

L'analyse factorielle multiple

Résumé

L'Analyse factorielle multiple (AFMULT) est une des méthodes les plus pratiques parmi les méthodes d'ordination simultanée de K tableaux. La fiche reprend les principaux travaux des auteurs de l'AFMULT, B. Escofier et J. Pages, pour décrire l'utilisation du module MFA (option Variable groups) qui contient cette méthode.

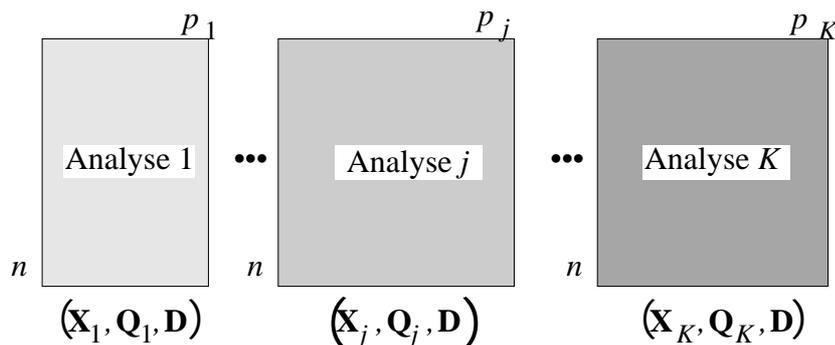
Plan

1 — L'objectif poursuivi	2
1.1 — Données de référence	2
1.2 — Notion de surpondération.....	3
2 — AFMULT et inertie simultanée	13
3 — L'AFMULT comme analyse inter-classes	17
4 — La représentation des groupes en AFMULT.....	21
Références	26

D. Chessel

1 — L'objectif poursuivi

L'AFMULT est une méthode d'ordination simultanée de K tableaux, comme STATIS et l'ACOM (fascicule 7), et en ce sens la fiche utilise le contenu du fascicule 6 de la présente documentation. Contrairement aux méthodes précédentes les auteurs de l'AFMULT n'ont envisagé que le cas des tableaux appariés par les individus et nous ne discuterons ici que de cette situation. On dispose de K tableaux ayant en commun les lignes-individus, chacun d'entre eux correspondant à un groupe de variables-colonnes. On dispose donc de K triplets statistiques $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$ ($1 \leq k \leq K$) :



L'objectif est d'ordonner simultanément K tableaux, de réaliser K analyses portant sur un même ensemble d'individus, ou encore de comparer K groupes de variables définies sur le même ensemble d'individus comme indiqué dans le premier texte présentant la méthode (Escofier & Pages 1982 a et b)¹. Pour vérifier le module écrit dans ADE-4, nous avons utilisé le jeu de données qui sous-tend la plupart des illustrations publiées. Plusieurs versions non identiques ont jalonné le travail de création de l'AFMULT, la plus complète est apparue dans un rapport détaillé de Escofier & Pages 1985². Les résultats de la présentation la plus récente (Escofier & Pages 1994)³ reprennent cet exemple et servent de référence. Dans la suite de cette fiche, nous ferons référence aux différents travaux de B. Escofier et J. Pages par l'abréviation E&P.

1.1 — Données de référence

Elles ont été recueillies par MM. Asselin et Morlat du centre de recherches INRA d'Angers (E&P 82b). Il s'agit d'analyse sensorielle, méthode centrale de la quantification de la qualité des produits alimentaires. Un certain nombre de dégustateurs attribuent une note à un produit donné pour une liste de critères précis. Lorsque chaque dégustateur analyse la même liste de produit on obtient des cubes de données Dégustateurs-Produits-Critères qui posent nombre de questions statistiques⁴

La variabilité individuelle des jugements n'est pas en jeu dans les données utilisées ici. 21 vins rouges du Val de Loire ont été analysés par 36 juges qui n'ont pas forcément répondu à toutes les questions du protocole : on connaît la note moyenne attribuée par l'ensemble des juges à un produit donné pour une variable donnée.

Les variables forment quatre groupes associées aux quatre périodes de la dégustation et aux sens mis en jeu. Le premier groupe décrit l'arôme du vin au repos par 5 variables :

- RInten : Intensité globale de l'arôme du vin au repos dans le verre
- RQual : Qualité globale des arômes
- RFruit : Note fruitée
- RFleur : Note fleurie

— REpice : Note épicée

Le deuxième groupe utilise trois variables associées à l'approche visuelle :

- VInten : Intensité colorée
- VNuance : Nuance violacée (Orangé=0 à violet=5)
- VSurf : Impression de surface (traces laissées sur le verre)

Le troisième groupe décrit l'arôme du vin après agitation et dégustation :

- AInten : Intensité de l'arôme du vin dans le verre
- AQuali : Qualité de l'arôme du vin dans le verre
- AFruit : Note fruitée
- AFleur : Note fleurie
- AEpice : Note épicée
- AVeg : Note végétale
- APhenol : Note phénolique
- ANoteb : Intensité globale de l'arôme du vin dans la bouche
- APersi : Persistance aromatique
- AQualb : Qualité des arômes du vin dans la bouche

Le quatrième groupe décrit les qualités gustatives du vin

- GInten : Intensité d'attaque
- GAcid : Acidité en bouche
- GAstr : Astringence en bouche
- GAlcool : Chaleur en bouche (alcool)
- GEqui : Équilibre entre dominantes
- GVelou : Velouté
- GAmert : Amertume
- GIfin : Intensité de fin en bouche
- GHarmo : Structure - harmonie

Les 21 vins soumis à la dégustation ont des noms formés par l'appellation d'origine (S = Saumur, B = Bourgueil, C = Chinon) les lettres mi (pour type de milieu) suivi des catégories 1, 2, 3 ou 4 suivi enfin d'une lettre a, b, c, ... pour plusieurs produits de même origine et même milieu de culture.

Pour une variable et un produit la valeur enregistrée est une moyenne de valeurs comprises entre 0 et 5 ou 1 et 5 suivant les cas. Les données sont accessibles sur la carte Afmult de ADE•Data.

L'AFMULT s'étend aux tableaux de variables tant qualitatives que quantitatives. Nous avons affaire ici à quatre tableaux d'ACP normées classiques.

1.2 — Notion de surpondération

Exécutons les quatre ACP initiales. Les valeurs (E&P 85 p. 29 & 30) sont éditées après multiplication par 1000 :

	1	2	3	4	5		1	2	3	4	5	6	7	8	9
Smi2a	3074	3000	2714	2280	1960	Smi2a	2963	2107	2429	2500	3250	2731	1926	2857	3143
Smi2b	2964	2821	2375	2280	1680	Smi2b	3036	2107	2179	2654	2926	2500	1926	2893	2964
Bmi2a	2857	2929	2560	1960	2077	Bmi2a	3222	2179	2250	2643	3321	2679	2000	3074	3143
Cmi3	2808	2593	2417	1913	2160	Cmi3	2704	3179	2185	2500	2333	1680	1963	2462	2038
Smi1a	3607	3429	3154	2154	2040	Smi1a	3464	2571	2536	2786	3464	3036	2071	3643	3643
Bmi1a	2857	3111	2577	2040	2077	Bmi1a	3286	2393	2643	2857	3286	2857	2179	3464	3500
Bmi1b	3214	3222	2962	2115	2040	Bmi1b	3393	2607	2607	2778	3464	2857	1929	3643	3556
Smi2c	3120	2852	2500	2200	2185	Smi2c	3250	2179	2630	2778	3179	2786	2000	3321	3296
Cmi2	2857	2815	2808	1923	2074	Cmi2	3286	2286	2407	2741	3143	2821	1964	3148	3286
Smi1b	3071	3107	2731	2120	1800	Smi1b	3519	2111	2536	2778	3444	3231	2071	3667	3786
Bmi2b	2643	2786	2536	1889	1808	Bmi2b	2179	2250	1964	2250	2464	1821	1679	2179	2107
Smi3a	2893	3000	2571	1846	1680	Smi3a	2893	2357	2250	2704	3214	2500	2185	2857	2963
Smi3b	3250	3286	2714	1926	1962	Smi3b	3321	2429	2571	2893	3192	2857	2214	3357	3071
Smi3c	3393	3179	2769	2038	1920	Smi3c	3357	2429	2607	2821	3107	2889	2037	3250	3393
Smi1c	3179	3286	2778	2231	1760	Smi1c	3393	2286	2500	2821	3500	3286	2000	3407	3643
Bmi1c	2929	3179	2852	2000	2037	Bmi1c	3393	2393	2357	2704	3321	3000	2000	3214	3214
Cmi1	3036	3179	3037	2231	1667	Cmi1	3071	2571	2321	2929	3333	2821	2143	3321	3250
Bmi2c	3107	3143	2846	2185	1962	Bmi2c	3185	2286	2643	2929	3286	2821	2107	3321	3296
Cmi3	3071	2926	2741	2000	1880	Cmi3	3071	2393	2321	2821	3143	2607	2143	3037	3074
Smi4a	3708	2926	2520	2040	2667	Smi4a	3333	2571	2667	2704	2769	2308	2667	3333	3000
Smi4b	3696	3192	2833	1826	2385	Smi4b	2963	2407	2643	2963	2571	2071	2222	3037	2741

Tableau 1

Tableau 4

	1	2	3		1	2	3	4	5	6	7	8	9	10
Smi2a	4321	4000	3269	Smi2a	3407	3308	2885	2320	1840	2000	1650	3259	2963	3200
Smi2b	3222	3000	2808	Smi2b	3370	3000	2560	2440	1739	2000	1381	2962	2808	2926
Bmi2a	3536	3393	3000	Bmi2a	3250	2929	2769	2192	2250	1750	1250	3077	2800	3077
Cmi3	2893	2786	2538	Cmi3	3160	2880	2391	2083	2167	2304	1475	2542	2583	2478
Smi1a	4393	4036	3385	Smi1a	3536	3360	3160	2231	2148	1762	1600	3615	3296	3462
Bmi1a	4464	4259	3407	Bmi1a	3179	3385	2800	2240	2148	1750	1476	3214	3148	3321
Bmi1b	4143	3929	3250	Bmi1b	3429	3500	3038	2200	2385	1826	1478	3250	3222	3385
Smi2c	4214	3857	3077	Smi2c	3654	3077	2520	2320	2444	2080	1905	3280	3160	2962
Cmi2	4037	3893	3280	Cmi2	3357	3346	3000	2040	2125	1875	1524	3146	2893	3308
Smi1b	4714	4536	3429	Smi1b	3357	3444	2885	2120	2346	1826	1550	3269	3080	3192
Bmi2b	2607	2536	2444	Bmi2b	2889	2800	2500	1962	2111	2080	1318	2680	2308	2556
Smi3a	3704	3407	3111	Smi3a	3222	3259	2926	2040	2042	2000	1773	3077	2704	2778
Smi3b	3857	3643	3259	Smi3b	3607	3385	2889	2115	2160	1955	1571	3286	3036	3222
Smi3c	4714	4500	3321	Smi3c	3481	3385	2962	2000	2200	2042	1545	3321	3071	3143
Smi1c	4222	4071	3462	Smi1c	3481	3423	2963	2269	2154	1957	1571	3481	3259	3269
Bmi1c	3889	3429	3143	Bmi1c	3286	3308	3115	2269	2000	1917	1400	3040	2960	3200
Cmi1	3786	3607	3357	Cmi1	3444	3500	3185	2160	2240	1913	1750	3520	3296	3462
Bmi2c	4071	3893	3462	Bmi2c	3357	3370	2846	2240	2280	1750	1524	3333	3037	3370
Cmi3	3679	3393	3192	Cmi3	3370	3360	2963	2308	1917	2000	1429	3250	2920	2880
Smi4a	4321	4107	3259	Smi4a	3727	2885	2600	2083	2609	2174	1650	3095	3136	2545
Smi4b	4321	4000	3333	Smi4b	3737	3080	2833	1773	2440	2292	1571	3437	2958	2600

Tableau 2

Tableau 3

Dans STATIS et ACOM, nous avons dans la même situation introduit des pondérations des variables de manière à ce que la somme des poids des variables d'un tableau soit égale à l'unité. Cette pratique est déconseillée en AFMULT :

It should be noted that weighting the groups with the inverse of the number of variables (the number of variables of group j is equal to the global inertia of this group since variables are centred and reduced) is not suitable. (E&P93 p. 133)

La raison principale est que l'AFMULT pour uniformiser le rôle des tableaux dans l'analyse simultanée ne cherche pas à ramener à une échelle commune les inerties totales mais les inerties projetées (première valeur propre) sur le premier axe de chaque tableau. Pour respecter ce choix, on procédera comme suit.

