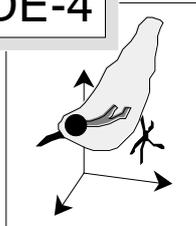


ADE-4



ACOM et mesure de la valeur typologique

Résumé

L'Analyse de co-inertie multiple (ACOM) généralise à K tableaux l'analyse de co-inertie pour deux tableaux. On la compare ici à STATIS en testant sa capacité à mesurer la valeur typologique des groupes faunistiques. Des logiques d'analyse distinctes conduisent à des résultats cohérents.

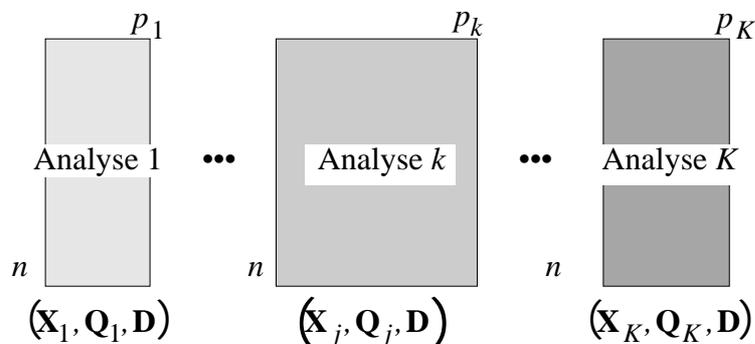
Plan

1 — Exécution de l'analyse	2
2 — Le nombre d'axes à conserver	3
3 — Comparaison des compromis	7
4 — Une nouvelle mesure de la valeur typologique.....	9
5 — Représentation des groupes et des espèces	12
Références	14

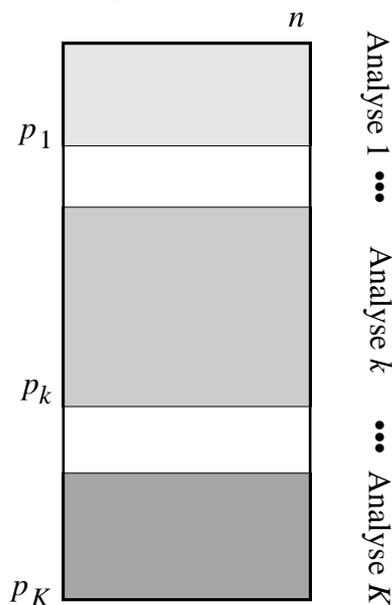
M. Hanafi & D. Chessel

1 — Exécution de l'analyse

Les données traitées sont strictement celles du tableau 1 de la fiche ACT-Statist 1 (fascicule 5). Elles sont publiées par Friday ¹. On dispose de K tableaux ayant en commun les lignes (16 étangs), les colonnes étant les taxa répartis en 10 groupes faunistiques. On dispose donc de K triplets statistiques $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$ ($1 \leq k \leq K$). Ceci est une présentation conceptuelle :



Concrètement, les tableaux sont rangés verticalement :



On décide de faire 10 ACP simultanées par l'exécution des options :

InitKTab			
Matrix input file	<input type="text" value="Fau"/>	91	16
Row indicator (Default = 1 class)	<input type="text" value="Blo"/>	10	1
Col indicator (Default = 1 class)	<input type="text"/>		
Option: output file name	<input type="text"/>		

CentringKtab		
---.ktp input file		Fau!.ktp
Matrix input file		Fau
Row weighting (default 1/nk)		
Column weighting (default 1/pj)		
Centring option (no default)		3
Option: output file name		

Les analyses séparées se font par :

Separate analyses		
---.kttta input file		Fau.kttta

Les valeurs propres des analyses séparées comme les cartes relevés sont décrites dans la fiche citée. On choisit alors de ne pas pondérer les tableaux en laissant faire l'hétérogénéité des inerties. En effet l'inertie d'un tableau mesure la variabilité inter-stationnelles des descripteurs et il y aurait un effet dommageable à uniformiser l'hétérogénéité des tableaux pour ensuite mesurer la valeur typologique de ses descripteurs. On sait que la structure commune (compromis) est de dimension 2. On demande 4 axes dans l'ACOM pour voir ce qui va se passer :

ACOM		
---.kttta input file		Fau.kttta
Table weights (default = 1/Inertia)		2
Selected axis number (default = 8)		4

2 — Le nombre d'axes à conserver

```

Generalised Ktab-PCA analysis
Input file Fau
Row number: 91, Column number: 16, Block number: 10
Selected weighting option
  -> Uniform weights
vp level 1=3.1366e-01
vp level 1=1.8504e-01
vp level 1=7.8760e-02

```

Logiquement, il y a dans une analyse de co-inertie multiple au maximum r facteurs possibles, si r est le minimum du rang des sous-tableaux :

```

Input file: Fau.sapa
Row: 10 Col: 2
  1 | 0.6832 | 11.0000 |
  2 | 0.9994 | 7.0000 |
  3 | 0.8705 | 13.0000 |
  4 | 1.8076 | 4.0000 |
  5 | 0.5063 | 12.0000 |
  6 | 1.3814 | 15.0000 |
  7 | 0.5576 | 4.0000 |
  8 | 1.1823 | 3.0000 |
  9 | 1.0068 | 8.0000 |
 10 | 1.4043 | 6.0000 |

```

Le tableau 8 est de rang 3 (ACP avec 16 relevés et 3 espèce), il y donc au maximum trois axes de co-structure. Il s'agit là d'une différence sensible entre ACOM et STATIS. Par la notion de coinertie, il n'y a pas de coinertie sans inertie (pas plus qu'il n'y a de

covariance entre variables sans variance pour chacune d'elles). On peut alors se poser une question extrême. Que se passe-t'il dans ces deux méthodes *quand il n'y a qu'une variable par groupe*. Une option de KtabUtil permet de faire cette expérience (dont les mathématiques indiquent le résultat, bien sûr !). Prenons le fichier Meteo (23-7) qui a servi à tant d'illustrations et transposons le tableau Meteo dans X (7-23) :

