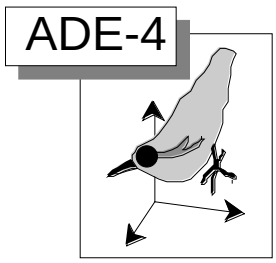


MCA

MCA : Correlation ratio - cmta

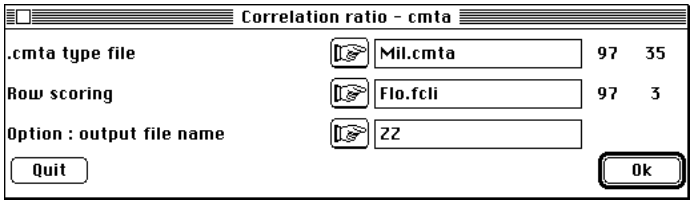
Utilitaire de statistique descriptive multivariée.

L'objectif est de décrire le lien existant entre des variables quantitatives et un tableau d'analyse des correspondances multiples (MCA : Multiple Correspondence Analysis).

Soit **A** un tableau de variables qualitatives avec *n* lignes et *v* colonnes (*v* est le nombre de variables). La variable *j* a *m*(*j*) modalités. Il doit être lu par le module CategVar : Read Categ File. Le fichier --.cat qui en résulte est le point d'entrée dans le module MCA : Multiple Correspondence Analysis qui autorise l'introduction des poids des lignes (individus) et assure l'analyse. Ces poids des lignes peuvent être uniformes (c'est le cas le plus fréquent) mais peuvent aussi dériver d'une autre analyse comme en co-inertie (CoInertia : Matching two statistical triplets).

Un tableau de variables quantitatives **X** contient des scores numériques des individus. Pour chaque colonne de **X** (score) on veut mesurer l'association avec chaque variable qualitative de **A** et l'association moyenne avec toutes les variables de **A**. Ceci se fait par le biais du rapport de corrélation qui est le rapport de la variance inter-classes à la variance totale. Les classes sont définies par les modalités d'une variable de **A** : la variance totale est calculée avec les poids des individus de l'ACM, la variance inter-classe (ou variance des moyennes par classes) est calculée avec les poids des classes qui en dérivent (le poids d'une classe est la somme des poids des points qui sont dedans).

L'option utilise une seule fenêtre de dialogue :

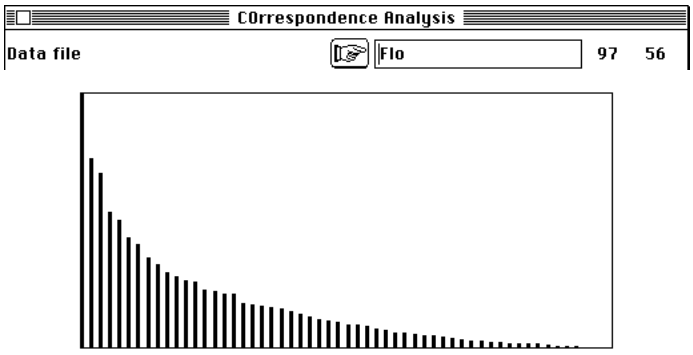


Nom du fichier d'accès à l'analyse des correspondances multiples.

Nom du fichier des scores.

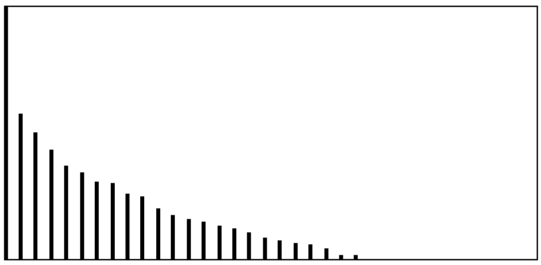
Nom du fichier binaire de sortie (création en option). Si aucun nom n'est utilisé, les résultats sont seulement consignés dans le listing.

Utiliser les cartes Mafragh et Mafragh+2 de la pile ADE-4•Data[1]. On obtient un tableau floristique Flo (97 relevés-lignes et 56 espèces-colonnes) et un tableau Mil (97 relevés-lignes et 11 variables qualitatives-colonnes). Faire l'AFC du tableau floristique (COA : COrrespondence Analysis):



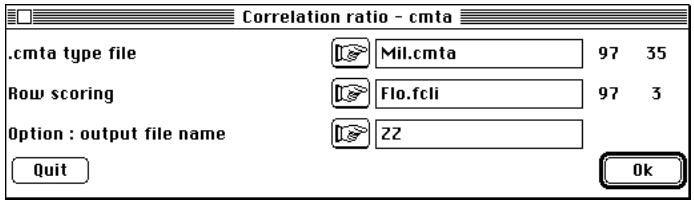
On garde 3 facteurs. Lire le tableau Mil (CategVar : Read Categ File) :

Introduire les poids des relevés dérivés du tableau floristique dans l'analyse du tableau de milieu (MCA : Multiple Correspondence Analysis) :



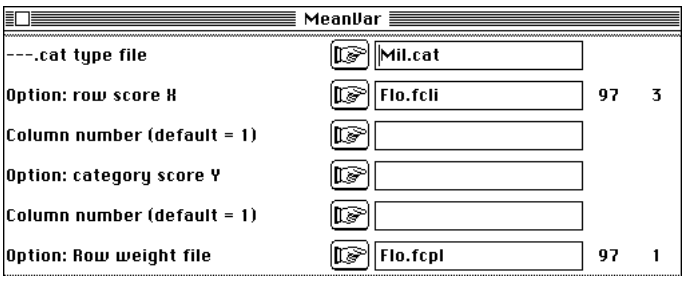
On garde un facteur.