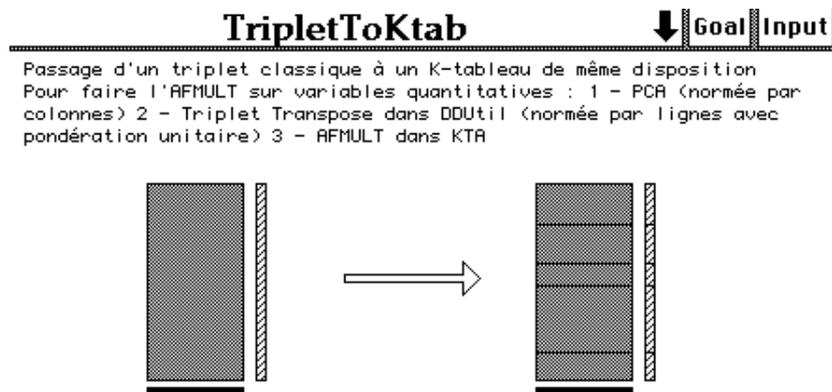
Implanter le tableau juxtaposé Z à 21 lignes (vins) et 27 colonnes (5+3+10+9 variables). En faire l'ACP normée :

Correlation matrix PCA		
Matrix input file	 Z	21 27

Permuter lignes et colonnes du triplet ainsi créé (Z.cnta le tableau, Z.cnpc le poids des colonnes, Z.cnpl le poids des lignes, Z.cnpa le fichier des paramètres) par l'option Triplet Transpose du module DDUtil :

Triplet Transpose		
Input file	 Z.cnta	21 27
Generic output file name	 A	

On obtient le triplet formé de A.trta (tableau transposé de Z.cnta), A.trpl (les poids des lignes identique à Z.cnpc) et A.trpc (les poids des colonnes identique à Z.cnpl). Les tableaux sont juxtaposés les uns sur les autres comme dans tous les modules de la partie K-tableaux. L'option TripletToKTab de KTabUtil prépare les K-tableaux à partir d'un triplet sans changer les centrages et les pondérations :



TripletToKTab		
---ta type file	 A.trta	27 21
Row indicator (Default = 1 class)	 Zbloc	4 1
Col indicator (Default = 1 class)		
Output file name	 B	

Dans Zbloc on a simplement, comme d'habitude les effectifs des lignes du K-tableau (ici les variables) par blocs. On peut alors faire les analyses séparées :

Separate analyses						
---.kttta input file			B.kttta		27 21	
Row bloc: 1 - Col bloc: 1 - Total inertia:				5		
Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum	
01 +2.2419E+00	+0.4484	+0.4484	02 +1.5164E+00	+0.3033	+0.7517	
03 +8.1542E-01	+0.1631	+0.9147	04 +2.8189E-01	+0.0564	+0.9711	
05 +1.4438E-01	+0.0289	+1.0000				

Row bloc: 2 - Col bloc: 1 - Total inertia:				3		
Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum	
01 +2.8348E+00	+0.9449	+0.9449	02 +1.5096E-01	+0.0503	+0.9952	
03 +1.4271E-02	+0.0048	+1.0000				

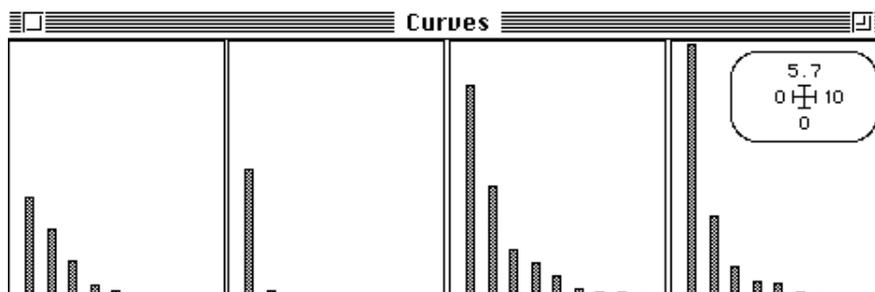
Row bloc: 3 - Col bloc: 1 - Total inertia:				10		
Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum	
01 +4.7015E+00	+0.4702	+0.4702	02 +2.4830E+00	+0.2483	+0.7185	
03 +1.0460E+00	+0.1046	+0.8231	04 +7.6504E-01	+0.0765	+0.8996	
05 +4.9583E-01	+0.0496	+0.9491	06 +1.9965E-01	+0.0200	+0.9691	
07 +1.1983E-01	+0.0120	+0.9811	08 +9.3627E-02	+0.0094	+0.9905	
09 +6.6606E-02	+0.0067	+0.9971	10 +2.8872E-02	+0.0029	+1.0000	

Row bloc: 4 - Col bloc: 1 - Total inertia:				9		
Num. Eigenval.	R.Iner.	R.Sum	Num. Eigenval.	R.Iner.	R.Sum	
01 +5.6420E+00	+0.6269	+0.6269	02 +1.7910E+00	+0.1990	+0.8259	
03 +6.7408E-01	+0.0749	+0.9008	04 +3.5051E-01	+0.0389	+0.9397	
05 +3.1272E-01	+0.0347	+0.9745	06 +1.1901E-01	+0.0132	+0.9877	
07 +5.9032E-02	+0.0066	+0.9943	08 +3.0702E-02	+0.0034	+0.9977	
09 +2.0934E-02	+0.0023	+1.0000				

A une toute petite différence près sur le tableau 3 (due aux difficultés de relecture du listing des valeurs brutes) on retrouve les résultats édités dans E&P85 p. 36).

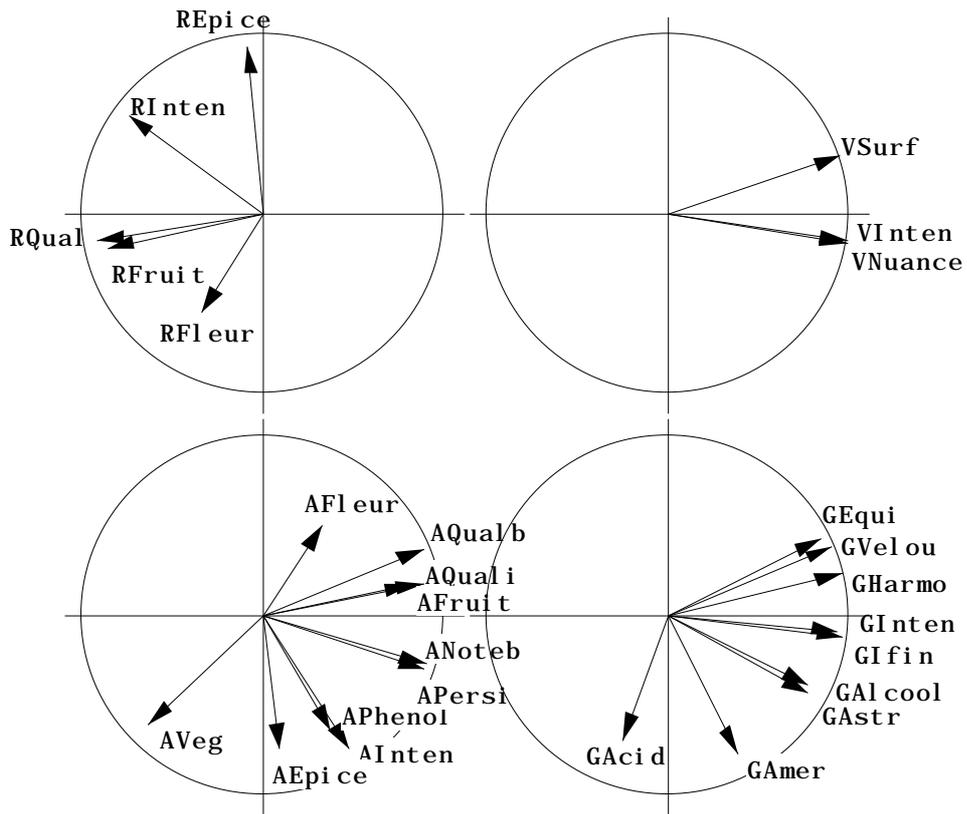
Eigenvalues are stored by column into file B.savp
 Row : 10 - Col (tables): 4

Le tableau 2 est presque de rang 1. Les valeurs propres sont représentées par Curves :

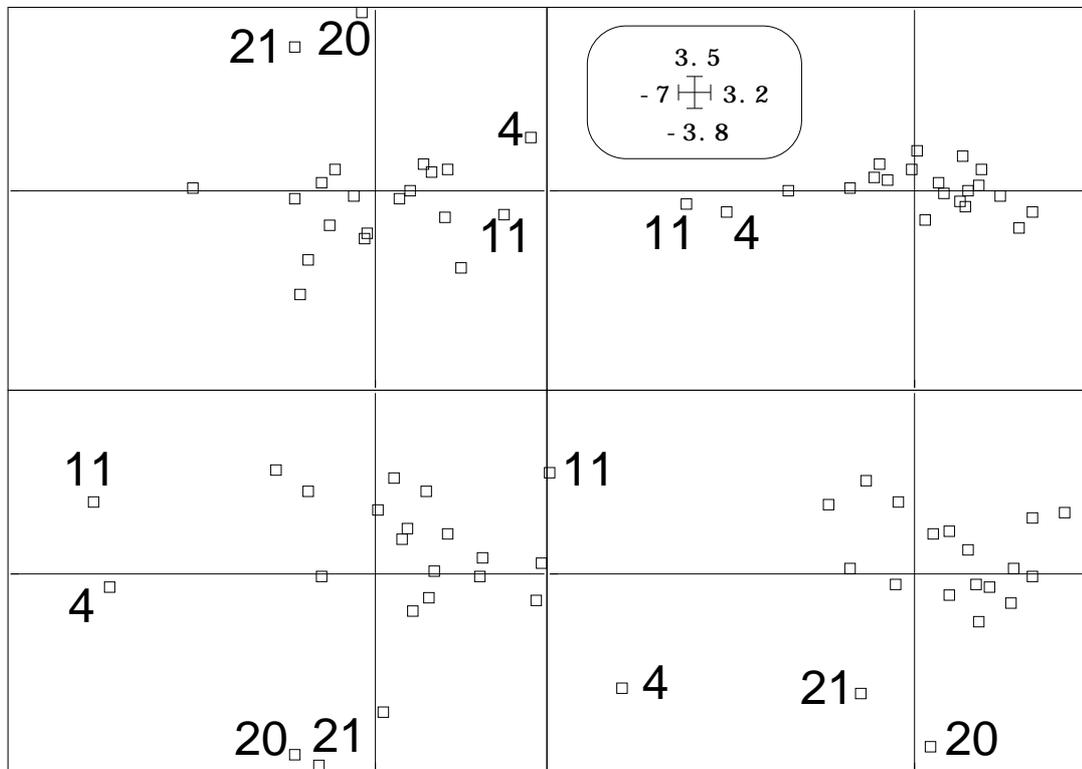


File B.saTlli contains row scores
 in separate analyses
 It has 27 rows and 4 columns
 It is to be used with --Tll.label and --Tll.cat files

Les plans 1-2 des analyses séparées sont obtenues par Scatters :



File B.saTcco contains column scores
 in separate analyses
 It has 84 rows and 4 columns
 It is to be used with --Tcc.label and --Tcc.cat files



On reconnaît les orientations aléatoires des axes d'inertie et la présence régulière de points systématiquement éloignés de l'origine. L'analyse du tableau complet permet de progresser.