Transpose		
Input file		Meteo 23 7
Output file		X

Dans KTabUtil :

TabToKtab		
Input file		X 7 23
Option: TabToKtab		
Input file: X		
-> Rows: 7, columns: 23		
-> Content: 7 tables with 1 rows and 23 columns		

On obtient à partir d'un tableau à 23 lignes et 7 colonnes un *K*-tableaux de 7 tableaux à 1 colonne et 23 lignes (modèle théorique) rangés sous forme de 7 lignes et 23 colonnes (modèle pratique). On peut normaliser par tableau (à une seule variable), c'est-à-dire normaliser tout court (mais sur une ligne, unique variable d'un tableau) :

NormalizeKtab		
---.ktp input file		X!.ktp
Matrix input file		X 7 23
Row weighting (default 1/nk)		
Column weighting (default 1/pj)		
Row-Col option (no default)		2
Option: output file name		Y

On compare alors l'ACOM de ces 7 tableaux d'ACP normée à une colonne :

ACOM		
---.ktt input file		Y.ktt 7 23
Table weights (default = 1/Inertia)		
Selected axis number (default = 8)		

et STATIS sur la même configuration :

Operator averaging		
---.ktt input file		Y.ktt 7 23
1 = Option COUV (default = RV)		

L'ACP normée elle-même de Meteo sert de référence :

Correlation matrix PCA		
Matrix input file		Meteo 23 7

Cette ACP affiche 7 valeurs propres :

Total inertia: 7							
Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+4. 2665E+00	+0. 6095	+0. 6095	02	+2. 0534E+00	+0. 2933	+0. 9028
03	+2. 9745E-01	+0. 0425	+0. 9453	04	+1. 9896E-01	+0. 0284	+0. 9738
05	+1. 1449E-01	+0. 0164	+0. 9901	06	+4. 3129E-02	+0. 0062	+0. 9963
07	+2. 6047E-02	+0. 0037	+1. 0000				

STATIS en donne 23, car elle diagonalise la matrice compromis des 7 matrices de produits scalaires (23-23). Ces valeurs propres sont nulles à partir de la huitième :

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+1. 6544E+00	+0. 6309	+0. 6309	02	+7. 1400E-01	+0. 2723	+0. 9032
03	+1. 1638E-01	+0. 0444	+0. 9475	04	+6. 6620E-02	+0. 0254	+0. 9730
05	+4. 5147E-02	+0. 0172	+0. 9902	06	+1. 5946E-02	+0. 0061	+0. 9963
07	+9. 8318E-03	+0. 0037	+1. 0000	08	+0. 0000E+00	+0. 0000	+1. 0000
09	+0. 0000E+00	+0. 0000	+1. 0000	10	+0. 0000E+00	+0. 0000	+1. 0000
...							

L'ACOM n'en donne qu'une, qui est exactement le taux d'inertie de la première de l'ACP :

Selected weighting option
 -> Total inertia from separate analysis
 vp level 1 = 6.0950e-01

L'analyse indique fortement la nature unicolonne des tableaux, ce qui n'est pas le cas de STATIS :

Row bloc: 1 - Col bloc: 1 - Total inertia: 1							
Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+1. 0000E+00	+1. 0000	+1. 0000				

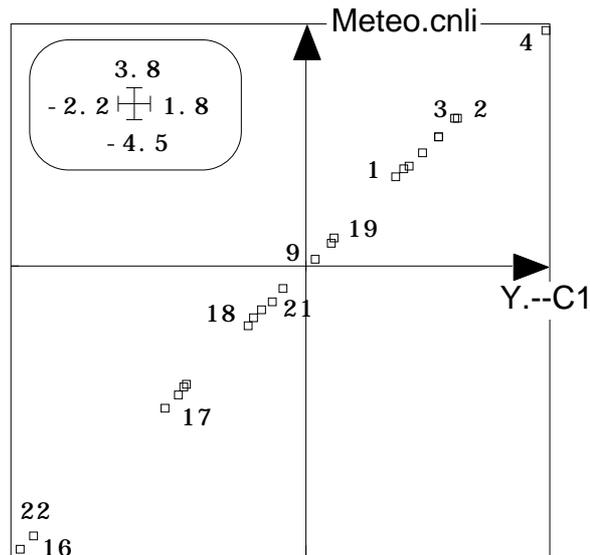
Noter, également, que si la matrice des corrélations de l'ACP normée vaut :

```
----- Correlation matrix -----
[ 1] 1000
[ 2]  827 1000
[ 3]  913  783 1000
[ 4]  597  822  724 1000
[ 5]   -6 -433  -34 -495 1000
[ 6] -254 -638 -271 -637  778 1000
[ 7]  -94 -504 -117 -497  841  934 1000
```

