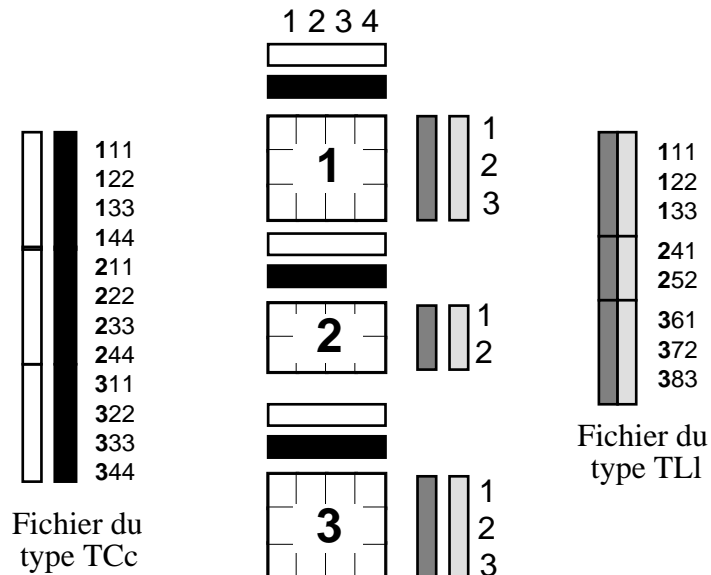
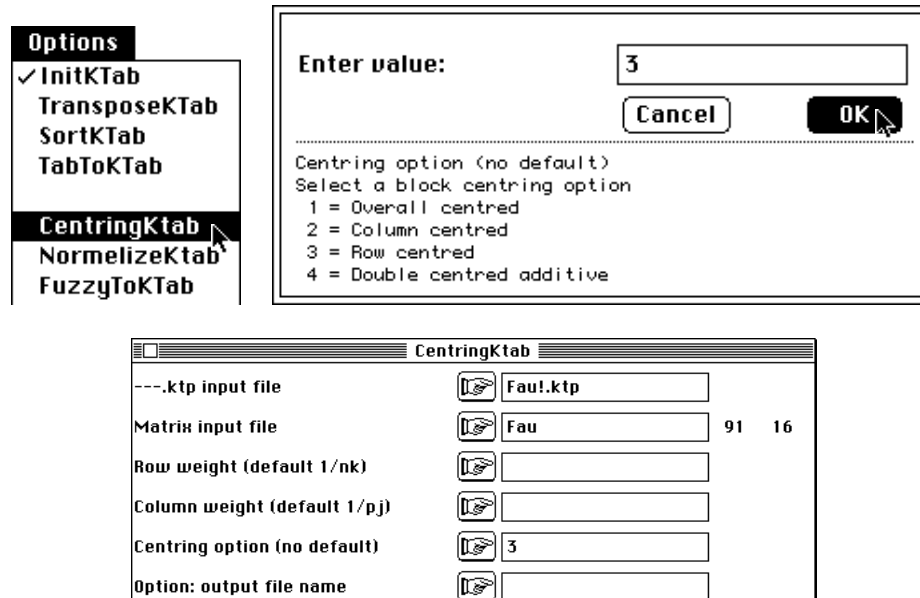
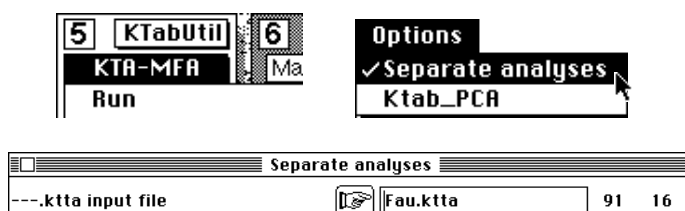


Figure 17 : Pour stocker des coordonnées des lignes et des colonnes de chaque sous-tableau, on utilise des fichiers du type Tableau-Ligne-ligne ou Tableau-Colonne-colonne. Le suffixe TLI ou TCc est utilisé systématiquement pour signaler leur contenu. Les indicatrices permettent toutes les combinaisons de graphiques.