Classical Principal Component Analysis (Hotteling 1933)

```

Input file: Z
---- Row weights:
File Z.cnpl contains the row weights. It has 21 rows and 1 column.
Each row has 0.047619 weight (Sum = 1)
---- Column weights:
File Z.cnpc contains the column weights. It has 21 rows and 1 column.
Each column has unit weight (Sum = 27)
---- Table:
File Z.cnta contains the centred and normed table. Zero mean and unit
variance for each column. It has 21 rows and 27 columns
---- Info: means and variances
File Z.cnma contains the descriptive of the analysis
It contains successively:
    Number of rows: 21
    Number of columns: 27
    means and variances:
    Col.: 1 Mean: 3111.14 Variance: 79793.2
    ...
    Col.: 27 Mean: 3147.95 Variance: 186094
-----
File Z.cn+r contains the matrix
from statistical triplet Z.cnta
It has 27 rows and 27 columns
----- matrix -----

```

```

[ 1] 1000
[ 2] 547 1000
...
[ 27] 305 673 536 525 -144 805 798 840 392 766 657
      388 132 -703 247 771 830 808 877 -372 632 616
      909 948 175 910 1000

```

DiagoRC: General program for two diagonal inner product analysis

```

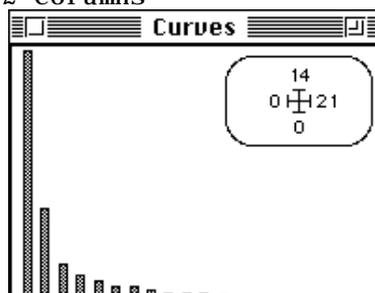
Input file: Z.cnta
--- Number of rows: 21, columns: 27
-----
Total inertia: 27
-----

```

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+1.3707E+01	+0.5077	+0.5077	02	+5.0922E+00	+0.1886	+0.6963
03	+2.0017E+00	+0.0741	+0.7704	04	+1.3316E+00	+0.0493	+0.8197
...							
19	+1.3431E-02	+0.0005	+0.9996	20	+1.0373E-02	+0.0004	+1.0000
21	+0.0000E+00	+0.0000	+1.0000				

File Z.cnvp contains the eigenvalues and relative inertia for each axis

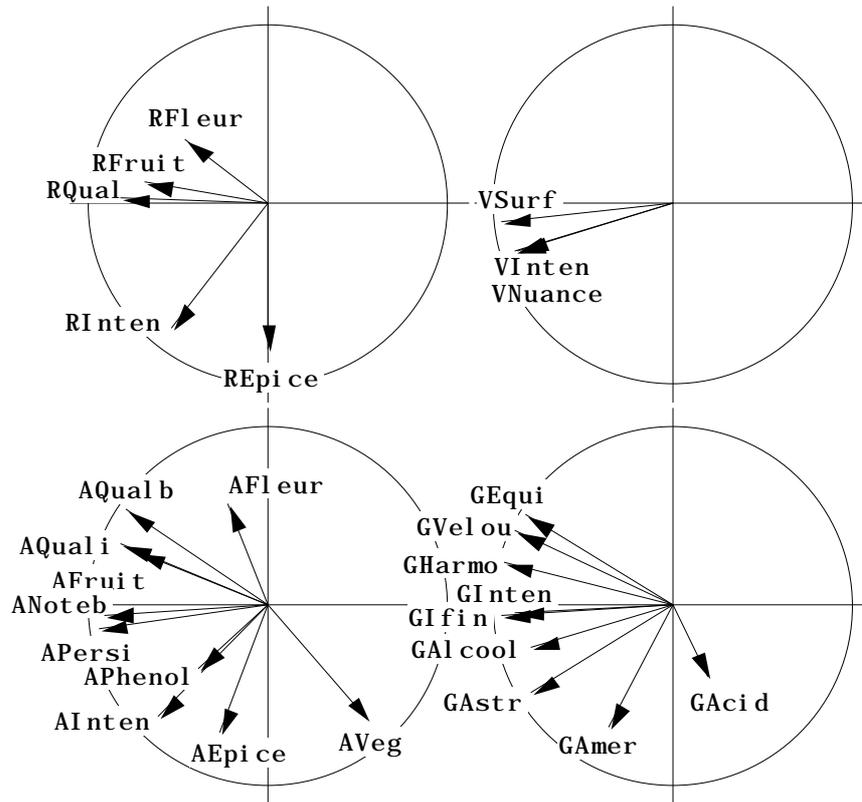
--- It has 27 rows and 2 columns



La dimension 2 pour le dépouillement semble s'imposer.

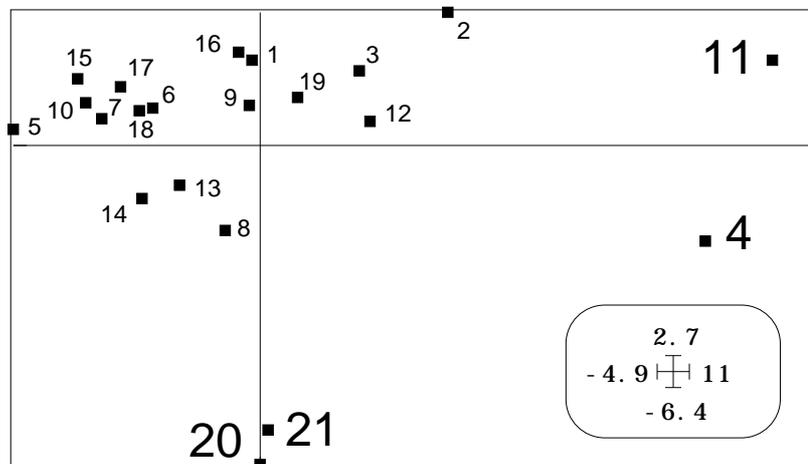
File Z.cnco contains the column scores
 --- It has 27 rows and 2 columns
 File : Z.cnco

Col.	Mini	Maxi
1	-9.460e-01	5.532e-01
2	-8.156e-01	5.682e-01



On pourrait se contenter de cette approche pour parler des variables.

File Z.cnli contains the row scores
 --- It has 21 rows and 2 columns
 File : Z.cnli



On cherche, évidemment, à concilier l'approche par analyse séparées qui donne des couples de cartes lignes-colonnes pour chaque groupe et l'approche analyse globale qui permet encore les cartes par groupe de variables (cartes cohérentes entre elles) mais impose la carte unique des individus. Les principes de l'AFMULT que nous allons aborder ont une propriété très particulière.

Ils sont précis et relativement sophistiqués. La présentation la plus achevée est celle de E&P84⁵. La solution pour les obtenir est elle-même très simple. Si on s'en tient à cette dernière propriété, on pourrait croire que la méthode n'opère pas un progrès sensible sur l'analyse globale du tableau de départ. En effet, l'AFMULT s'appuie sur une analyse globale du tableau de départ dans laquelle on ne modifie que le poids des variables, par l'opération de surpondération.

Pratiquement, on voit cette opération dans le listing du programme. L'option Variable groups du module MFA est un programme d'AFMULT qui respecte au mieux les propositions des auteurs. On l'utilise avec l'option par défaut :

Variable groups			
---.kttta input file		B.kttta	27 21
Table weights (default = 1/lambda1)			

```
Multiple factor analysis
Escofier & Pages 1984
Input file B
Row number: 27, Column number: 21, Block number: 4
Selected weighting option
-> First eigenvalue from separate analysis
```

Le programme crée un nouveau triplet statistique à partir l'analyse globale. B.++ta est le tableau dupliqué de B.kttta, lui-même identique à A.trta, lui-même transposé de Z.cnta. Les fichiers associés au triplet transposé (A.trpa, A.trpc, A.trpl et A.trta) peuvent être détruits. Les poids des colonnes du nouveau triplet (B.++pc) sont exactement ceux de B.ktpc, lui-même identique à Z.cnpl (après l'échange lignes-colonnes). Nous pouvons donc faire de l'AFMULT avec une pondération arbitraire des individus. Seule la pondération des lignes (B.++pl) n'est pas identique à la pondération de départ B.ktpl (issu de Z.cnpc après l'échange lignes-colonnes). Les poids des lignes (donc des variables initiales) associées au groupe j ($1 \leq j \leq 4$) sont simplement divisés par la première valeur propre de l'analyse séparée du tableau j .

La surpondération des lignes (variables de départ) engendre un triplet standard (B.++ta, tableau d'ACP normée transposé, B.++pc poids des individus d'une ACP normée, B.++pl poids des variables d'une ACP normée multipliés par une constante par bloc de variables). Ce nouveau triplet est simplement analysé par la procédure standard.

```
DiagoRC: General program for two diagonal inner product analysis
Input file: B.++ta
--- Number of rows: 27, columns: 21
-----
Total inertia: 7.01072
-----
```

Num.	Eigenval.	R. Iner.	R. Sum	Num.	Eigenval.	R. Iner.	R. Sum
01	+3.4620E+00	+0.4938	+0.4938	02	+1.3666E+00	+0.1949	+0.6888
03	+6.1536E-01	+0.0878	+0.7765	04	+3.7222E-01	+0.0531	+0.8296
05	+2.7021E-01	+0.0385	+0.8682	06	+2.0236E-01	+0.0289	+0.8970
...							
13	+2.8692E-02	+0.0041	+0.9895	14	+2.1967E-02	+0.0031	+0.9927
15	+1.9176E-02	+0.0027	+0.9954	16	+1.0924E-02	+0.0016	+0.9970
17	+9.1444E-03	+0.0013	+0.9983	18	+6.3589E-03	+0.0009	+0.9992

```

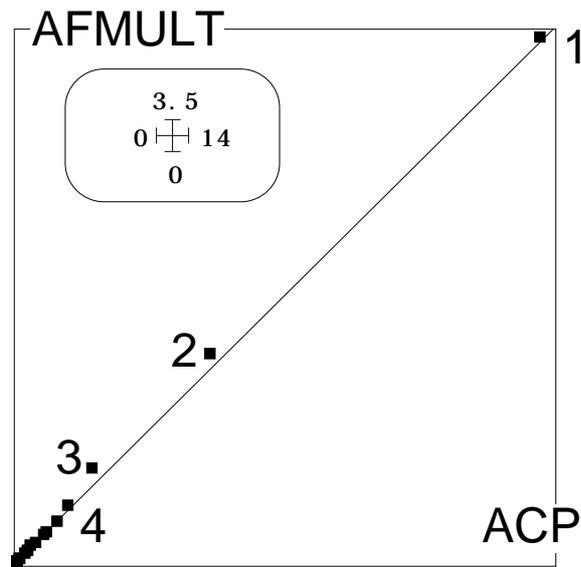
19 +3.3120E-03 +0.0005 +0.9997 | 20 +2.4245E-03 +0.0003 +1.0000 |
21 +0.0000E+00 +0.0000 +1.0000

```

On obtient les valeurs propres éditées dans E&P85 p. 38, en remarquant qu'ici la diagonalisation s'est faite dans la plus petite dimension (21 individus) alors que le travail cité donne 27 valeurs propres nulles à partir du rang 21.