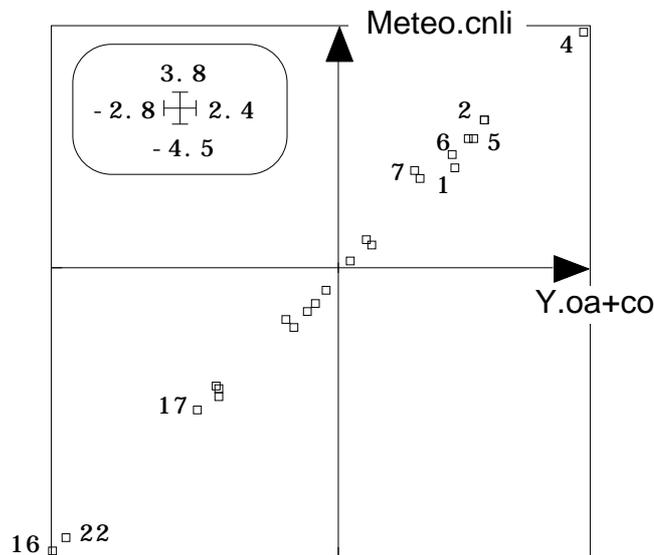
celle des corrélations vectorielles entre tableaux (RV dans STATIS) est formée des carrés des corrélations ordinaires qui précèdent :

```
----- Correlation matrix -----
[ 1] 1000
[ 2]  684 1000
[ 3]  834  612 1000
[ 4]  357  676  525 1000
[ 5]   0  188   1  245 1000
[ 6]  64  408   74  406  606 1000
[ 7]   9  254  14  247  707  872 1000
```

On vérifiera que la composante de synthèse de l'ACOM est à une constante près la première coordonnées de l'ACP normée (le premier pas de l'ACOM est identique à l'ACP elle-même mais ne garde qu'un axe et il n'y a pas de deuxième pas) :



La coordonnée de synthèse dans STATIS est très proche des précédentes sans leur être mathématiquement liée :



Sur un tableau vu comme multitableau à composantes unicolonnes, les relations sont donc étroites entre les points de vue mais les principes de fonctionnement distincts n'autorisent pas à confondre les méthodes. STATIS est indépendant de la géométrie des nuages par sous-tableau alors que l'ACOM ne dépend que de cette géométrie. C'est pourquoi la représentation des individus par tableau dans STATIS est un échec mais dans l'ACOM se passe bien. C'est aussi pourquoi l'ACOM ne peut reconnaître une structure multidimensionnelle avec des tableaux unidimensionnels.

Il est donc logique de prévoir le nombre d'axes conservé dans l'ACOM à partir des analyses séparées, en sachant qu'il est exclu de conserver plus d'axes que ne pourrait en définir une seule des analyses de base.

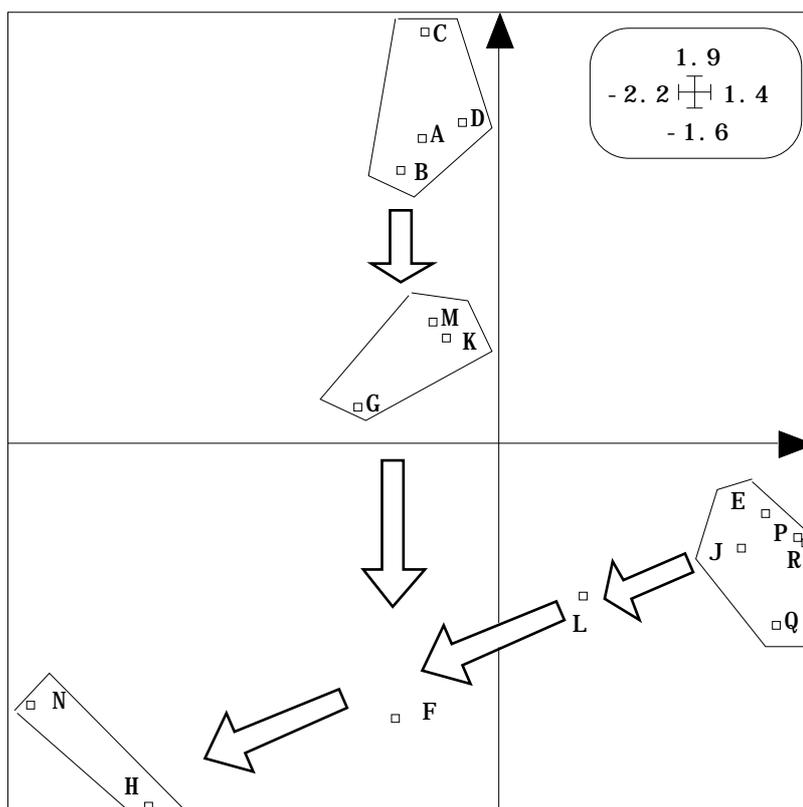
En ce sens STATIS et ACOM ne sont pas comparables au sens strict et il convient de ne pas espérer dire "celle-ci est meilleure que celle-là", réflexe assez répandu, qui a l'avantage de limiter la difficulté en pensant avoir une bonne méthode à tout faire.

On pourrait faire encore plusieurs remarques sur ce cas d'intérêt théorique. On ajoutera simplement que la coordonnée d'un tableau sur l'axe 1 de l'ACOM est exactement la coordonnée de la variable dans l'ACP normée mais que l'analyse ne donne qu'une coordonnée à un tableau qui n'a qu'une dimension. STATIS donne aussi des coordonnées aux tableaux dans sa version classique, mais nous avons décidé de ne pas reproduire cette pratique qui induit une question délicate entre la notion de compromis (moyenne de structures) et celle de typologie d'opérateurs (variabilité entre structures).