On utilise la présente option pour mesurer la relation entre variables de milieu et scores des relevés dans l'ordination floristique :

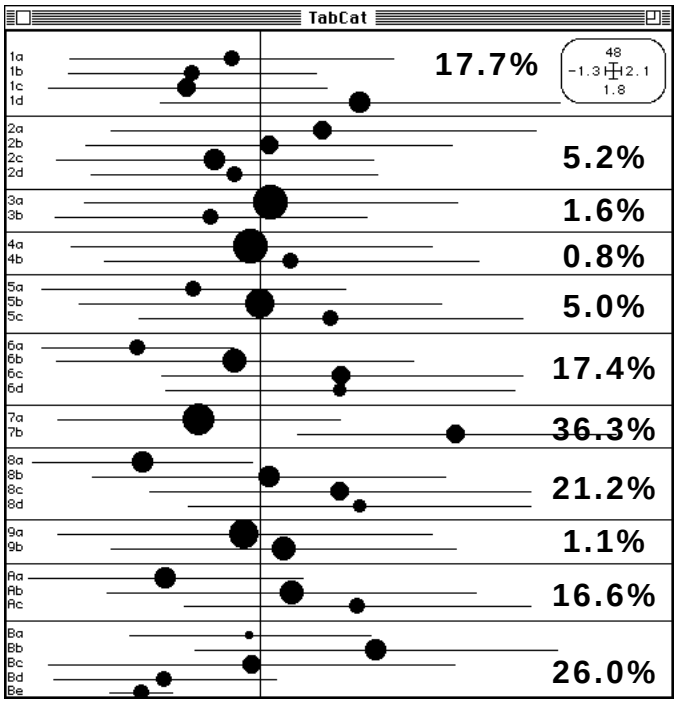



```
Categorical variables: file Mil.cmta
Number of rows: 97, variables: 11, categories: 35

Row scores: file Flo.fcli
Number of rows: 97, columns: 3

Variable :   1--> r=    0.177   0.003   0.026
Variable :   2--> r=    0.052   0.067   0.077
Variable :   3--> r=    0.016   0.025   0.012
Variable :   4--> r=    0.008   0.004   0.055
Variable :   5--> r=    0.050   0.018   0.054
Variable :   6--> r=    0.174   0.101   0.032
Variable :   7--> r=    0.363   0.000   0.031
Variable :   8--> r=    0.212   0.180   0.040
Variable :   9--> r=    0.011   0.135   0.045
Variable :  10--> r=    0.166   0.118   0.013
Variable :  11--> r=    0.260   0.122   0.011

Mean     :    --> r=    0.135   0.070   0.036

File ZZ contains the correlation ratios between
each categorical variable and each score
It has 11 rows and 3 columns
```

La signification de ces statistiques se comprend bien en utilisant le graphique fourni par TabCat : MeanVar :

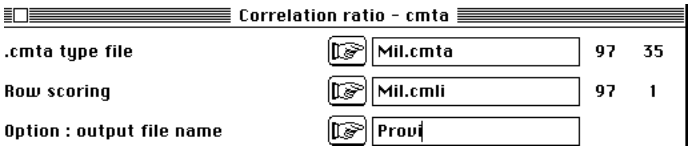


Chaque relevé est positionné par le premier score factoriel de l'AFC, chaque modalité de chaque variable prend sur ce score une position moyenne, la variance de ces moyennes mesure la liaison entre le score et chaque variable :



Les rapports de corrélation ont été rajoutés à la figure.

Noter que l'usage de l'option sur les fichiers :




```
Categorical variables: file Mil.cmta
Number of rows: 97, variables: 11, categories: 35

Row scores: file Mil.cmli
Number of rows: 97, columns: 1

Variable :   1--> r=    0.458
Variable :   2--> r=    0.096
Variable :   3--> r=    0.321
Variable :   4--> r=    0.124
•••
Variable :   9--> r=    0.335
Variable :  10--> r=    0.624
Variable :  11--> r=    0.245 <•••

Mean     :    --> r=    0.378 <<---
```

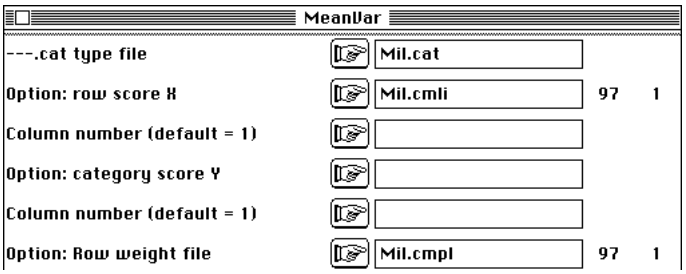
donne exactement les résultats de l'ACM :

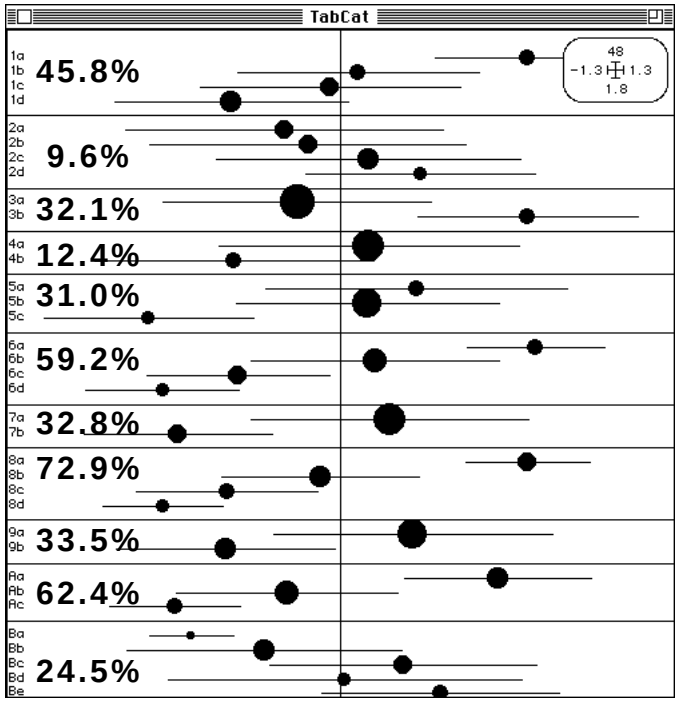
```
Variable : 11
> Categ=   1 Weight=  0.059  -0.939
> Categ=   2 Weight=  0.344  -0.481
> Categ=   3 Weight=  0.250   0.405
> Categ=   4 Weight=  0.168   0.031
> Categ=   5 Weight=  0.179   0.637
----------------------> r=    0.245 <•••
```