File B. ++vp contains the eigenvalues and relative inertia for each axis

--- It has 21 rows and 2 columns



L'inertie totale du nouveau triplet (AFMULT) égale la somme sur les groupes de la quantité nombre de variables du groupe divisé par la première valeur propre de l'analyse séparée du groupe. La principale propriété est que la première valeur propre ne peut dépasser le nombre de groupes (4). La répartition de l'inertie entre axes d'ACP et axes d'AFMULT est peu modifiée par la surpondération (ci-dessus, en abscisse valeurs propres de l'ACP, en ordonnée valeur propre de l'AFMULT de même rang). L'analyse du nouveau triplet donne normalement des coordonnées lignes et colonnes :

File B. ++co contains the column scores

--- It has 21 rows and 3 columns

File : B. ++co

Col.	Mini	Maxi
1	-2.696e+00	4.886e+00
2	-3.362e+00	1.383e+00
3	-1.858e+00	1.179e+00

File B. ++li contains the row scores

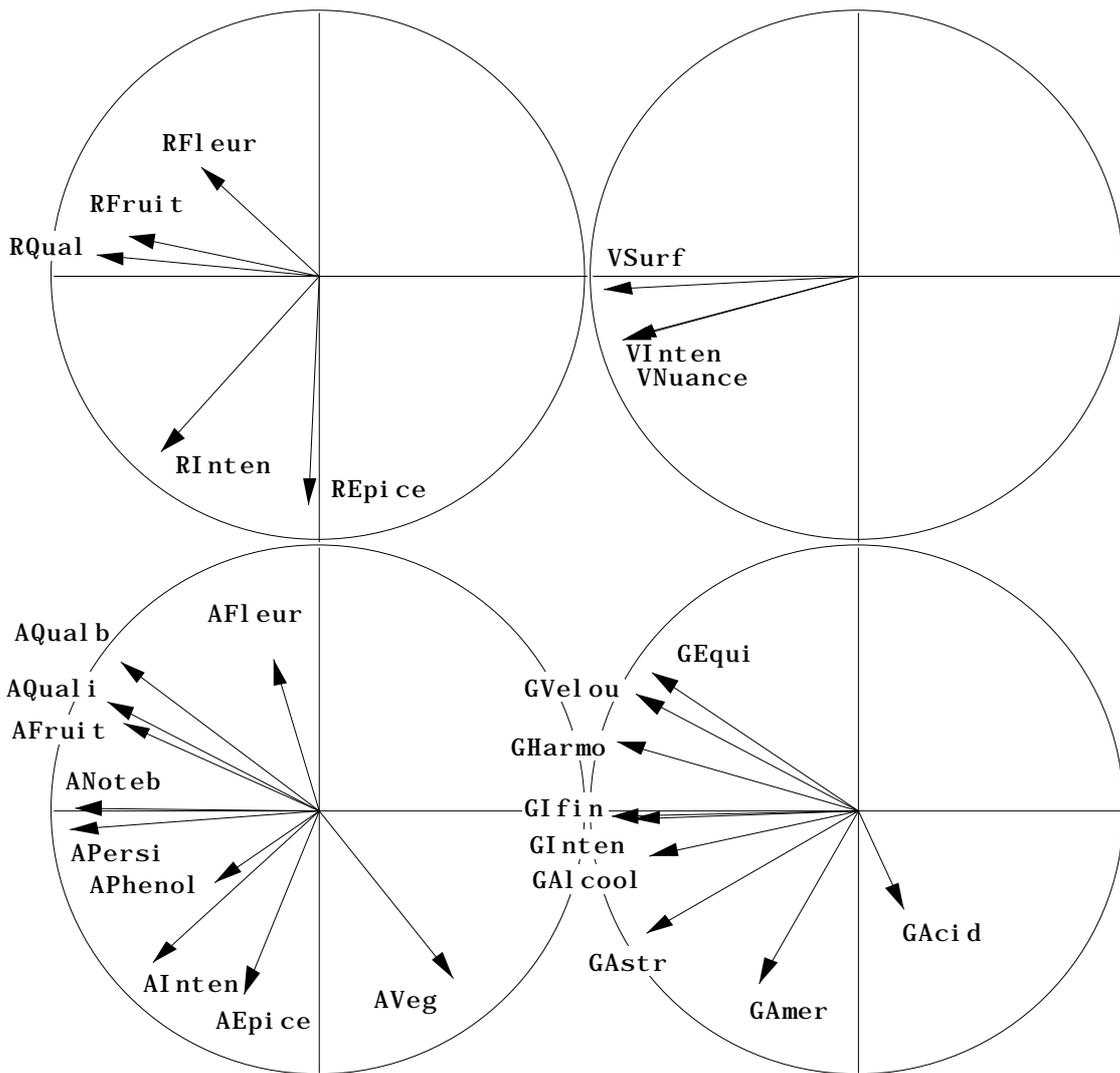
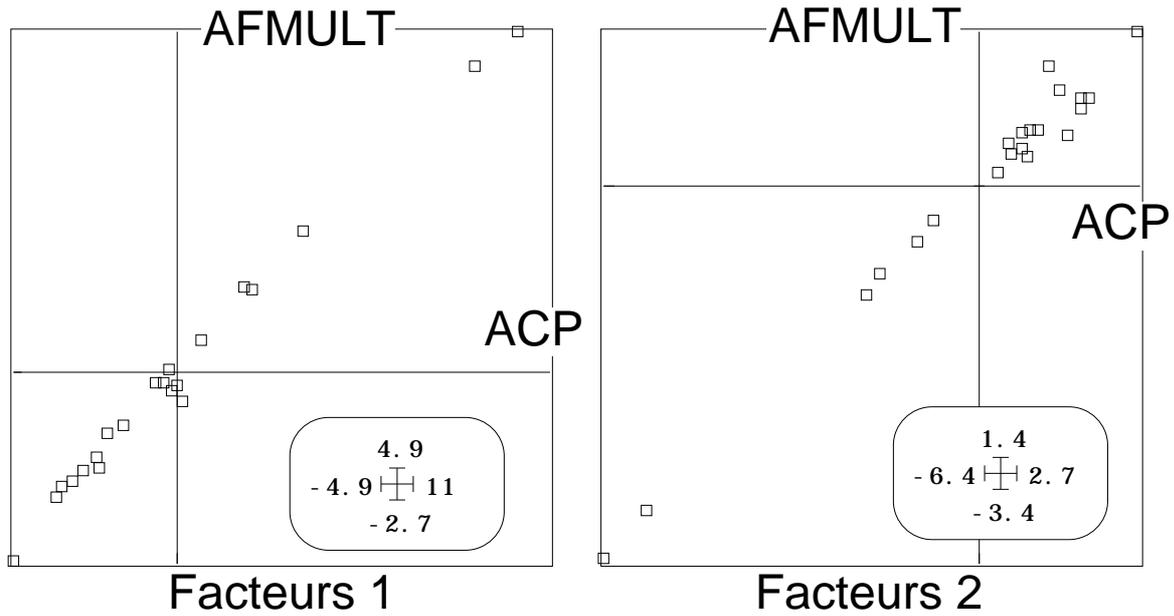
--- It has 27 rows and 3 columns

File : B. ++li

Col.	Mini	Maxi
1	-9.501e-01	5.064e-01
2	-8.650e-01	5.728e-01
3	-6.373e-01	5.375e-01

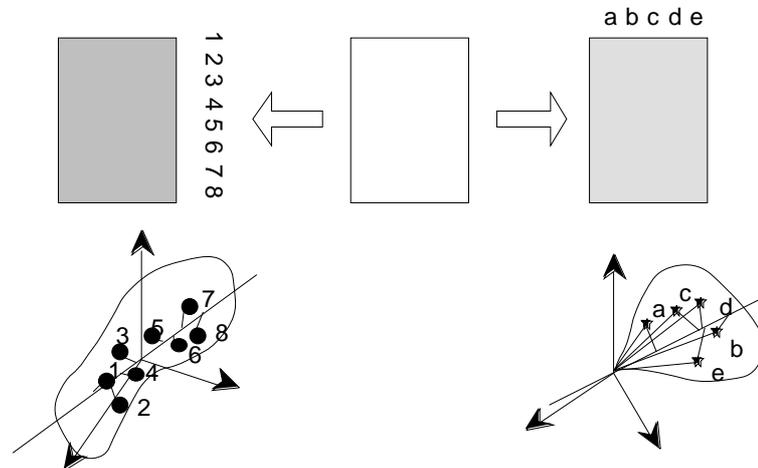
Comme on pouvait s'y attendre, les coordonnées des individus (ci-dessous, lignes de l'ACP globale dans Z.cnli, colonnes de l'AFMULT dans B.++co) sont peu modifiées

par l'altération de la métrique par surpondération. En bas, les cartes des variables (B.++li) :

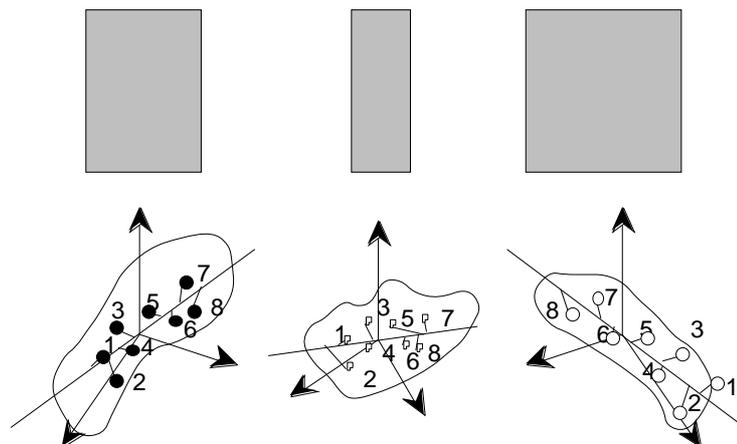


2 — AFMULT et inertie simultanée

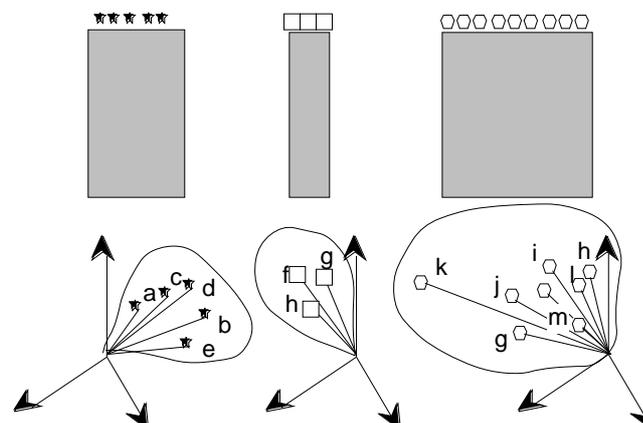
On aurait tort de croire que l'AFMULT est une simple surpondération du triplet de l'analyse globale. L'analyse du triplet à pondération modifiée donne la solution de plusieurs problèmes dont l'exemple choisi ne montre pas immédiatement toute la portée. Nous avons déjà rappelé qu'un tableau individus-variables génère deux nuages de points :



K tableaux génèrent donc $2K$ nuages de points. Les K nuages d'individus sont dans K espaces distincts et l'ACOM en explore la géométrie par le principe de co-inertie :



Par contre les K nuages de variables sont dans un même espace, donc susceptibles d'une approche géométrique simultanée :



On voit qu'en partant des objets créés par les mêmes données plusieurs approches sont concurrentes. L'analyse linéaire d'un tableau a ceci de caractéristique : elle aborde les deux nuages de points dualement. C'est la base de son succès : tout résultat obtenu sur un des deux objets se retrouve exprimé directement sur l'autre et réciproquement. Mais dès qu'il y a plus d'un tableau la symétrie lignes-colonnes est détruite et plusieurs voies sont ouvertes suivant qu'on aborde la question par l'une ou l'autre.