L'opération du centrage des données est essentielle. Le centrage par espèces s'impose dans les tableaux faunistiques. Ici, les relevés sont en lignes. Dans le même module, utiliser :



On obtient un triplet statistique standard, avec une pondération uniforme par bloc de lignes, une pondération uniforme pour les colonnes et un tableau centré par taxon (ligne). On dispose d'un multi-tableau prêt à l'emploi. On peut exécuter toutes les analyses séparées par :



```

Row bloc: 1 - Col bloc: 1 - Total inertia: 0.683239
Num. Eigenval.   R. Iner.   R. Sum   | Num. Eigenval.   R. Iner.
R. Sum |
01  +2.4127E-01 +0.3531 +0.3531 | 02  +1.8054E-01 +0.2642
+0.6174 |
03  +9.2671E-02 +0.1356 +0.7530 | 04  +7.7108E-02 +0.1129
+0.8659 |
...

```

```

-----
Row bloc: 10 - Col bloc: 1 - Total inertia: 1.4043
Num. Eigenval.   R. Iner.   R. Sum   | Num. Eigenval.   R. Iner.
R. Sum |
01  +5.3600E-01 +0.3817 +0.3817 | 02  +4.2840E-01 +0.3051
+0.6867 |
...

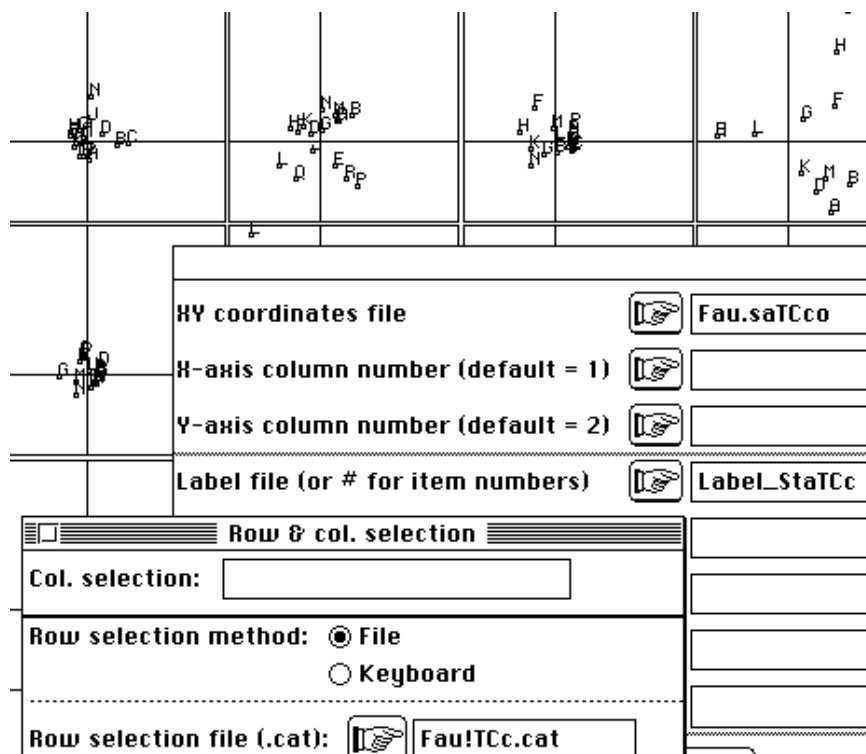
```

```

-----
Eigenvalues are stored by column into file Fau.savp
Row : 16 - Col (tables): 10
File Fau.saTlli contains row scores in separate analyses
It has 91 rows and 4 columns
It is to be used with --Tll.label and --Tll.cat files
File Fau.saTcco contains column scores in separate analyses
It has 160 rows and 4 columns
It is to be used with --Tcc.label and --Tcc.cat files
File Fau.sapa contains total inertia and rank for each
separate analyses
It has 10 rows and 2 columns

```

On reconnaît le déroulement de 10 analyses d'inertie. Le fichier Fau.savp et le module Curves donne la figure 5. Le module Scatters, option Labels sur les fichiers Fau.Tcco et Label_StaTcco, trié par Fau!Tcco.cat donne la figure 7 :



Pour obtenir les résultats principaux de la fiche, utiliser :



File Fau. oa+vp contains the eigenvalues of compromise analysis
It has 16 rows and 1 columns

Utiliser Curves pour représenter les valeurs propres (figure 8).