3 — Comparaison des compromis

Le compromis porte sur les colonnes du *K*-tableaux pratique, donc sur les lignes du *K*-tableaux théorique (relevés). Nous avons conservé des codes normalisés (moyennes nulles et variances unité).

```
File Fau.--C1 contains column scores (unit norm)
It has 16 rows and 3 columns
File : Fau.--C1
----- Minimum/Maximum:
Col.: 1 Mini = -2.1035 Maxi = 1.3698
Col.: 2 Mini = -1.58 Maxi = 1.8081
```



Ces coordonnées sont dupliquées dix fois (10 tableaux) dans :

```
File Fau.--TC1rep contains column scores (unit norm)
Duplication of the previous file
It has 160 rows and 3 columns
It is to be used with --Tcc.label and --Tcc.cat files
```

Chaque tableau fabrique son propre système de coordonnée comparable au précédent :

File Fau.--TC1 contains normalized column scores
 from projections of separate clouds
 It has 160 rows and 3 columns
 It is to be used with --TCc.label and --TCc.cat files
 It is to be matched with Fau.--TC1rep files
 File : Fau.--TC1

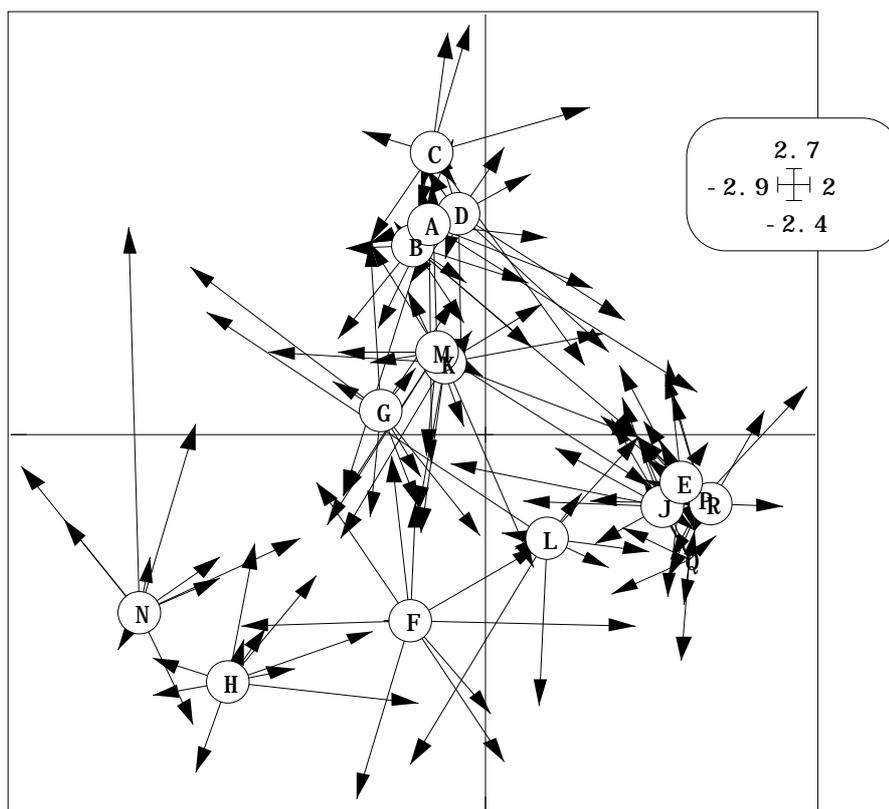
----- Minimum/Maximum:
 Col.: 1 Mini = -2.8191 Maxi = 1.9518
 Col.: 2 Mini = -2.3209 Maxi = 2.6238
 Col.: 3 Mini = -3.1276 Maxi = 2.4303

Fau.--TC1 provient de la normalisation par bloc de :

File Fau.--TCco contains column scores
 from projections of separate clouds
 It has 160 rows and 3 columns
 It is to be used with --TCc.label and --TCc.cat files
 File : Fau.--TCco

----- Minimum/Maximum:
 Col.: 1 Mini = -1.748 Maxi = 1.2596
 Col.: 2 Mini = -1.5843 Maxi = 1.348

La superposition peut se faire par groupe ou d'un seul coup :



Pas tout à fait identique à celle du compromis de STATIS, la typologie des stations, vue synthétiquement par l'ACOM, est cependant très proche. Cette carte de synthèse, vue par l'une ou l'autre des méthodes contient manifestement toute l'information commune aux tableaux. Les deux méthodes utilisent pourtant pour y arriver des chemins bien différents.

La voie empruntée par STATIS est celle des opérateurs d'inertie, donc des matrices de produits scalaires. Mesurer l'adéquation d'un tableau au compromis ne peut pas se faire par le biais des nuages de points mais se fait clairement par le biais des nuages

d'opérateurs. La géométrie dans l'espace des matrices de produits scalaires nous a conduit à définir une valeur typologique comme le cosinus carré de l'angle entre la matrice des produits scalaires d'un tableau et l'approximation interprétée de la matrice des produits scalaires moyenne (compromis).

Le principe qui supporte l'ACOM n'a rien à voir.