En particulier, la moyenne des rapports de corrélation sur le premier axe de l'ACM est la première valeur propre :

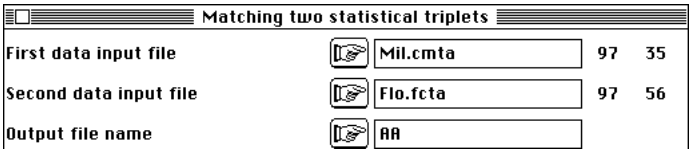
```
          Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
     --->>01   +3.7833E-01 +0.1734 +0.1734   |02   +2.1751E-01 +0.0997 +0.2731 |
          03   +1.8953E-01 +0.0869 +0.3600   |04   +1.6462E-01 +0.0755 +0.4354 |
```

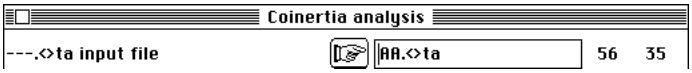
L'ACM donne le code relevés qui maximise la moyenne des rapports de corrélation, ce qui s'exprime dans la figure :

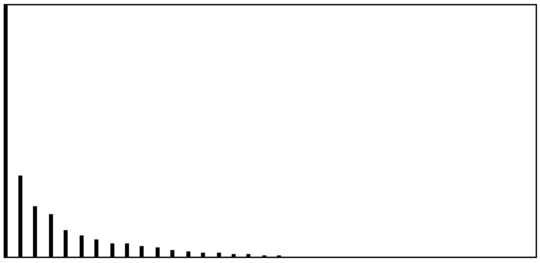




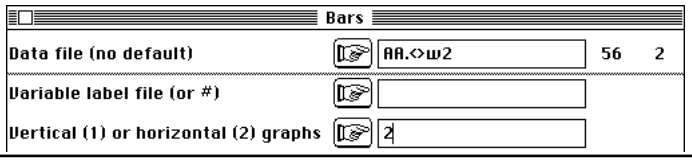
La très grande différence entre ordination floristique et ordination mésologique mérite d'être explicitée par l'analyse de co-inertie (CoInertia : Matching two statistical triplets) :

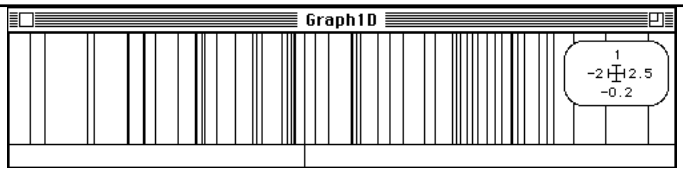




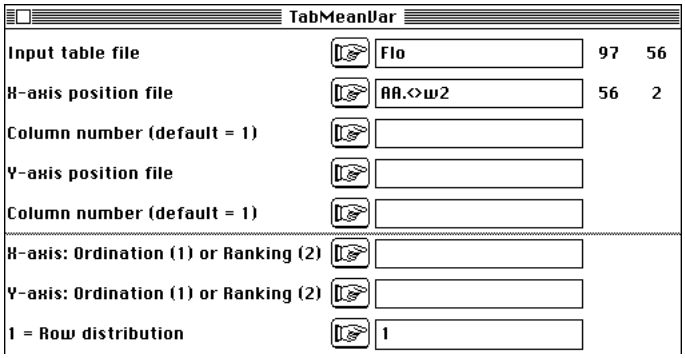


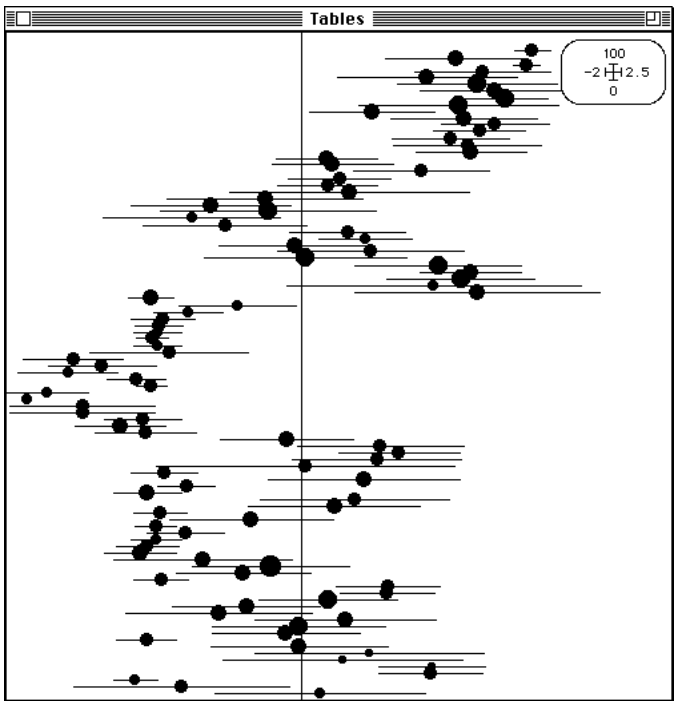
Les espèces sont positionnées sur le premier axe par un code numérique de variance 1:



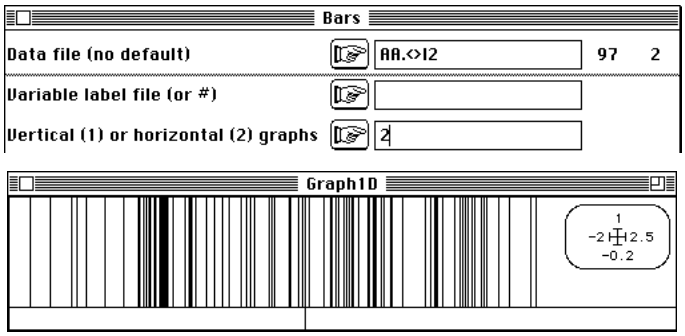


Chaque relevé prend une position moyenne par son contenu spécifique :

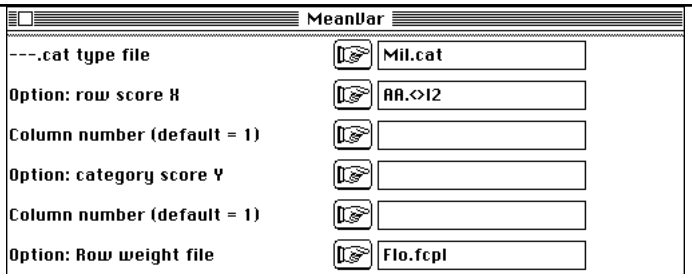


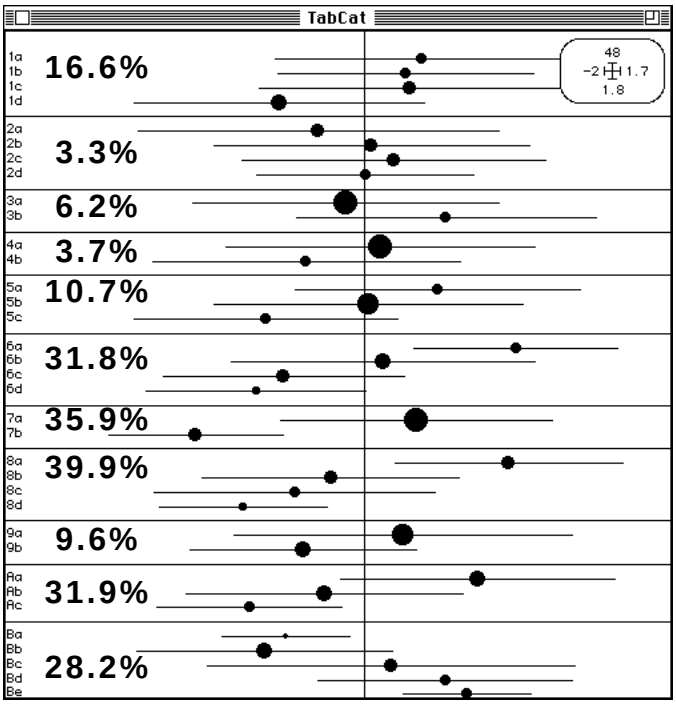


Ce code relevé est utilisé directement :

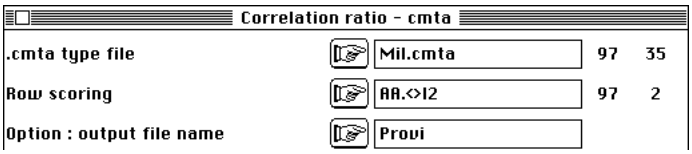


Il permet de positionner les modalités :





La présente option permet d'ajouter les rapports de corrélation :




```
Categorical variables: file Mil.cmta
Number of rows: 97, variables: 11, categories: 35

Row scores: file AA.<>l2
Number of rows: 97, columns: 2

Variable :   1--> r=    0.166   0.209
Variable :   2--> r=    0.033   0.218
Variable :   3--> r=    0.062   0.007
Variable :   4--> r=    0.037   0.016
Variable :   5--> r=    0.107   0.003
Variable :   6--> r=    0.318   0.047
Variable :   7--> r=    0.359   0.074
Variable :   8--> r=    0.399   0.066
Variable :   9--> r=    0.096   0.075
Variable :  10--> r=    0.319   0.041
Variable :  11--> r=    0.282   0.274

Mean     :    --> r=    0.198   0.094
```

Noter alors que la variance de départ (espèces) est égale à 1. La variances du code relevés vaut 0.7638 :

```
-------------------------------------------------------------------
|Num| Covaria.| Varian1 | varian2 | Correla.| INER1   | INER2   |
-------------------------------------------------------------------
|  1|   0.3889|   0.3444|   0.7638|   0.7583|   0.3783|   0.8691|
```

```
-------------------------------------------------------------
|  2|    0.221|   0.2029|   0.5212|   0.6794|   0.2175|   0.6491|
-------------------------------------------------------------
```

Cette variance ne peut atteindre le maximum 0.869 proposé par l'AFC du tableau floristique :

```
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +8.6915E-01 +0.1043 +0.1043   |02   +6.4911E-01 +0.0779 +0.1823 |
```

En moyenne (Cf. ci-dessus) une proportion de 0.198 est inter-modalité. 0.198 est inférieur au maximum 0.378 proposé par l'ACM du tableau mésologique :

```
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +3.7833E-01 +0.1734 +0.1734   |02   +2.1751E-01 +0.0997 +0.2731 |
```

Au total, en partant d'une variance de 1 pour les espèces, on obtient une variance inter-modalités moyenne de 0.7638*0.198 = 0.1512, ce qui est la première valeur propre de l'analyse de co-inertie. En effet, l'analyse de co-inertie maximise le produit de la variance du score des relevés (obtenu par averaging sur un code espèce de variance 1, critère d'AFC) par la moyenne des rapports de corrélation (critère d'ACM).

```
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +1.5124E-01 +0.4343 +0.4343   |02   +4.8819E-02 +0.1402 +0.5745 |
```

L'analyse de co-inertie fait donc, comme toujours, deux analyses simultanément. La présente option facilite son interprétation en permettant de redécomposer le critère optimisé dans le cas d'une ACM.