File Fau. oaTl1 contains standard column scores with lambda norm
It has 91 rows and 2 columns

It is to be used with --Tl1.label and --Tl1.cat files
File : Fau. oaTl1

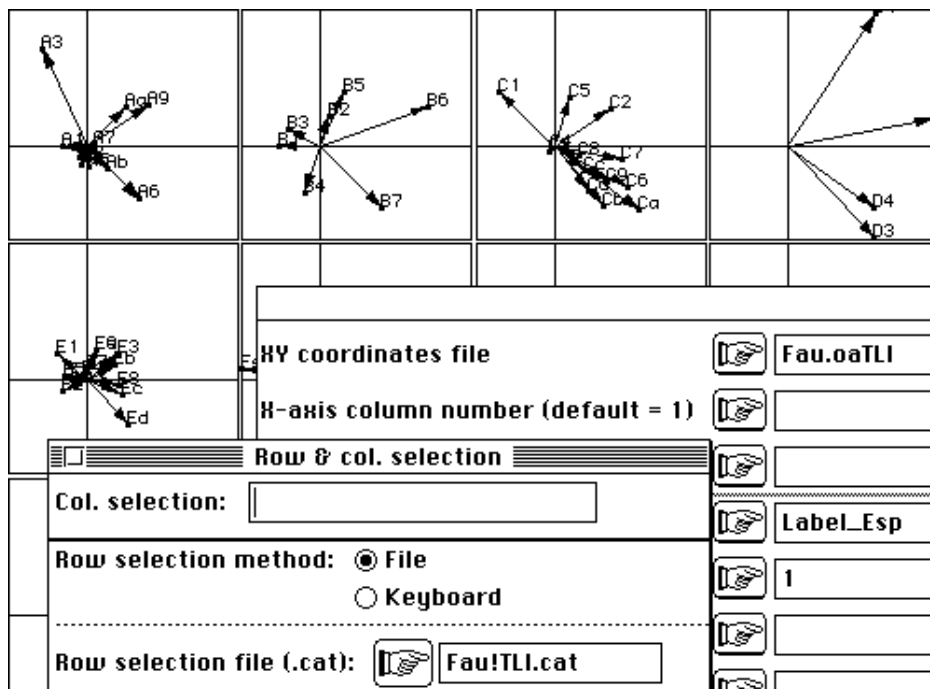
----- Minimum/Maximum:
Col.: 1 Mini = -0.7252 Maxi = 1.372
Col.: 2 Mini = -0.87256 Maxi = 1.2558

Utiliser Scatters pour obtenir la figure 9.

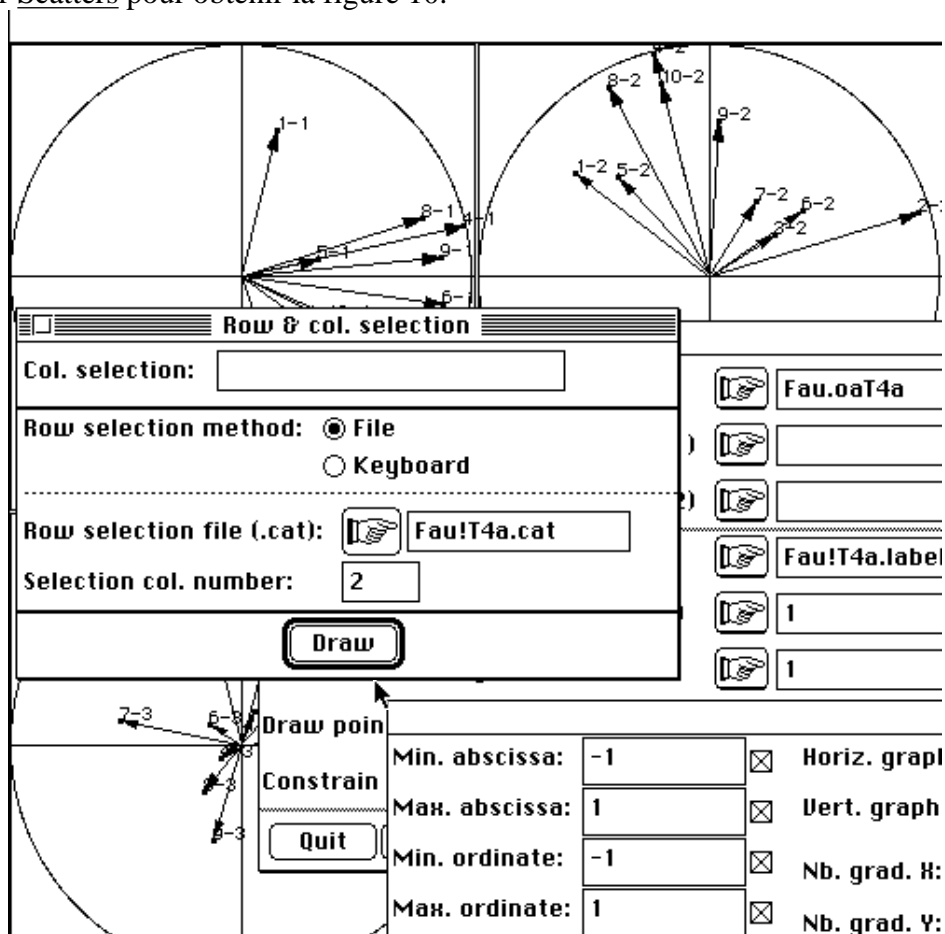
File Fau. oaT4a contains scores of axes of separate analysis
It has 40 rows and 2 columns
It is to be used with --T4a.label and --T4a.cat files
File : Fau. oaT4a