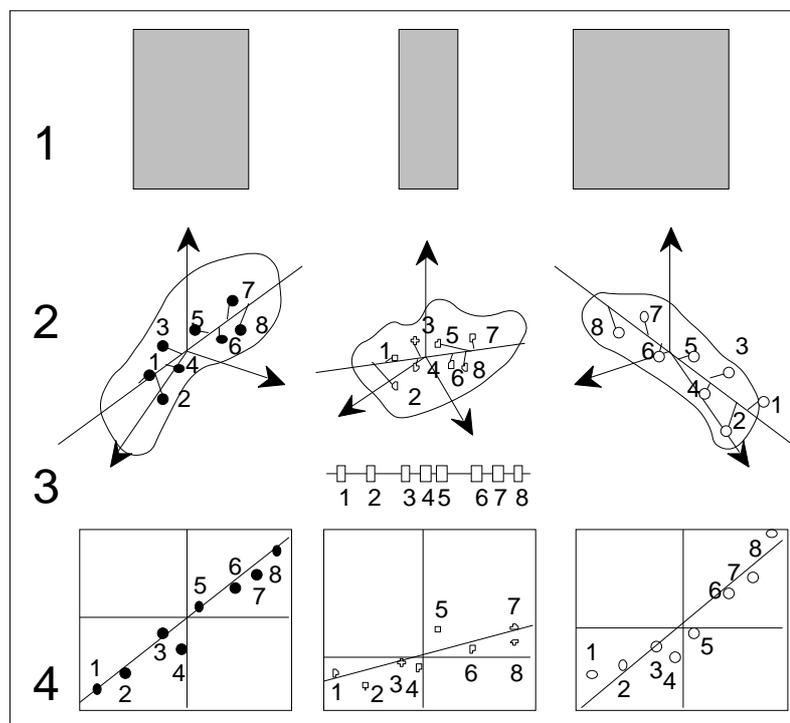


Figure 2 : Principe de la coinertie. 1 — K tableaux ont les mêmes lignes. 2— Ils définissent K nuages de points dans K espaces euclidiens. Les points sont pondérés de la même manière dans chaque nuage. Dans chacun des espaces on cherche un vecteur normé (axe) sur lequel on projette le nuage. 3 — On définit un code numérique de référence de variance unitaire. 4 — Les axes et le code de référence optimise la somme pondérée des carrés des covariances entre le code de référence et les coordonnées de chaque projection. On recommande sous la contrainte d'orthogonalité sur les axes et sur les codes.

Pour optimiser une somme de carrés de covariances, il faut fabriquer de la variance (donc de l'inertie projetée) et de la corrélation, donc des projections proches entre elles (puisque les plus proches possible d'un code donné).

4 — Une nouvelle mesure de la valeur typologique

Un groupe d'espèce qui a une forte valeur typologique, de ce point de vue, est un groupe qui permet de définir dans le nuage de points correspondant la plus grande séparation des positions des relevés (inertie projetée) et la plus grande corrélation entre ces positions et les positions de référence qui intègrent la capacité typologique globale.

Ces deux éléments sont visibles séparément. On voit l'inertie projetée dans les coordonnées de Fau.--TCco (figure 3), on voit la ressemblance avec les codes de référence dans la superposition des coordonnées de Fau.TCrep et Fau.TC1 (figure 4).

Sur la figure 4, on notera que la difficulté associée à la reconstitution des coordonnées dans STATIS (qui est parfaite et sans signification dès que le nombre de taxons dans un groupe est élevé) est ici résolue. La reconstitution se fait avec des coordonnées normalisées et si le nombre de taxons est élevé et si, de plus, la reconstitution est médiocre, on a l'indice d'une incapacité absolue du groupe à refaire l'ordination compromis.

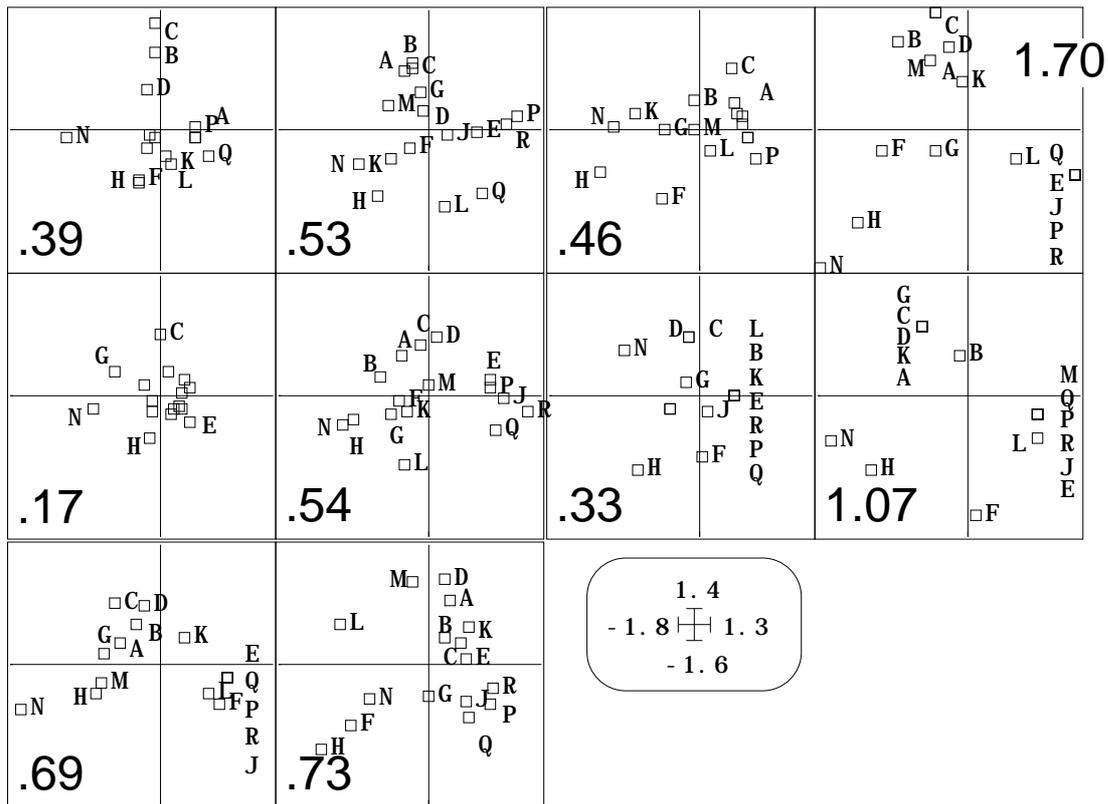


Figure 3 : Plans de projections des nuages de points sur leurs axes de co-inertie. Indication numérique de l'inertie projetée (var+ du niveau 2 dans le listing). L'inertie projetée varie dans un rapport de 1 à 10 entre deux groupes faunistiques qui ont respectivement entre 4 et 13 espèces, le groupe possédant le moins d'espèces ayant la plus forte valeur.