L'analyse globale et l'AFMULT partage donc un même objet, celui du nuage des variables. Notons p le nombre total de variables, P l'ensemble des p premiers entiers, à savoir $\{1, 2, \dots, p\}$. Notons p_1, p_2, \dots, p_K les effectifs des variables par tableau et P_1, P_2, \dots, P_K les parties de P correspondant à chaque groupe. Cela permet d'écrire $j \in P_k$ ($1 \leq j \leq p$) ($1 \leq k \leq K$) pour dire que la variable j dans la numérotation complète appartient au tableau k . \mathbf{x}^j désigne la variable j , vecteur de \mathbb{R}^n (n est le nombre d'individus), de poids initial m_j dans l'analyse globale (ici l'unité). \mathbb{R}^n est muni du produit scalaire \mathbf{D} associé à la pondération des individus.

On note encore \mathbf{Q}_k la matrice diagonale à p_k lignes et p_k colonnes qui contient les poids m_j des variables du groupe k . On note enfin α_k la première valeur propre de l'analyse séparée du triplet $(\mathbf{X}_k, \mathbf{Q}_k, \mathbf{D})$, où \mathbf{X}_k est le tableau du groupe k .

Si \mathbf{z} est un vecteur de \mathbb{R}^n \mathbf{D} -normé, l'inertie projetée associée aux p_k colonnes de \mathbf{X}_k est :

$$I_k(\mathbf{z}) = \sum_{j \in P_k} p_j (\mathbf{x}^j | \mathbf{z})_{\mathbf{D}}^2 = \|\mathbf{X}_k^t \mathbf{D} \mathbf{z}\|_{\mathbf{Q}_k}^2 = \left(\mathbf{X}_k^t \mathbf{D} \mathbf{z} \mid \mathbf{X}_k^t \mathbf{D} \mathbf{z} \right)_{\mathbf{Q}_k} = \left(\mathbf{X}_k \mathbf{D} \mathbf{X}_k^t \mathbf{Q}_k \mathbf{z} \mid \mathbf{z} \right)_{\mathbf{D}} = \left(\mathbf{W}_k \mathbf{D} \mathbf{z} \mid \mathbf{z} \right)_{\mathbf{D}}$$

On appellera valeur du lien entre le vecteur \mathbf{z} et le tableau \mathbf{X}_k cette inertie projetée ramenée au maximum potentiel, c'est-à-dire α_k . D'où la notation (E&P84 p. 43, d'abord noté L^2 , E&P94 p. 125) :

$$L(\mathbf{z}, k) = \sum_{j \in P_k} \frac{p_j}{\alpha_k} (\mathbf{x}^j | \mathbf{z})_{\mathbf{D}}^2 = \frac{1}{\alpha_k} \|\mathbf{X}_k^t \mathbf{D} \mathbf{z}\|_{\mathbf{Q}_k}^2 = \frac{1}{\alpha_k} \left(\mathbf{W}_k \mathbf{D} \mathbf{z} \mid \mathbf{z} \right)_{\mathbf{D}}$$

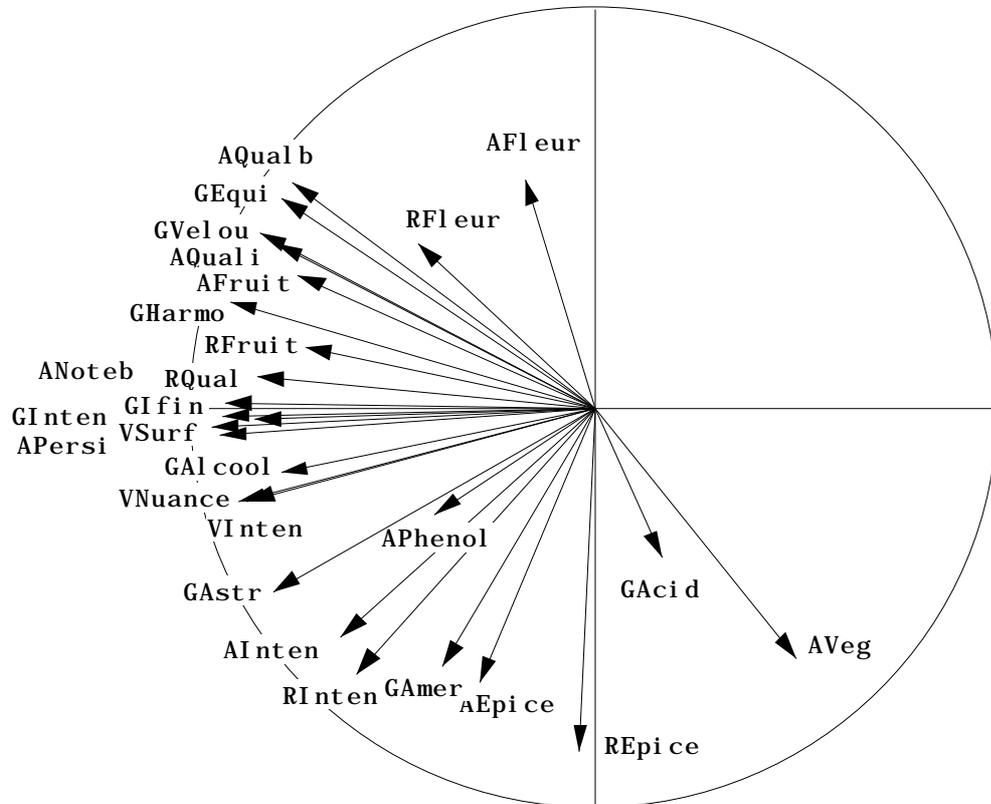
Ce lien ne peut donc dépasser l'unité (pourcentage d'un optimum). La valeur globale du vecteur \mathbf{z} , sur l'ensemble des nuages de variables (des tableaux) est la somme des liens, soit :

$$L(\mathbf{z}) = \sum_{k=1}^K L(\mathbf{z}, k) = \sum_{k=1}^K \frac{1}{\alpha_k} \left(\mathbf{W}_k \mathbf{D} \mathbf{z} \mid \mathbf{z} \right)_{\mathbf{D}} = \sum_{k=1}^K \frac{1}{\alpha_k} \left(\mathbf{W}_k \mathbf{D} \mathbf{z} \mid \mathbf{z} \right)_{\mathbf{D}}$$

L'optimum est atteint pour la première composante principale normée du schéma surpondéré de l'AFMULT et cet optimum ne peut dépasser le nombre de tableau (somme de k valeurs inférieures ou égales à 1). On cherche ensuite un second vecteur \mathbf{z} , orthogonal au précédent qui optimise la même quantité et on trouve la seconde composante de l'AFMULT. En cherchant un plan qui optimise l'inertie projetée, on trouve le sous-espace engendré par les deux premières composantes.

Tout se passant dans le même espace \mathbb{R}^n , il est logique de représenter simultanément toutes les variables alors qu'on a figuré (ci-dessus) les nuages de variables par tableau.

On retrouve la figure 3 de E&P94 p.135 :



On pourrait se croire dans une ACP normée ordinaire mais le tableau de la décomposition des valeurs propres entre liens (Tab 4, E&P94 p. 134) indique qu'on a clairement utilisé un critère prenant en compte la structure des données.

Projected inertia of variable groups

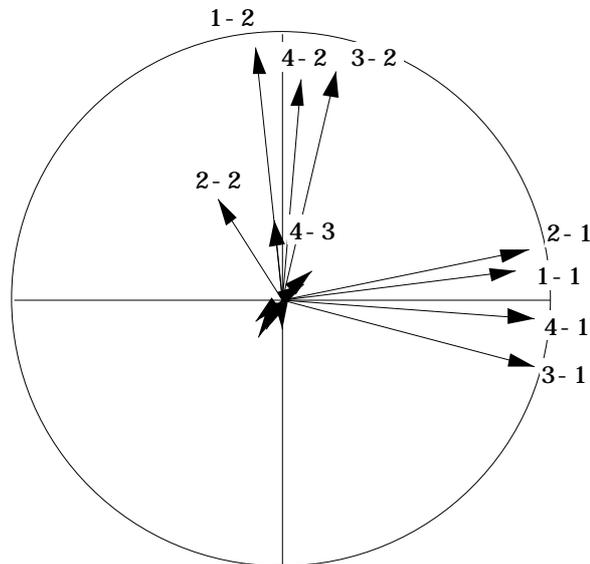
	fac 1	fac 2	fac 3
Tab 1	7.821e-01	6.198e-01	3.735e-01
Tab 2	8.547e-01	4.013e-02	1.438e-02
Tab 3	9.248e-01	4.688e-01	1.800e-01
Tab 4	9.004e-01	2.379e-01	4.743e-02
Total	3.462e+00	1.367e+00	6.154e-01

These values are stored in the file B.++iner
It has 4 rows and 3 columns

Tout se passant dans le même espace \mathbb{R}^n , il est aussi logique de projeter sur le plan des composantes de l'AFMULT les composantes principales des sous-nuages de variables par tableau. Nous avons déjà utilisé cette pratique de nombreuses fois.

File B.++T4a contains scores of axes of separate analysis
It has 16 rows and 3 columns
It is to be used with --T4a.label and --T4a.cat files
File : B.++T4a
| Col. | Mini | Maxi |

1	-2.377e-01	9.480e-01
2	-2.597e-01	9.438e-01
3	-3.542e-01	9.132e-01



On retrouve la figure 7 (E&P94 p. 139), en notant ici qu'on utilise le fait que l'orientation d'une composante principale n'est pas significative et qu'on peut changer son signe pour obtenir éventuellement les vecteurs dans la même direction. La première composante est commune aux quatre groupes tandis que la seconde ne concerne pas le tableau 2 dont on sait qu'il se résume par un seul axe.

En tant qu'analyse d'inertie d'un triplet original, on peut également explorer le rôle des variables par une analyse d'inertie du module DDUtil :

Rows: Inertia analysis							
Input file				B.++up		21	2
----- Relative contributions -----							
Num	Fac 1	Fac 2	Fac 3	Remains	Weight	Cont.	
1	3490	4451	5	2051	4460	636	
2	6976	56	1254	1711	4460	636	
3	5126	227	2889	1757	4460	636	
4	1924	1676	4060	2337	4460	636	
5	14	7483	164	2338	4460	636	
6	7764	567	198	1468	3527	503	
7	7436	545	201	1816	3527	503	
8	9026	24	7	941	3527	503	
9	3932	3311	457	2298	2126	303	
10	6255	1678	487	1578	2126	303	
11	5378	1108	2442	1071	2126	303	
12	295	3281	4050	2372	2126	303	
13	818	4775	0	4406	2126	303	
14	2564	3957	9	3468	2126	303	
15	1555	712	570	7161	2126	303	
16	8432	1	53	1512	2126	303	
17	8646	50	371	932	2126	303	
18	5601	3163	20	1214	2126	303	
19	7122	10	378	2489	1772	252	
20	287	1406	1083	7222	1772	252	
21	6346	2123	312	1217	1772	252	

22	6072	276	107	3543	1772	252
23	5991	2756	65	1187	1772	252
24	6887	1912	188	1011	1772	252
25	1407	4232	77	4282	1772	252
26	8541	5	131	1321	1772	252
27	8144	699	331	824	1772	252

La dernière aide à l'interprétation, qui ne figure pas dans le programme AFMULT des auteurs renvoie à la discussion des relations entre l'AFMULT et analyse canonique généralisée. Manipulant le même objet des k nuages de variables, l'ACG^{6 7} cherche des vecteurs $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_K$ dans les sous espaces de \mathbb{R}^n engendré par les groupes de variables et un vecteur \mathbf{z} qui optimisent :

$$R^2(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_K / \mathbf{z}) = \sum_{k=1}^K \varphi_k \text{Cos}^2(\mathbf{z}, \mathbf{z}_k)$$

En substituant un critère d'inertie projetée à un critère de corrélation pure l'AFMULT est par rapport à l'ACG ce que la régression PLS est à la régression, l'analyse inter-classe est à l'analyse discriminante, l'analyse de co-inertie à l'ACPVI, une version plus stable numériquement (particulièrement pour des variables de groupes très corrélées).