----- Minimum/Maximum:

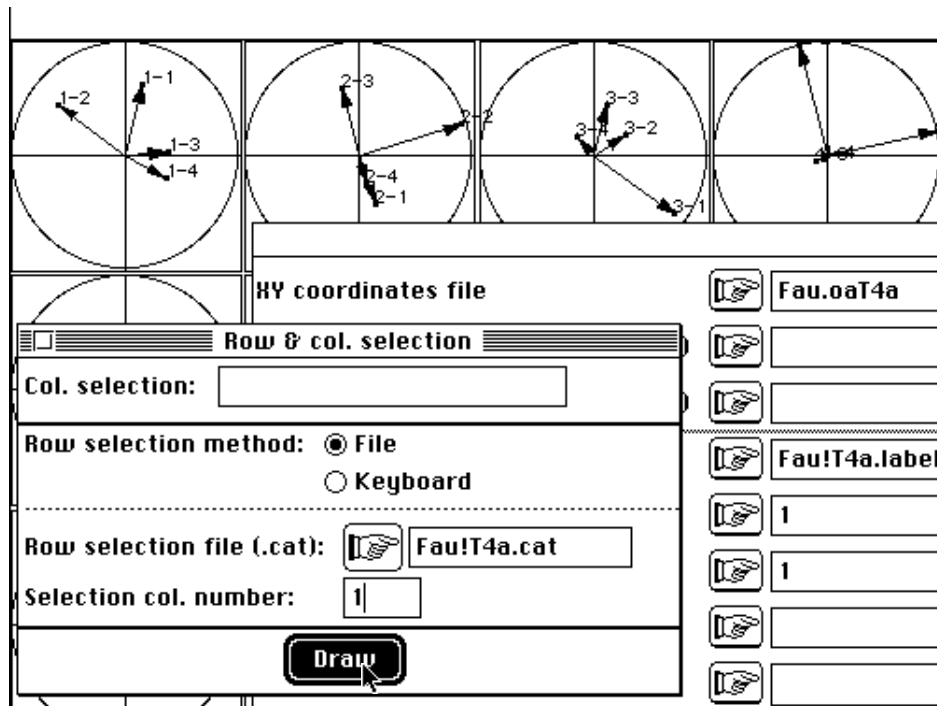
Col.: 1 Mini = -0.57936 Maxi = 0.95846
 Col.: 2 Mini = -0.51114 Maxi = 0.95477



Utiliser Scatters pour obtenir la figure 10.



Changer un paramètre pour obtenir la figure 11.



File Fau. oa+co contains column scores (analysis of the compromise)

It has 16 rows and 2 columns

File : Fau. oa+co

----- Minimum/Maximum:

Col.: 1 Mini = -1.5449 Maxi = 2.019

Col.: 2 Mini = -1.6343 Maxi = 1.4285

Utiliser Scatters sur Fau. oa+co pour obtenir la figure 8.