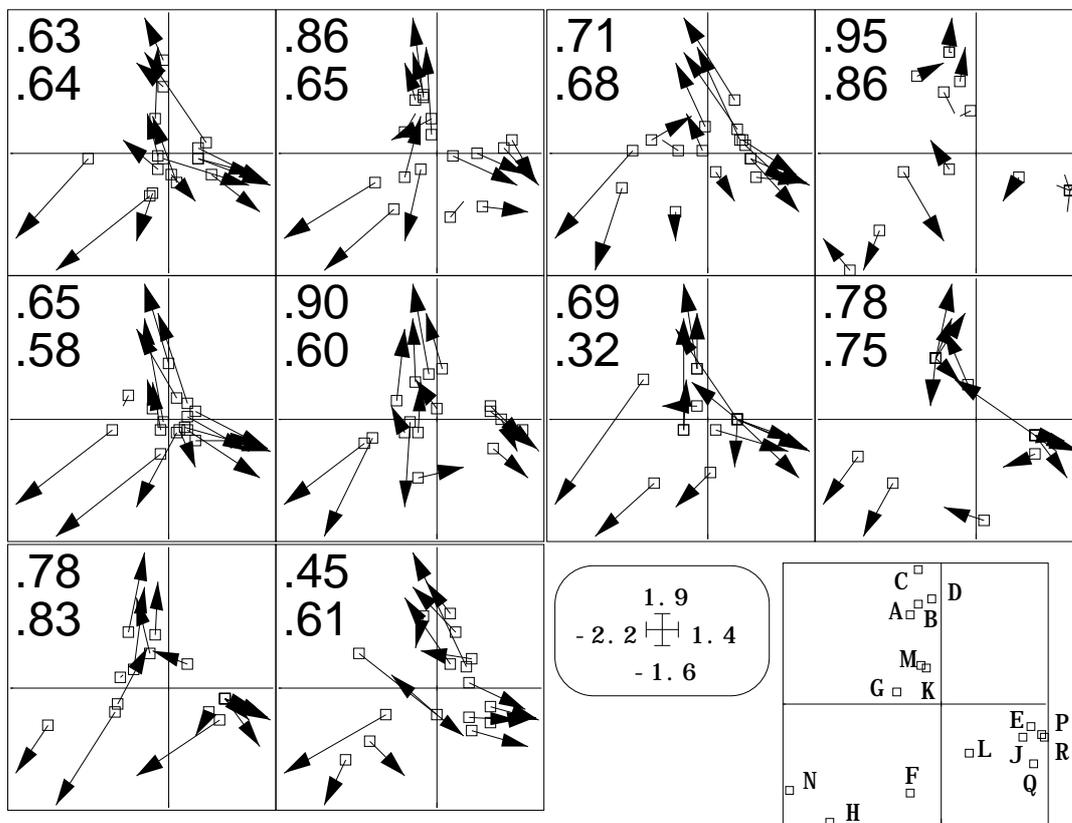


Figure 4 : Ajustement des projections par nuage (normalisée) et de la position de référence. Indication des carrés de corrélation par axes.

Suivant les cas, les rôles respectifs de l'inertie et de la corrélation dans la co-inertie varieront sans doute fortement. En particulier ici, les nombres de taxons étant grands par rapport au nombre de stations, en ne tenant compte que de la corrélation (qui figure dans les RV) on aurait une vue optimiste. Pour tenir compte des deux, on prendra comme indice de valeur typologique la quantité optimisée, soit la somme des carrés des covariances.

On consultera le listing qui donne 0.25 (.1008 + .1452) pour le groupe 1 :

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	2.413e-01	2.413e-01	1.601e-01	1.601e-01	1.008e-01	6.296e-01
2	1.805e-01	4.218e-01	2.281e-01	3.882e-01	1.462e-01	6.408e-01

Table n°: 1 * Rows: 11 * Col: 16 * Rank: 11
* Weight (x1000): 100 * RV (x1000): 726

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	3.243e-01	3.243e-01	2.731e-01	2.731e-01	2.344e-01	8.583e-01
2	2.778e-01	6.021e-01	2.573e-01	5.304e-01	1.665e-01	6.469e-01

Table n°: 2 * Rows: 7 * Col: 16 * Rank: 7
* Weight (x1000): 100 * RV (x1000): 849

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	3.717e-01	3.717e-01	3.454e-01	3.454e-01	2.449e-01	7.090e-01
2	1.470e-01	5.187e-01	1.182e-01	4.637e-01	8.022e-02	6.785e-01

Table n°: 3 * Rows: 13 * Col: 16 * Rank: 13
* Weight (x1000): 100 * RV (x1000): 769

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	1.005e+00	1.005e+00	9.694e-01	9.694e-01	9.173e-01	9.463e-01
2	7.005e-01	1.705e+00	7.332e-01	1.703e+00	6.308e-01	8.603e-01

Table n°: 4 * Rows: 4 * Col: 16 * Rank: 4
* Weight (x1000): 100 * RV (x1000): 827

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	1.436e-01	1.436e-01	9.423e-02	9.423e-02	6.134e-02	6.509e-01
2	9.812e-02	2.417e-01	7.563e-02	1.699e-01	4.354e-02	5.758e-01

Table n°: 5 * Rows: 13 * Col: 16 * Rank: 12
* Weight (x1000): 100 * RV (x1000): 731

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	4.264e-01	4.264e-01	4.047e-01	4.047e-01	3.624e-01	8.955e-01
2	2.037e-01	6.302e-01	1.389e-01	5.436e-01	8.272e-02	5.955e-01

Table n°: 6 * Rows: 22 * Col: 16 * Rank: 15
* Weight (x1000): 100 * RV (x1000): 879

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	1.951e-01	1.951e-01	1.724e-01	1.724e-01	1.153e-01	6.688e-01
2	1.696e-01	3.647e-01	1.601e-01	3.325e-01	5.115e-02	3.195e-01

Table n°: 7 * Rows: 4 * Col: 16 * Rank: 4
* Weight (x1000): 100 * RV (x1000): 535