On retrouve la même information par les options Discrimin : Anova1-FF et StatUtil : R2-Anova1-FF mais **uniquement pour les cas de la distribution uniforme des poids des lignes**. L'approche inférentielle associée (ANOVA) est alors privilégiée.

1 Belair, G. de & Bencheikh-Lehocine, M. (1987) Composition et déterminisme de la végétation d'une plaine côtière marécageuse : La Mafragh (Annaba, Algérie). Bulletin d'Écologie : 18, 4, 393-407.

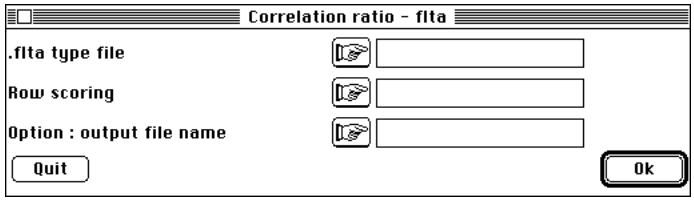
Belair, G. de. (1981) Biogéographie et aménagement : la plaine de La Mafragh (Annaba, Algérie). Thèse de 3° cycle. Université Paul Valéry, Montpellier. 1-150.

MCA : Correlation ratio - flta

Utilitaire de statistique descriptive multivariée (version de MCA : Correlation ratio - cmta pour variables floues).

L'objectif est de décrire le lien existant entre des variables quantitatives et un tableau d'analyse des correspondances floues (MCA : Fuzzy Correspondence Analysis). Bien que mathématiquement les deux options utilisent strictement le même calcul, la signification des rapports de corrélations avec des variables floues est plus complexe. Les utilitaires graphiques associés ne sont pas encore disponibles.

L'option utilise une seule fenêtre de dialogue :

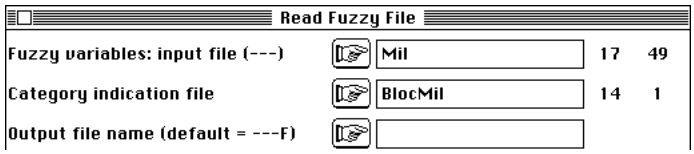


Nom du fichier d'accès à l'analyse des correspondances floues.

Nom du fichier des scores.

Nom du fichier binaire de sortie (création en option). Si aucun nom n'est utilisé, les résultats sont seulement consignés dans le listing.

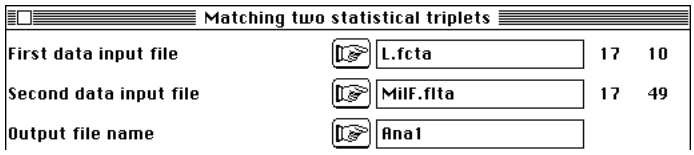
Utiliser la carte Amphibiens de la pile ADE-4•Data [1]. On obtient un tableau faunistique à 17 lignes-relevés et 10 colonnes-espèces et un tableau d'habitat à 17 lignes-relevés et 49 colonnes-modalité d'habitat (réparties en 14 variables de milieu). Lire le fichier de variables floues (FuzzyVar : Read Fuzzy File) :



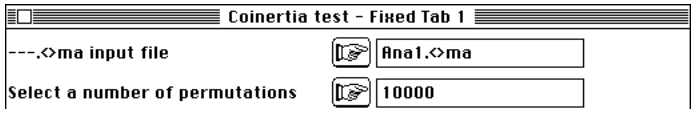
Faire l'analyse des correspondances du tableau faunistique (COA : COrrespondence Analysis) :

Faire l'ACF du tableau Mil (MCA : Fuzzy Correspondence Analysis) :

Coupler les deux tableaux (CoInertia : Matching two statistical triplets) :

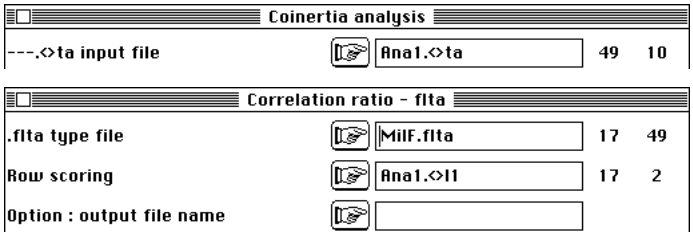


Tester l'existence d'une co-structure (CoInertia : Coinertia test - Fixed Tab 1) :



```
Warning: non uniform weight. The results from permutations
are valid only if the row weights come from the fixed table.
-----
The fixed table is table 1: L.fcta
number of random matching: 10000   Observed: 0.244645
Histogramm:  minimum = 0.155525, maximum = 0.520037
number of simulation X<Obs: 3629 (frequency: 0.362900)
number of simulation X>=Obs: 6371 (frequency: 0.637100)

   |
   |*****
   |******************
   |****************************************
●-->|**************************************************
   |**********************************************
   |****************************************
   |*****************************
   |******************
   |*************
   |***********
   |********
   |*****
   |***
   |**
   |*
   |
   |
   |
   |
```

Il n'y a aucune corrélation significative entre les deux typologies. L'interprétation par les rapports de corrélation n'a donc pas de sens. Exécutons cependant le calcul :




```
Categorical variables: file MilF.flta
Number of rows: 17, variables: 14, categories: 49

Row scores: file Ana1.<>l1
Number of rows: 17, columns: 2

Variable :   1--> r=    0.244   0.121
Variable :   2--> r=    0.399   0.486
Variable :   3--> r=    0.463   0.203
Variable :   4--> r=    0.364   0.101
Variable :   5--> r=    0.277   0.116
Variable :   6--> r=    0.281   0.013
Variable :   7--> r=    0.286   0.073
Variable :   8--> r=    0.321   0.112
Variable :   9--> r=    0.048   0.133
Variable :  10--> r=    0.031   0.456
Variable :  11--> r=    0.418   0.183
Variable :  12--> r=    0.135   0.174
Variable :  13--> r=    0.053   0.609
Variable :  14--> r=    0.225   0.001

Mean     :    --> r=    0.253   0.199
```

Il n'y a pas de signification statistique à ces nombres mais ils nous permettent de comprendre exactement le mode de calcul. Prenons la variable 5 (r = 0.277). La variable floue brute extraite directement des données d'origine est éditée en A (ci-dessous). On passe en pourcentage par lignes (variable floue) en B1. On remplace la ligne (0, 0, 0) par la moyenne pondérée des autres en B2. La pondération utilisée (issue de l'AFC) est éditée en C. On calcule pour chaque modalité dans chaque site une note d'association qui est simplement le produit de la fréquence de la modalité dans le site par le poids du site (D).