File Fau. oaTCrep contains column scores (analysis of the compromise)

Duplication of general scores

It has 160 rows and 2 columns

It is to be used with --Tcc.label and --Tcc.cat files

File : Fau. oaTCrep

----- Minimum/Maximum:

Col.: 1 Mini = -1.5449 Maxi = 2.019

Col.: 2 Mini = -1.6343 Maxi = 1.4285

File Fau. oaTCest contains columns scores issued from each table by

It has 160 rows and 2 columns

It is to be used with --Tcc.label and --Tcc.cat files

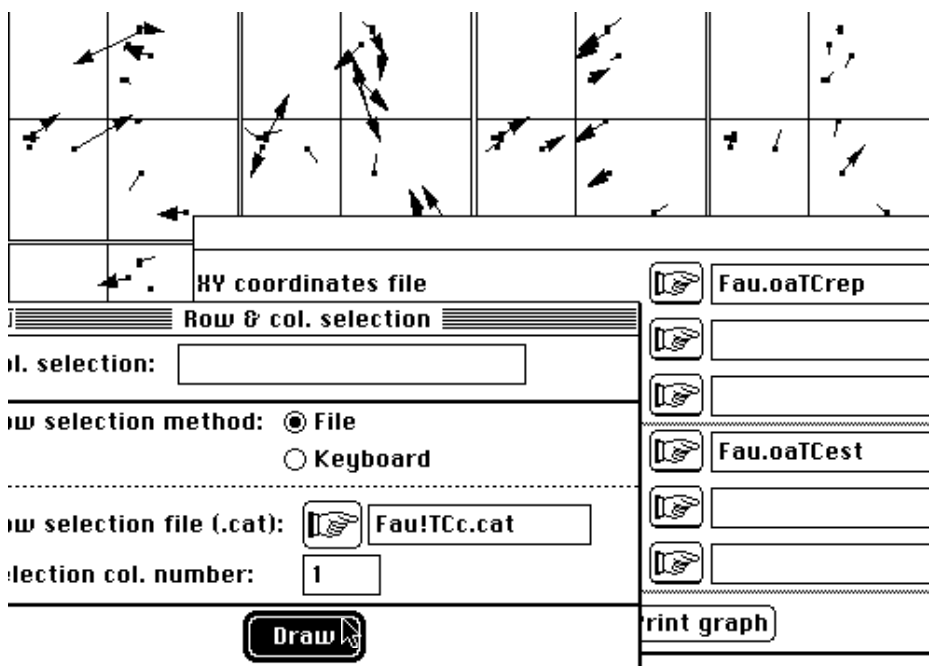
File : Fau. oaTCest

----- Minimum/Maximum:

Col.: 1 Mini = -1.8247 Maxi = 2.4203

Col.: 2 Mini = -1.8821 Maxi = 1.6897

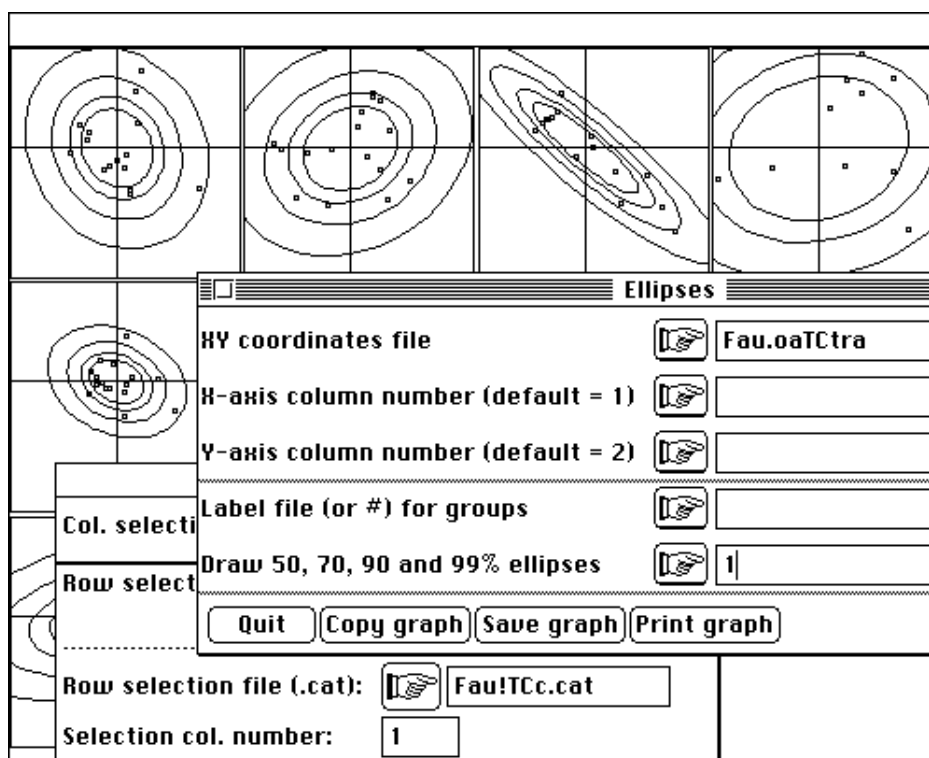
Utiliser Scatters, option Match two scatters pour obtenir la figure 13.