Dans la solution de l'ACG le vecteur \mathbf{z}_k est nécessairement le projeté du vecteur \mathbf{z} sur le sous-espace engendré par les colonnes de \mathbf{X}_k et le Cos^2 est celui de l'angle entre \mathbf{z} et le sous-espace. Pour information, le module indique quelle est la valeur de ce paramètre pour les composantes de l'AFMULT :

```
R2 coefficients = cos2 between canonical scores and subspaces
```

	fac 1	fac 2	fac 3
Tab 1	8.196e-01	9.400e-01	9.359e-01
Tab 2	9.152e-01	1.781e-01	1.294e-01
Tab 3	9.752e-01	9.288e-01	8.658e-01
Tab 4	9.432e-01	9.067e-01	3.223e-01

Dans \mathbb{R}^n , l'AFMULT a donc des propriétés d'optimalité clairement définies et simples d'emploi. Qu'en est-il pour la représentation des individus ?

3 — L'AFMULT comme analyse inter-classes

Nous savons qu'en tant qu'analyse d'un triplet particulier l'AFMULT renvoie à une analyse duale des individus. Une ligne du tableau juxtaposant les K tableaux (présentation théorique), est une colonne du K -tableau (présentation pratique). Les lignes du tableau k sont dans l'espace \mathbb{R}^{p_k} et seule l'ACOM s'intéresse directement à ces K nuages de points. Les lignes du tableau global sont dans \mathbb{R}^p où $p = p_1 + p_2 + \dots + p_K$. Le nuage des n points de \mathbb{R}^p donne la carte ordinaire des produits (B.++co) que nous avons déjà tracée.

La question est de savoir si l'on peut faire une carte des vins par tableau et examiner la cohérence des positions d'un point pour ses multiples représentations. La solution proposée par les auteurs est simple en théorie et remarquablement efficace.

En effet, elle utilise une propriété de base des espaces vectoriels de dimension finie. Une ligne de tableau global peut s'écrire comme somme de K vecteurs de \mathbb{R}^p .

Soit un vecteur $\mathbf{x} = (x_1, x_2, \dots, x_p)$. Notons $\mathbf{x}[k] = (a_1, a_2, \dots, a_p)$ le vecteur défini par ses composantes a_j et la règle $a_j = x_j$ si $j \in P_k$ et $a_j = 0$ sinon. Pour obtenir $\mathbf{x}[k]$, on part du vecteur \mathbf{x} et on annule toutes les valeurs sauf celles des variables du groupe k . Il est évident que $\mathbf{x} = \mathbf{x}[1] + \mathbf{x}[2] + \dots + \mathbf{x}[K]$. En partant du nuage des n points \mathbf{x}_i (points définis par l'ensemble des variables) on obtient K nuages de n points du type $\mathbf{x}_i[k]$. Ceci consiste à projeter sur le sous-espace engendré par les vecteurs de la base canonique de \mathbb{R}^p de rang $j \in P_k$ ($1 \leq j \leq p$) ($1 \leq k \leq K$).

Cette opération immerge les K nuages distincts dans un même espace où se trouve déjà le nuage de référence. La somme $\mathbf{x}_i = \mathbf{x}_i[1] + \mathbf{x}_i[2] + \dots + \mathbf{x}_i[K]$ se conserve par projection et au prix d'une dilatation sans importance se réécrit :

$$\mathbf{x}_i = K \frac{\mathbf{x}_i[1] + \mathbf{x}_i[2] + \dots + \mathbf{x}_i[K]}{K}$$

Dés qu'on possède un vecteur \mathbf{u} normé pour la métrique :

$$\mathbf{Q} = \begin{pmatrix} \frac{1}{\alpha_1} \mathbf{Q}_1 & 0 & & 0 \\ 0 & \frac{1}{\alpha_2} \mathbf{Q}_2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \frac{1}{\alpha_K} \mathbf{Q}_K \end{pmatrix}$$

on peut projeter sur \mathbf{u} les K nuages de n points et prendre pour chaque point le centre de gravité de ses K représentations. Quelle propriété utiliser pour choisir le vecteur \mathbf{u} ?

On veut exprimer les différences entre points donc maximiser la variance des projections des centres de gravités. C'est exactement ce que fait l'analyse du triplet modifié de l'AFMULT. En effet :

$$\mathbf{u} = \mathbf{u}[1] + \mathbf{u}[2] + \dots + \mathbf{u}[K]$$

La coordonnée de la projection de $\mathbf{x}_i[k]$ sur \mathbf{u} vaut :

$$\mathbf{u}^t \mathbf{Q} \mathbf{x}_i[k] = (\mathbf{x}_i[k])^t \mathbf{Q} \mathbf{u} = \frac{1}{\alpha_k} \mathbf{x}_i^t \mathbf{Q}_k \mathbf{u}[k]$$

La coordonnée moyenne des représentations sur \mathbf{u} du point i vaut alors :

$$\mathbf{u}^t \mathbf{Q} \mathbf{x}_i = (\mathbf{x}_i)^t \mathbf{Q} \mathbf{u} = \frac{1}{K} \sum_{k=1}^K \frac{1}{\alpha_k} \mathbf{x}_i^t \mathbf{Q}_k \mathbf{u}[k] = \frac{1}{K} \mathbf{x}_i^t \mathbf{Q} \mathbf{u}$$

La variance de ces moyennes, ou variance inter est donc :

$$\frac{1}{K^2} \sum_{n=1}^K p_i (\mathbf{x}_i^t \mathbf{Q} \mathbf{u})^2 = \frac{1}{K^2} \|\mathbf{X} \mathbf{Q} \mathbf{u}\|_D^2$$

A une constante près, la solution est donnée directement par l'analyse du triplet de l'AFMULT (E&P84). Les coordonnées des projections des $\mathbf{x}_i[k]$ sont dans B.++TCO, les coordonnées des centres de gravité associés à un individu sont dans B.++C1 et elles sont répétées pour les représentations graphiques dans B.++TCrep.

File B.++TCO contains columns scores issued from each table
 It has 84 rows and 3 columns
 It is to be used with --Tcc.label and --Tcc.cat files
 File : B.++TCO

Col.	Mini	Maxi
1	-2.029e+00	2.802e+00
2	-1.852e+00	9.742e-01
3	-9.730e-01	6.544e-01

File B.++C1 contains general scores
 It has 21 rows and 3 columns
 File : B.++C1

Col.	Mini	Maxi
1	-1.254e+00	2.273e+00
2	-9.826e-01	4.043e-01
3	-3.643e-01	2.313e-01

File B.++TCrep contains column scores
 Duplication of general scores
 It has 84 rows and 3 columns
 It is to be used with --Tcc.label and --Tcc.cat files
 File : B.++TCrep

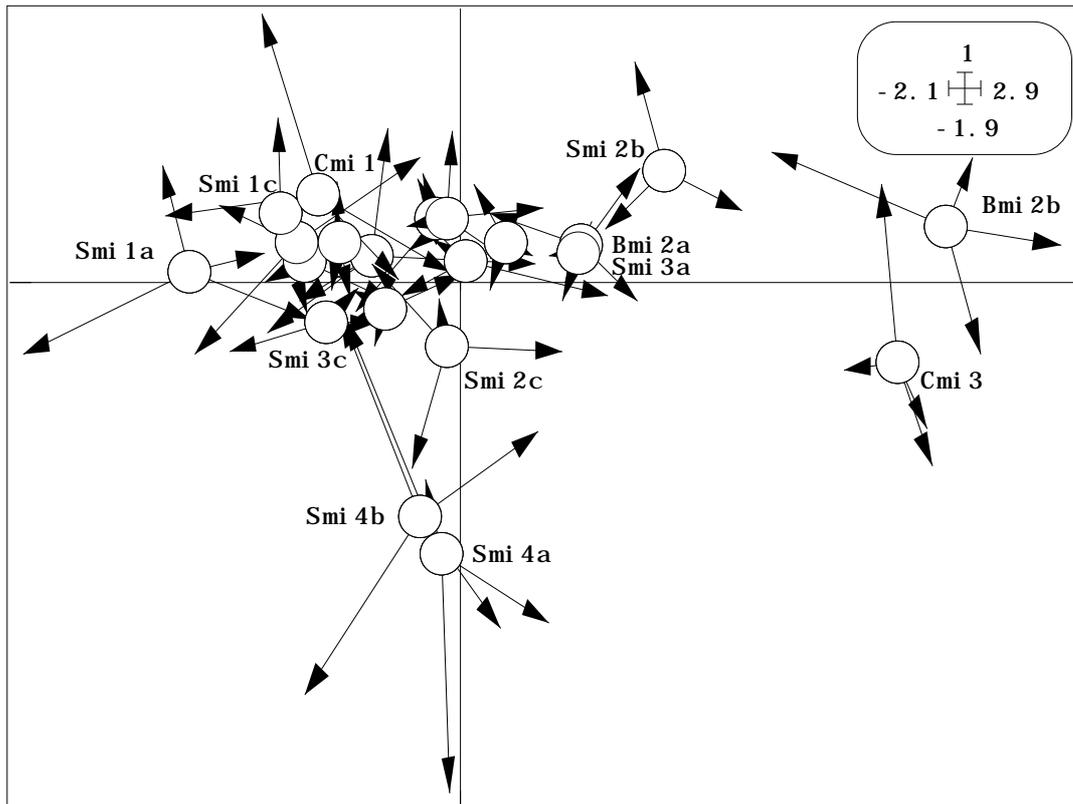
Col.	Mini	Maxi
1	-1.254e+00	2.273e+00
2	-9.826e-01	4.043e-01
3	-3.643e-01	2.313e-01

On attend donc le même résultat avec Scatters ou ScatterClass et les paramètres :

Match two scatters			
HY coordinates file		B.++TCrep	84 3
H-axis column number (default = 1)			
Y-axis column number (default = 2)			
Second HY coordinates file		B.++TCO	84 3
Label file (or #) for rows		Code_Indi	

Stars			
XY coordinates file		B.++TC0	84 3
X-axis column number (default = 1)			
Y-axis column number (default = 2)			
Categories file (.cat)		B!TCc.cat	

Ceci est bien le cas (Cf. fig 5, E&P94 p. 137 ou E&P85 p. 47) :