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	7.023e-01	7.023e-01	6.341e-01	6.341e-01	4.973e-01	7.843e-01
2	3.716e-01	1.074e+00	4.347e-01	1.069e+00	3.247e-01	7.469e-01

Table n°: 8 * Rows: 3 * Col: 16 * Rank: 3
* Weight (x1000): 100 * RV (x1000): 820

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	1.951e-01	1.951e-01	1.724e-01	1.724e-01	1.153e-01	6.688e-01
2	1.696e-01	3.647e-01	1.601e-01	3.325e-01	5.115e-02	3.195e-01

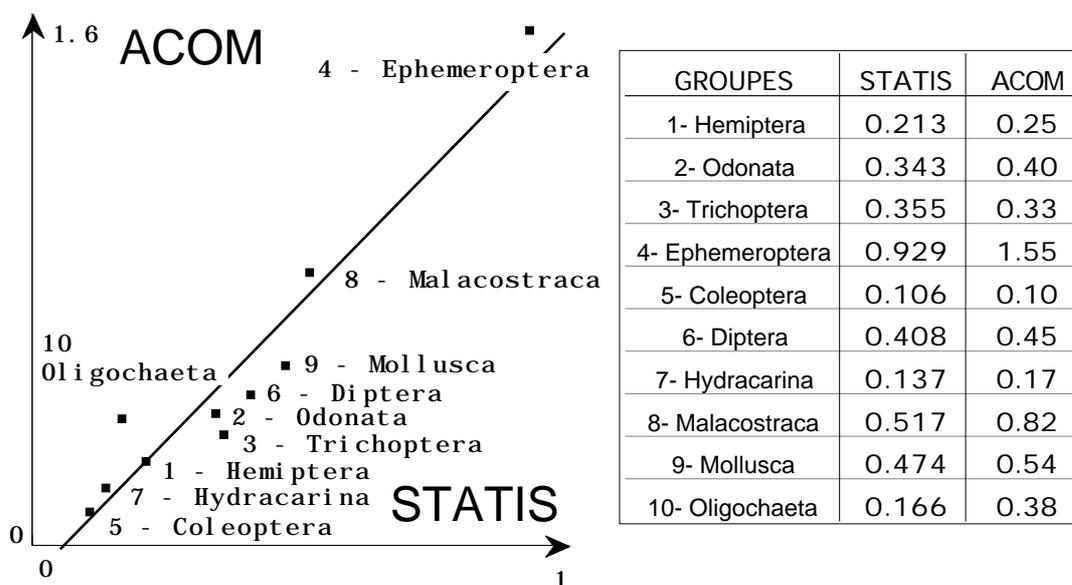
Table n°: 9 * Rows: 8 * Col: 16 * Rank: 8
* Weight (x1000): 100 * RV (x1000): 787

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	5.612e-01	5.612e-01	5.501e-01	5.501e-01	4.287e-01	7.792e-01
2	1.604e-01	7.215e-01	1.391e-01	6.892e-01	1.161e-01	8.345e-01

Table n°: 10 * Rows: 6 * Col: 16 * Rank: 6
* Weight (x1000): 100 * RV (x1000): 785

N	Iner	Iner+	Var	Var+	Sca. Pro.2	Cos2
1	5.360e-01	5.360e-01	3.899e-01	3.899e-01	1.741e-01	4.464e-01
2	4.284e-01	9.644e-01	3.412e-01	7.312e-01	2.086e-01	6.113e-01

On obtient une nouvelle mesure de la valeur typologique qui doit être comparés à celle qui est proposée par STATIS :



La compatibilité de ces deux mesures basée sur des principes très différents est acceptable.

5 — Représentation des groupes et des espèces

File Fau. --TL1 contains row scores (unit norm for each block)
It has 91 rows and 3 columns
File : Fau. --TL1

----- Minimum/Maximum:
Col.: 1 Mini = -2.2311 Maxi = 1.7264
Col.: 2 Mini = -1.6657 Maxi = 2.4036

On notera encore (figure 5) une différence sensible avec STATIS (fiche Statis-1, figure 9, p.12). Il n'y a aucune contradiction entre les deux représentations des espèces, mais les cartes ont des fonctions distinctes. Celle de STATIS procède d'une démarche qui va de la typologie de synthèse des relevés à la typologie induite des taxons. On y voit donc les différences de valeurs typologiques sur la carte des espèces d'un groupe.

Celle de l'ACOM procède de la démarche qui va d'une typologie des espèces vers une typologie des relevés spécifique à chaque tableau mais coordonnée à un compromis. Les cartes des espèces sont donc normalisées (moyenne des carrés des coordonnées des espèces unitaire). On verra donc la valeur typologique sur la carte des relevés (figures 3 et 4) et non sur celle des espèces (figure 5).