File Fau.aaTCtra contains columns scores issued from each table by
 Modified Status trajectories
 It has 160 rows and 2 columns
 It is to be used with --Tcc.label and --Tcc.cat files
 File :Fau.aaTCtra

----- Minimum/Maximum:
 Col.: 1 Mini = -1.3153 Maxi = 1.5948
 Col.: 2 Mini = -1.7037 Maxi = 1.3168

Utiliser Scatters, option Ellipses pour obtenir la figure 15.



Une dernière remarque concerne l'option de départ :

1 = Option COVV (default = RV)



En utilisant l'option par défaut, on obtient un résultat voisin :

Number	Rows	Weights	NS norm2	Cos2
1	11	3.051e-01	1.072e-01	0.271
2	7	3.256e-01	2.317e-01	0.380
3	13	3.406e-01	1.794e-01	0.433
4	4	3.407e-01	1.507e+00	0.790
5	13	2.845e-01	4.273e-02	0.135
6	22	3.529e-01	2.957e-01	0.426
7	4	2.839e-01	8.614e-02	0.186
8	3	3.175e-01	6.430e-01	0.578
9	8	3.393e-01	3.621e-01	0.543
10	6	2.583e-01	5.361e-01	0.173

Ceci est dû au fait que l'inertie initiale des différents tableaux est du même ordre de grandeur, du moins que l'inertie maximale (tableau 4) correspond à une structure maximale. Ce n'est pas toujours le cas, une inertie très importante pouvant correspondre à une structure faible pour des raisons numériques. La méthode aura tendance à prendre comme référence un tableau sans intérêt si on utilise, dans ce cas, l'option COVV. C'est pourquoi l'option RV est proposée par défaut. Entre COVV et RV on raisonne comme entre une ACP centrée (les variances jouent un rôle déterminant) ou normée (le rôle des variances est éliminé). Si plusieurs tableaux ont une inertie totale qui dépend des unités de mesures, l'option RV s'impose.

Remerciements : Merci à E. Baran et S. Champely de leurs commentaires fructueux sur la première version de cette fiche.

Références

¹ L'Hermier des Plantes, H. (1976) *Structuration des tableaux à trois indices de la statistique. Théorie et applications d'une méthode d'analyse conjointe*. Thèse de 3^o cycle, USTL, Montpellier.

² Lavit, Ch. (1988) *Analyse conjointe de tableaux quantitatifs*. Masson, Paris. 1-240.

³ Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1994) The ACT (Statis method). *Computational Statistics and Data Analysis* : 18, 97-119.

⁴ Escoufier, Y. (1973) Le traitement des variables vectorielles. *Biometrics* : 29, 750-760.

Escoufier, Y. (1976) Opérateur associé à un tableau de données. *Annales de l'INSEE* : 22-23, 165-178.

Robert, P. & Escoufier, Y. (1976) A unifying tool for linear multivariate statistical methods : the RV coefficient. *Applied Statistics* : 25, 3, 257-265.

Escoufier, Y. (1977) Operators related to a data matrix. In : *Recent developments in Statistics*. Barra, J.R. & Coll. (Eds.) North-Holland. 125-131.

Escoufier, Y. (1980) L'analyse conjointe de plusieurs matrices de données. In : *Biométrie et Temps*. Jolivet, M. (Ed.) Société Française de Biométrie, Paris. 59-76.

⁵ Amanieu, M., Guelorget, O. & Nougier-Soule, J. (1981) Analyse de la diversité de la macrofaune benthique d'une lagune littorale méditerranéenne. *Vie Milieu* : 31, 303-312.

⁶ Friday, L.E. (1987) The diversity of macroinvertebrate and macrophyte communities in ponds. *Freshwater Biology* : 18, 87-104.

⁷ Place, M.C. (1980) *Contribution algorithmique à la mise en œuvre de la méthode STATIS*. Thèse de troisième cycle, Université de Montpellier II.