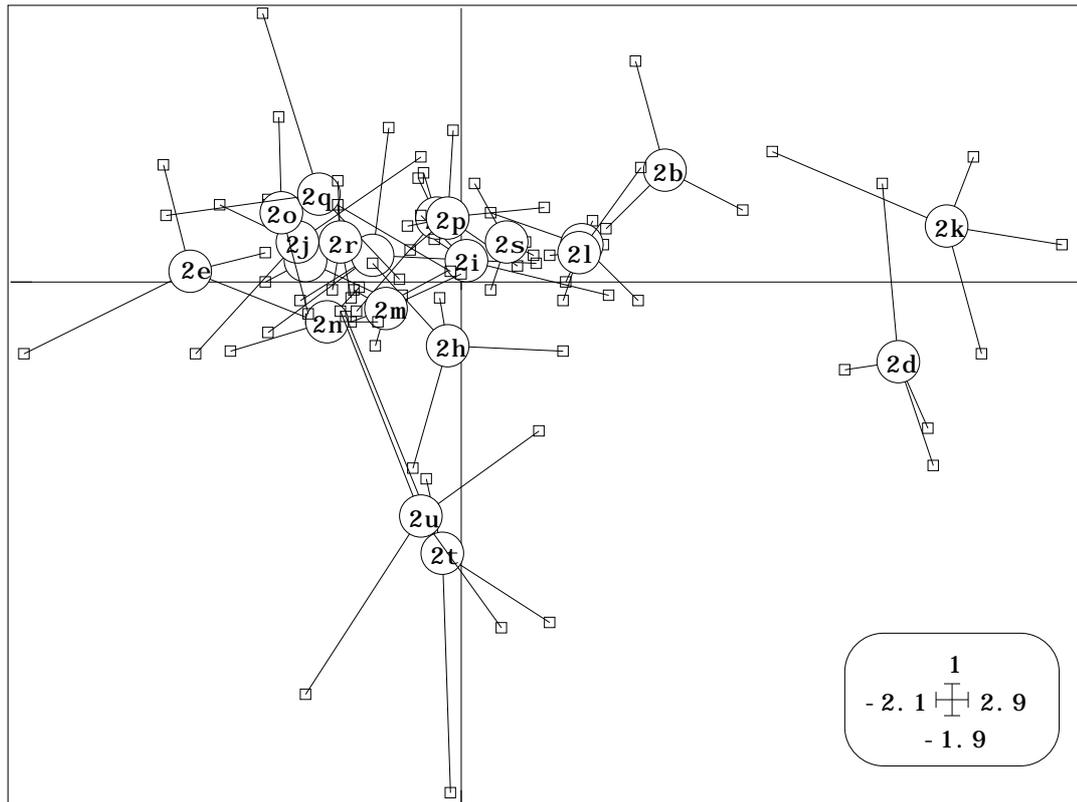


La précision de la reproduction par un tableau de la typologie moyenne engendrée par l'ensemble des tableaux peut alors être décrite par les coefficients de corrélation entre coordonnées propres à chaque tableau et coordonnées générale :

R corrélation coefficients between general scores and one table scores

	fac 1	fac 2	fac 3
Tab 1	8.880e-01	9.564e-01	8.867e-01
Tab 2	9.261e-01	2.212e-01	1.581e-01
Tab 3	9.687e-01	8.936e-01	8.958e-01
Tab 4	9.500e-01	8.684e-01	2.990e-01

On retrouve exactement les résultats du tableau 3 de E&P94 p. 133. La cohérence avec l'approche par les Cos^2 dans l'autre espace est sans défaut. On peut donc considérer l'AFMULT comme une analyse inter-classes, une classe étant définie par un individu, les éléments de la classe par les différentes façons de mesurer cet individu.

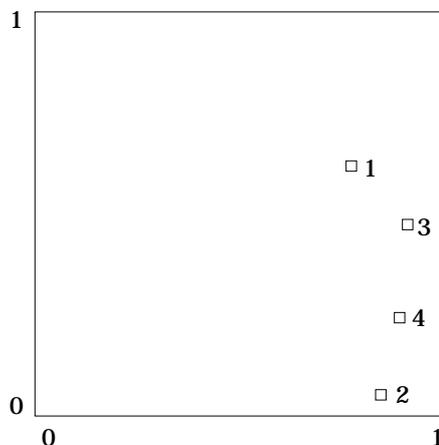


Cette analyse permet, en outre, de mélanger tableaux de variables quantitatives et qualitatives (E&P86⁸). Tout porte alors à croire la conclusion des auteurs (E&P89⁹) :

First applications of M.F.A. were highly encouraging as regards to the practical value of the method. ... The last example clearly shows that to take into account variable groups is not only a technical problem, susceptible of being solved by an appropriate method, but it also can be considered as a methodological problem which enriches the field of data Analysis.

4 — La représentation des groupes en AFMULT

Pour représenter les groupes de variables l'AFMULT propose d'utiliser les inerties projetées du fichier B.++iner (Fig 6, E&P94 p. 138) :



Ceci nous intéresse particulièrement dans le cas des groupes faunistiques que nous avons plusieurs fois étudié.

Transpose			
Input file		Fau	91 16
Output file		FTR	

Par transposition, reprendre le tableau faunistique à 16 stations lignes et 91 espèces -colonnes. En faire l'ACP centrée :

Covariance matrix PCA			
Matrix input file		FTR	16 91

Transposer le triplet obtenu :

Triplet Transpose			
Input file		FTR.cpta	16 91
Generic output file name		Fprovi	

Définir le K-tableau :

TripletToKTab			
.-ta type file		Fprovi.trta	91 16
Row indicator (Default = 1 class)		Blo	10 1
Col indicator (Default = 1 class)			
Output file name		A	

Exécuter l'AFMULT :

Variable groups			
---.kttta input file		A.kttta	91 16
Table weights (default = 1/lambda1)			

On trouve un plan 1-2 important, comme dans les autres approches.

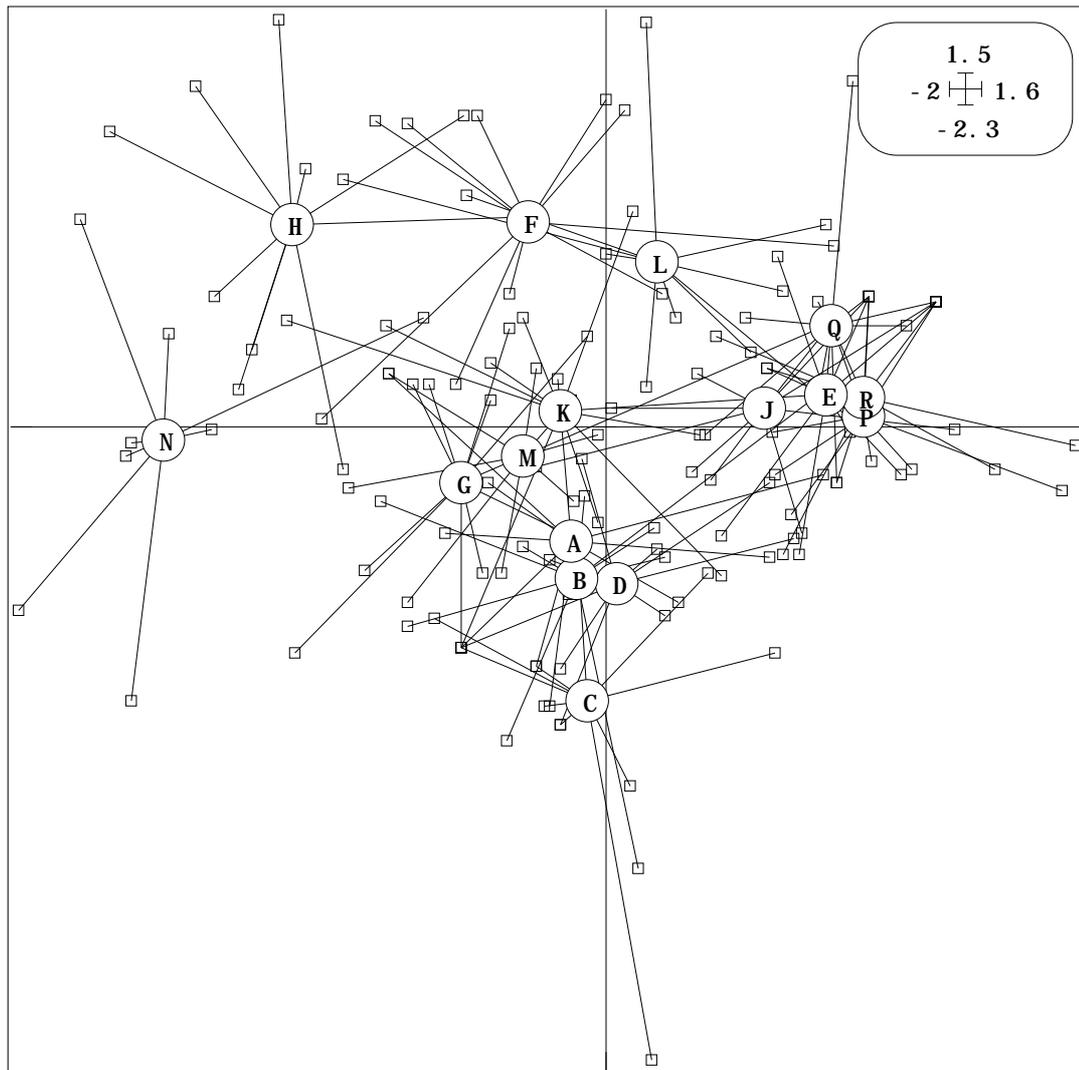
Num.	Ei genval .	R. Iner.	R. Sum	Num.	Ei genval .	R. Iner.	R. Sum
01	+6.4595E+00	+0.2506	+0.2506	02	+4.5864E+00	+0.1779	+0.4285
03	+2.4735E+00	+0.0960	+0.5245	04	+2.1728E+00	+0.0843	+0.6088
05	+1.7652E+00	+0.0685	+0.6773	06	+1.4725E+00	+0.0571	+0.7344
07	+1.2501E+00	+0.0485	+0.7829	08	+1.2041E+00	+0.0467	+0.8296
09	+1.0561E+00	+0.0410	+0.8706	10	+9.0159E-01	+0.0350	+0.9056
11	+6.2810E-01	+0.0244	+0.9299	12	+6.0318E-01	+0.0234	+0.9533
13	+4.8871E-01	+0.0190	+0.9723	14	+4.3706E-01	+0.0170	+0.9892
15	+2.7725E-01	+0.0108	+1.0000	16	+0.0000E+00	+0.0000	+1.0000

On obtient une bonne vérification des calculs en observant que le sous-espace engendré par le groupe 6 (22 espèces) donne un Cos2 de 1, ce qui aurait été obtenu dans toute circonstance, n'a aucune signification et indique clairement qu'une alternative à l'analyse canonique généralisée est indispensable :

R2 coefficients = cos2 between canonical scores and subspaces
 |-----|-----|-----|

	fac 1	fac 2
Tab 1	8. 761e- 01	9. 521e- 01
Tab 2	8. 988e- 01	8. 942e- 01
...		
Tab 6	1. 000e+00	1. 000e+00

La carte des stations du type inter-classe est très satisfaisante et répond clairement à toutes les difficultés soulevées dans STATIS par la représentation simultanée (chaque station est positionné sur la même carte par chaque groupe faunistique) :



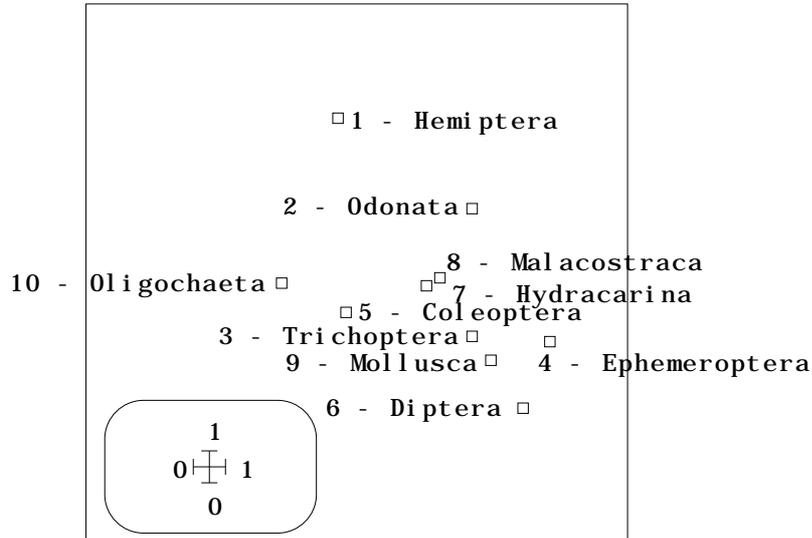
On peut comparer avec la figure de la page 8 dans la fiche “ACOM et valeur typologique” ou la figure p.21 de la fiche “Méthode ACT-STATIS 1”. On s’attend donc à obtenir une carte des groupes compatibles avec les indications de valeurs typologiques des groupes déjà connue. Il n’en est rien, ni numériquement, ni graphiquement :