Les notes sont sommées par modalités ce qui redonne les poids des modalités (0.339, 0.407, 0.254) puis divisées par cette somme marginale ce qui fournit une distribution de fréquence (somme par colonne unitaire) sur l'ensemble des sites pour chaque modalité en E. La première colonne de Ana1.<>l1 est un code numérique édité en F.

Les trois distributions (E) donnent sur le code (F) trois moyennes (0.371, 0.035, -0.551) soit une variance inter de $0.339*0.371^2 + 0.407*0.0035^2 + 0.254*(-0.551)^2 = 0.124$. Divisée par la variance totale du code (F) pour la distribution des poids (C) qui vaut 0.4499 on retrouve r = 0.124/0.4499 = 0.277.

L'essentiel de cette valeur provient du site 7 qui utilise la modalité rare 3. On ne peut lui accorder aucune signification biologique.

A

	1	2	3
1	1	3	0
2	1	3	0
3	3	0	0
4	1	3	0
5	3	0	0
6	3	0	0
7	0	1	3
8	2	2	0
9	0	1	3
10	1	3	0
11	0	0	0
12	1	3	0
13	1	3	0
14	1	3	0
15	3	0	0
16	1	3	0
17	1	3	0

B1

	1	2	3
1	0.250	0.750	0.000
2	0.250	0.750	0.000
3	1.000	0.000	0.000
4	0.250	0.750	0.000
5	1.000	0.000	0.000
6	1.000	0.000	0.000
7	0.000	0.250	0.750
8	0.500	0.500	0.000
9	0.000	0.250	0.750
10	0.250	0.750	0.000
11	0.000	0.000	0.000
12	0.250	0.750	0.000
13	0.250	0.750	0.000
14	0.250	0.750	0.000
15	1.000	0.000	0.000
16	0.250	0.750	0.000
17	0.250	0.750	0.000

B2

	1	2	3
1	0.250	0.750	0.000
2	0.250	0.750	0.000
3	1.000	0.000	0.000
4	0.250	0.750	0.000
5	1.000	0.000	0.000
6	1.000	0.000	0.000
7	0.000	0.250	0.750
8	0.500	0.500	0.000
9	0.000	0.250	0.750
10	0.250	0.750	0.000
11	0.339	0.407	0.254
12	0.250	0.750	0.000
13	0.250	0.750	0.000
14	0.250	0.750	0.000
15	1.000	0.000	0.000
16	0.250	0.750	0.000
17	0.250	0.750	0.000

C

	1
1	0.015
2	0.015
3	0.059
4	0.015
5	0.044
6	0.059
7	0.265
8	0.059
9	0.044
10	0.074
11	0.088
12	0.088
13	0.029
14	0.059
15	0.029
16	0.029
17	0.029

D

	1	2	3
1	0.004	0.011	0.000
2	0.004	0.011	0.000
3	0.059	0.000	0.000
4	0.004	0.011	0.000
5	0.044	0.000	0.000
6	0.059	0.000	0.000
7	0.000	0.066	0.199
8	0.029	0.029	0.000
9	0.000	0.011	0.033
10	0.018	0.055	0.000
11	0.030	0.036	0.022
12	0.022	0.066	0.000
13	0.007	0.022	0.000
14	0.015	0.044	0.000
15	0.029	0.000	0.000
16	0.007	0.022	0.000
17	0.007	0.022	0.000

E

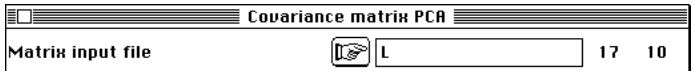
	1	2	3
1	0.011	0.027	0.000
2	0.011	0.027	0.000
3	0.173	0.000	0.000
4	0.011	0.027	0.000
5	0.130	0.000	0.000
6	0.173	0.000	0.000
7	0.000	0.163	0.782
8	0.087	0.072	0.000
9	0.000	0.027	0.130
10	0.054	0.135	0.000
11	0.088	0.088	0.088
12	0.065	0.163	0.000
13	0.022	0.054	0.000
14	0.043	0.108	0.000
15	0.087	0.000	0.000
16	0.022	0.054	0.000
17	0.022	0.054	0.000

F

	1
1	0.119
2	0.119
3	0.361
4	0.119
5	0.260
6	0.361
7	-0.602
8	2.185
9	-0.264
10	0.204
11	-0.531
12	-0.078
13	-0.482
14	0.225
15	0.662
16	0.060
17	0.060

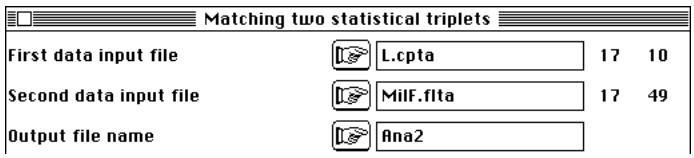
Ces coefficients ont surtout une valeur relative. Pour 17 sites un coefficient de 0.25 n'a pas la même signification que pour 200 sites. On ne peut leur accorder de valeurs absolue mais seulement le statut d'aide à l'interprétation.