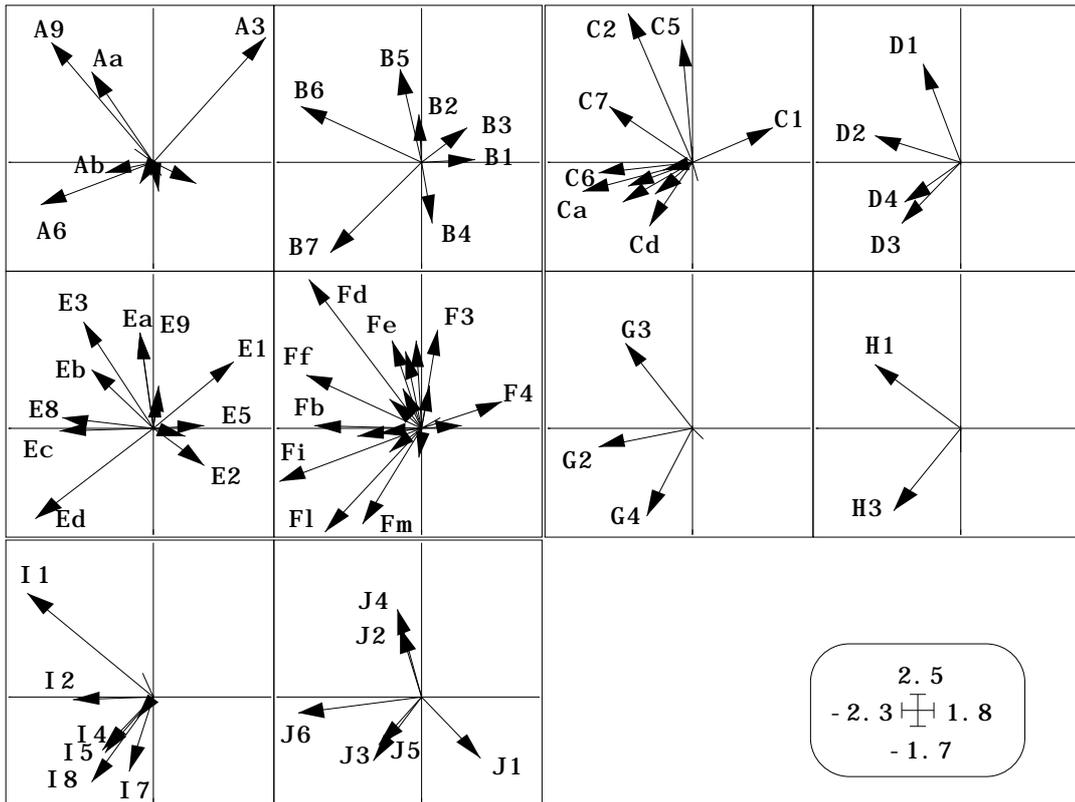


Figure 5 : Cartes des taxons par groupes (composantes des vecteurs \mathbf{u}_j^1 et \mathbf{u}_j^2 de la théorie.

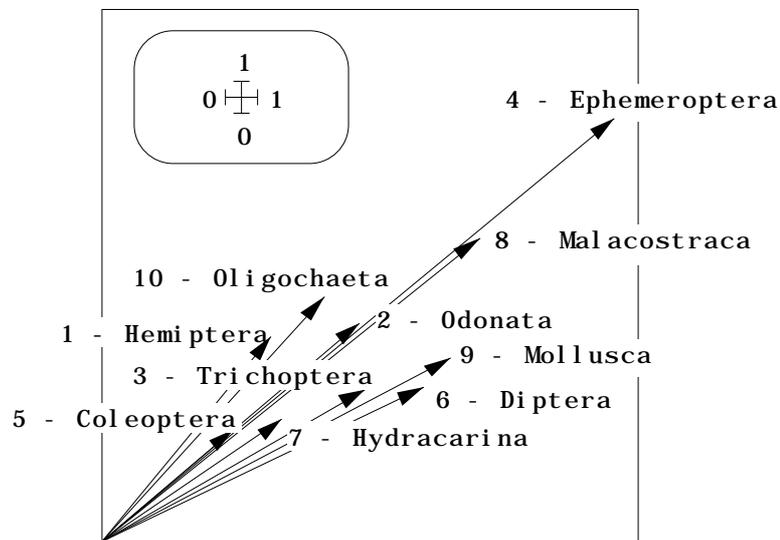


Figure 6 : Cartes des groupes (représentation par les carrés des covariances $(\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^1 | \mathbf{v}_1)_{\mathbf{D}}^2$ et $(\mathbf{X}_j \mathbf{Q}_j \mathbf{u}_j^2 | \mathbf{v}_2)_{\mathbf{D}}^2$).

Les résultats sont cohérents mais d'interprétation distincte. On retrouve ici encore le même résultat de plusieurs manières, ce qui ne pose pas de problème particulier dans le champ de l'expérimentation, mais qui interdit à nouveau de croire qu'il y a de bonnes et de mauvaises méthodes. On notera aussi combien l'interprétation est sous la

dépendance des principes théoriques, dont il est difficile d'ignorer l'existence. L'ACOM rajoute une représentation des groupes (figure 6) :

```
File Fau.--tab contains table scores
from projections of principal components (XkDkuk)
It has 10 rows and 3 columns
File : Fau.--tab
----- Minimum/Maximum:
Col. : 1 Mini = 0.24767 Maxi = 0.95778
Col. : 2 Mini = 0.20867 Maxi = 0.79421
```

La figure 6 est particulièrement explicite. Dans un cas complexe, on voit qu'il sera possible de préciser la participation de certains groupes à la typologie de synthèse sur certains aspects. Au total, nous avons maintenant des outils fortement structurés au plan théorique, d'usage aisé au plan pratique, qui permettent de discuter sur le fond de la variabilité d'une structure réalisée par un ensemble de descripteur, variabilité autour d'une structure de référence induite par l'ensemble des descripteurs étudiés.

Écologiquement, ces progrès pourraient contribuer aux discussions sur la fonction de la biodiversité. Ici, en effet, la richesse, descripteur habituel de la diversité β (entre les stations). Cela est peut-être lié au facteur écologique limitant (pH) qui est omniprésent dans cette observation, l'expérience méritant d'être reprise avec d'autres jeux de données. En effet, Foster² (1995) apporte des arguments très cohérents avec ces résultats.

Références

¹ Friday, L.E. (1987) The diversity of macroinvertebrate and macrophyte communities in ponds. *Freshwater Biology* : 18, 87-104.

² Foster, G.N. (1995) Evidence for pH insensitivity in some insects inhabiting peat pools in the loch fleet catchment. *Chemistry and Ecology* : 9, 207-215.