Projected inertia of variable groups

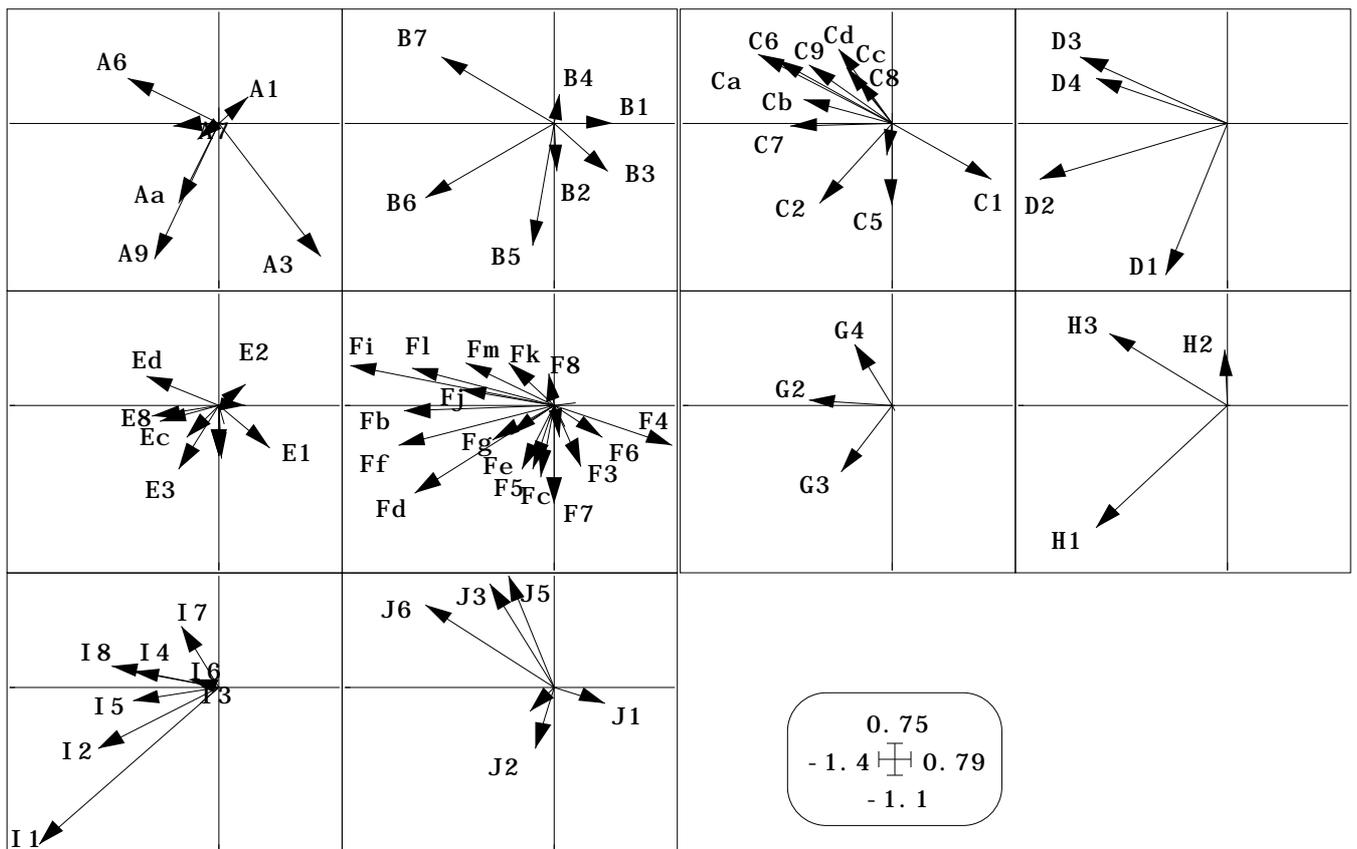
	fac 1	fac 2
Tab 1	4. 675e- 01	7. 856e- 01
Tab 2	7. 152e- 01	6. 169e- 01
Tab 3	7. 154e- 01	3. 769e- 01
...		

Tab 10	3.640e-01	4.784e-01
Total	6.459e+00	4.586e+00

These values are stored in the file A.++iner
It has 10 rows and 2 columns



La carte des espèces est pourtant proche des cartes obtenues par les autres analyses :



En fait la version retenue utilise une surpondération qui donne beaucoup d'importance aux tableaux associant un grand nombre de variables à une structure faible. Il ne devrait pas en être ainsi si la surpondération intégrait également le nombre d'espèces, ce que nous avons systématiquement fait dans les autres méthodes. Reprendre les données initiales et préparer la structure des données :

InitKtab			
Matrix input file		Fau	91 16
Row indicator (Default = 1 class)		Blo	10 1

Centrer par espèces :

CentringKtab			
---.ktp input file		Fau!.ktp	
Matrix input file		Fau	91 16
Row weighting (default 1/nk)			
Column weighting (default 1/pj)			
Centring option (no default)		3	
Option: output file name		Z	

Exécuter l'AFMULT :

Variable groups			
---.ktt input file		Z.ktt	91 16
Table weights (default = 1/lambda1)			

Observer qu'on retrouve exactement le même résultat. Cette opération n'a aucun effet, parce que la première valeur propre tient compte du changement de pondération des triplets initiaux. On pourra donc exécuter STATIS, ACOM ou AFMULT sur les mêmes données strictement.

Il convient alors de noter que l'inertie projetée, utilisée dans l'AFMULT, est fortement influencée par la surpondération. En utilisant un pourcentage de l'optimum (définition en fait valide surtout pour le premier axe), ce qui disparaît est évidemment cet optimum. Les fortes valeurs propres de rang 1 et 2 du groupe 4 (Ephéméroptères) jouent un rôle important dans la valeur donnée par STATIS et sont dissimulées par la mesure donnée par l'AFMULT. Faut-il en conclure que l'une serait meilleure que l'autre ?

La réponse est certainement pas. Il est plus que vraisemblable que l'une ou l'autre, en face d'un certain type de données, pour un certain objectif, seront prises en défaut. La synergie méthode-donnée, à laquelle les théoriciens sont insensibles, est en fait un élément clef de l'analyse des données.

STATIS a des principes mathématiques clairement établis. Ils s'en suit des propriétés pratiques précises. Vouloir compléter ces propriétés optimales par des pratiques empiriques (représentation simultanée des éléments communs, projections ou trajectoires) conduit à des difficultés. Des trois, c'est la seule qui ouvre la discussion vers la variabilité de la structure derrière la structure compromis.

L'ACOM a des principes mathématiques clairement établis. Ils s'en suit des propriétés pratiques précises. On peut y faire K analyses coordonnées du point de vue induction de structures des éléments spécifiques vers les éléments communs. La représentation simultanée des éléments communs n'a pas de caractère d'optimalité.

L'AFMULT a des principes mathématiques clairement établis. Ils s'en suivent des propriétés pratiques précises. Elle seule introduit une double canonicité dans la représentation simultanée des objets, mais n'est conçue que pour des individus en commun. Nous verrons ce qu'il en est dans l'autre cas. Sa force est celle de l'analyse d'un triplet classique.

On peut y introduire par exemple les propriétés mixtes axe-composante, qui correspondent tout à fait à l'introduction d'un critère d'inertie dans celui de l'analyse canonique généralisée. En effet, considérons un vecteur unitaire \mathbf{z} de \mathbb{R}^n \mathbf{D} -normé et un vecteur unitaire \mathbf{u} de \mathbb{R}^p \mathbf{Q} -normé. Sur \mathbf{u} on projette le nuage des points $\mathbf{x}_i[k]$. On peut s'intéresser à la covariance entre ces coordonnées et les composantes du vecteur \mathbf{z} :

$$Cov(\mathbf{z}, k) = \sum_{i=1}^n p_i \left((\mathbf{x}_i[k])^t \mathbf{Q} \mathbf{u} \right) \mathbf{z}_i = \frac{1}{\alpha_k} (\mathbf{X}_k \mathbf{Q}_k \mathbf{u} | \mathbf{z})_{\mathbf{D}}$$

La somme de ces covariances :

$$Cov(\mathbf{z}, k) = \sum_{k=1}^K \frac{1}{\alpha_k} (\mathbf{X}_k \mathbf{Q}_k \mathbf{u} | \mathbf{z})_{\mathbf{D}} = (\mathbf{X} \mathbf{Q} \mathbf{u} | \mathbf{z})_{\mathbf{D}}$$

est optimisée par le couple du premier axe principal et de la première composante principale. L'optimum est égal à la racine carrée de la première valeur propre.

Importer dans l'une des trois la description des critères optimisées par l'une des deux autres ne semblent pas du meilleur effet. Le plus simple est de prendre ce qu'il y a de meilleur dans chacune. Ce qui est probable, c'est que les résultats communs enregistrés par les trois correspondront le plus souvent à l'essentiel de l'information attendue. L'essentiel sera, au plan méthodologique, d'articuler les principes de chaque méthode, les objectifs de chaque analyse et les propriétés spécifiques de chaque jeu de données.

Références

¹ Escofier, B. & Pages, J. (1982a) Comparaison de groupes de variables définies sur le même ensemble d'individus. Rapport de recherche n°149, ISSN 0249-6399. INRIA, Domaine de Voluceau-Rocquencourt, BP 105, 78153 Le Chesnay cedex, France. 1-115.

Escofier, B. & Pages, J. (1982b) Comparaison de groupes de variables. 2ème partie : un exemple d'applications. Rapport de recherche n°165. INRIA, Domaine de Voluceau-Rocquencourt, BP 105, 78153 Le Chesnay cedex, France. 1-35.

² Escofier, B. & Pages, J. (1985) Mise en œuvre de l'analyse factorielle multiple pour les tableaux numériques qualitatifs ou mixtes. Rapport de recherche n°429. INRIA, Domaine de Voluceau-Rocquencourt, BP 105, 78153 Le Chesnay cedex, France. 1-54 + annexes.

³ Escofier, B. & Pages, J. (1994) Multiple factor analysis (AFMULT package). *Computational Statistics and Data Analysis* : 18, 121-140.

⁴ Tomassone, R. & Flanzky, C. (1977) Présentation synthétique de diverses méthodes d'analyse de données fournies par un jury de dégustateurs. *Ann. Technol. Agric* : 26, 373-418.

⁵ Escofier, B. & Pages, J. (1984) L'analyse factorielle multiple : une méthode de comparaison de groupes de variables. In : *Data Analysis and Informatics III*. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 41-55.

⁶ Carrol, J.D. (1968) A generalization of canonical analysis to three or more sets of variables. *Proceeding of the 76th Convention of the American Psychological Association* : 3, 227-228.

⁷ Lazraq, A., Cléroux, R. & Kiers, H.A.L. (1992) Mesures de liaison vectorielle et généralisation de l'analyse canonique. *Revue de Statistique Appliquée* : 39, 23-35.

⁸ Escofier, B. & Pages, J. (1986) Le traitement des variables qualitatives et des tableaux mixtes par analyse factorielle multiple. In : *Data Analysis and Informatics IV*. Diday, E. & Coll. (Eds.) Elsevier, North-Holland. 179-191.

⁹ Escofier, B. & Pages, J. (1989) Multiple factor analysis: results of a three-year utilization. In : *Multiway data analysis*. Coppi, R. & Bolasco, S. (Eds.) Elsevier Science Publishers B.V., North-Holland. 277-285.