Reprenons les mêmes données et faisons l'ACP centrée du tableau faunistique (PCA : Covariance matrix PCA) :

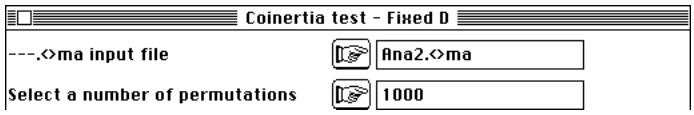


On garde le premier facteur. Refaire alors l'ACM floue sur la pondération uniforme (MCA : Fuzzy Correspondence Analysis) :

Coupler les deux analyses (CoInertia : Matching two statistical triplets) :



Tester la co-structure (CoInertia : Coinertia test - Fixed D) :

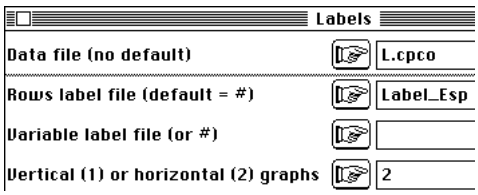


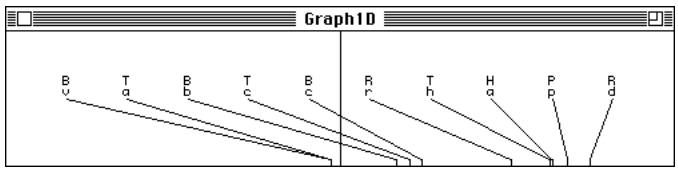

```
number of random matching: 1000   Observed: 0.577709
Histogramm:  minimum = 0.236051, maximum = 0.686794
number of simulation X<Obs: 985 (frequency: 0.985000)
number of simulation X>=Obs: 15 (frequency: 0.015000)

    |**
    |******
    |***************
    |*********************************
    |************************************
    |*************************************************
    |*****************************************************
    |*******************************************
    |****************************************
    |*****************************************
    |**********************
    |*****************
    |*********
    |***********
    |****
●-> |***
    |*
    |*
    |
    |
```

La signification statistique est sûre au risque de 5%. Effectuer l'analyse (CoInertia : Coinertia analysis) :

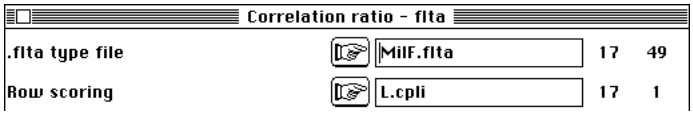
Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	0.5894	2.206	0.2246	0.8373	2.229	0.3574

Vérifier que le premier axe de co-inertie côté faune est presque exactement l'axe de son analyse en composante principale (correlation de -.998) et simplifier l'interprétation en réduisant la discussion à la relation F1 de l'ACP et variables floues. Noter que cet axe est un effet taille (Graph1D : Labels) :



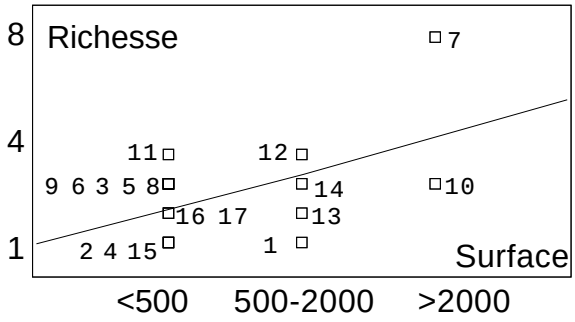


Utiliser alors la présente option :




```
Categorical variables: file MilF.flta
Number of rows: 17, variables: 14, categories: 49

Row scores: file L.cpli
Number of rows: 17, columns: 1

Variable :   1--> r=    0.506
Variable :   2--> r=    0.364
Variable :   3--> r=    0.121
Variable :   4--> r=    0.056
Variable :   5--> r=    0.266
Variable :   6--> r=    0.015
Variable :   7--> r=    0.113
Variable :   8--> r=    0.066
Variable :   9--> r=    0.095
Variable :  10--> r=    0.152
Variable :  11--> r=    0.074
Variable :  12--> r=    0.033
Variable :  13--> r=    0.190
Variable :  14--> r=    0.107

Mean     :    --> r=    0.154
```

Un de ces rapports l'emporte sur les autres. Il concerne la variable 1 (surface occupée par l'habitat). L'abondance totale des amphibiens dépend d'abord de la surface inventoriée.



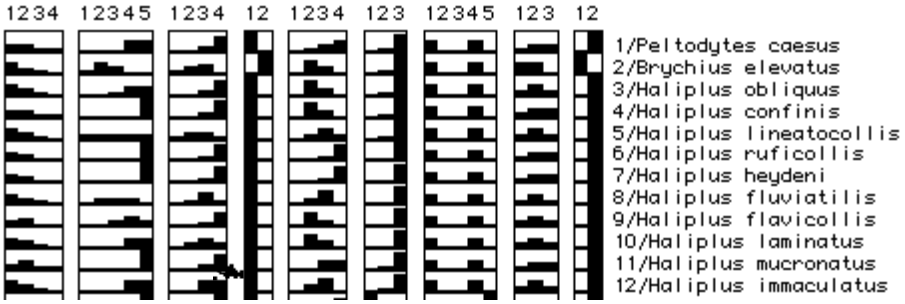
Ce n'est pas une découverte mais le seul élément que l'analyse statistique permet d'extraire des données.

1 Morand, A. & Joly, P. (1995) Habitat variability and space utilization by the amphibian communities of the French Upper-Rhone floodplain. *Hydrobiologia* : 300/301, 249-257.

MCA : Fuzzy Correspondence Analysis

Méthode d'analyse de données à un tableau.

L'option fait l'analyse des correspondances d'un tableau de variables floues[1], c'est-à-dire l'analyse des correspondances multiples étendue aux tableaux individus-modalités non disjonctifs complets[2] :



Soit **A** un tableau de variables floues avec t lignes et m colonnes (modalités). Soit v le nombre de variables. La variable j a $m(j)$ modalités.

$$m = \sum_{j=1}^{v} m(j)$$

Pour la $i^{\text{ème}}$ ligne et la $k^{\text{ème}}$ modalité de la variable j, on note la valeur a_{ij}^{k} ($1 \le i \le t$, $1 \le j \le v$, and $1 \le k \le m(j)$), la valeur du tableau des données brutes. Soit :

$$a_{ij}^{\bullet} = \sum_{k=1}^{m(j)} a_{ij}^{k} \text{ and } p_{ij}^{k} = \frac{a_{ij}^{k}}{a_{ij}^{\bullet}}$$

Soit le tableau $\mathbf{P} = \left[p_{ij}^{k} \right]$ et les paramètres :

$$p_{ij}^{\bullet} = \sum_{k=1}^{m(j)} p_{ij}^{k} = 1,\ p_{i\bullet}^{\bullet} = \sum_{j=1}^{v} p_{ij}^{\bullet} = v, \text{ and } p_{\bullet\bullet}^{\bullet} = \sum_{i=1}^{t} p_{i\bullet}^{\bullet} = tv$$

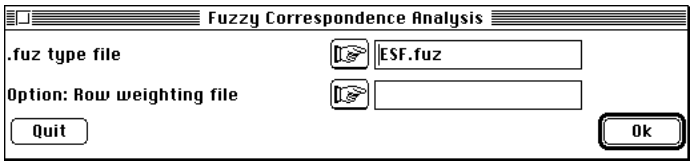
Les poids des lignes du tableau **P** sont uniformément égaux à $r_i = \frac{v}{tv} = \frac{1}{t}$ et les poids des colonnes du tableau **P** sont égaux à :

$$c_j^k = \frac{\sum_{i=1}^{t} p_{ij}^{k}}{tv} = \frac{1}{v}\left(\frac{1}{t}\sum_{i=1}^{t} p_{ij}^{k}\right) = \frac{1}{v}\overline{p_j^k}$$

Si une variable n'est pas renseignée pour un individu, elle est codée "0, 0, 0, 0" et remplacée par le profil moyen de toutes les espèces et positionnée automatiquement à l'origine par le centrage. Le tableau **P** est traité par une analyse des correspondances, c'est à dire l'analyse du schéma de dualité[3] formé du tableau $\left[p_{ij}^{k} / \overline{p_j^k} - 1 \right]$, des poids des colonnes c_j^k et des poids des lignes r_i.

L'option s'emploie obligatoirement après l'utilitaire de lecture des données FuzzyVar : Read Fuzzy File.

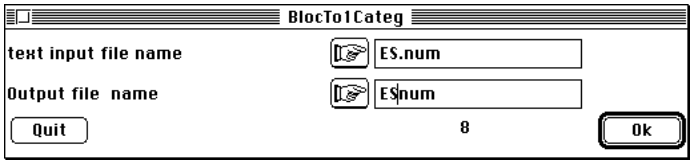
L'option utilise une seule fenêtre de dialogue :



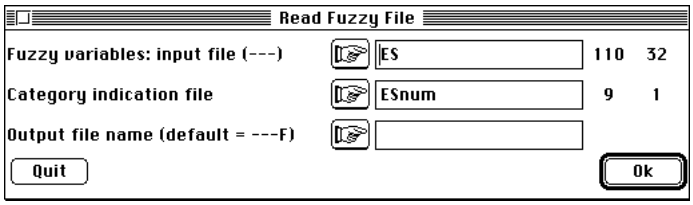
Nom du fichier de type ---.fuz créé par FuzzyVar : Read Fuzzy File.

Une pondération non uniforme est théoriquement possible si on donne ici un nom de fichier des poids.

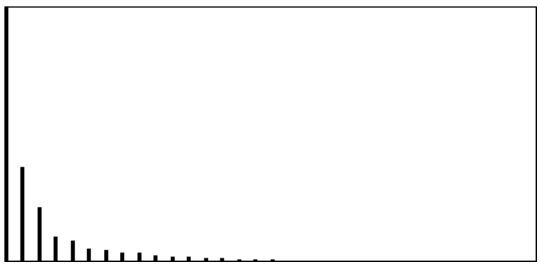
Utiliser la carte Coléoptères de la pile ADE-4•Data pour créer le fichier ES (110-32) par TextToBin : Char->Binary. Récupérer le fichier ES.num et le lire par TextToBin : BlocTo1Categ :



Lire le fichier ES avec FuzzyVar : Read Fuzzy File :



Utiliser la présente option :



```
fl/FuzzyMCA: Multiple correspondence analysis on fuzzy table
Input file: ESF.fuz for access to file ESF
Row number: 110, column number: 32
Uniform row weights

File ESF.flpl contains the row weights
It has 110 rows and 1 column

File ESF.flta contains the tabled processed by MCA
It has 110 rows and 32 columns (categories)

File ESF.flma contains
----- number of rows: 110
----- number of variables: 9
----- number of categories: 32
----- variable number of each category (vector of 32 values)

File ESF.flpc contains the column weights (1/V)*DM
It has 32 rows and 1 column
```

Le listing donne d'abord les poids des modalités (inchangés par rapport à l'édition par FuzzyVar : Read Fuzzy File si on a utilisé une pondération imposée uniforme, ce qui est le cas ici) :

```
Marginal distributions by variable:
----------------------------------------
Variable number 1 has 4 categories
----------------------------------------
[1]  Category:  1    Weight:    0.468
[2]  Category:  2    Weight:    0.261
[3]  Category:  3    Weight:    0.152
[4]  Category:  4    Weight:    0.119

Variable number 2 has 5 categories
•••
Variable number 9 has 2 categories
----------------------------------------
[31] Category:  1    Weight:    0.318
[32] Category:  2    Weight:    0.682
----------------------------------------
```

On a ensuite les résultats généraux de l'analyse d'un triplet :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: ESF.flta
--- Number of rows: 110, columns: 32
----------------------
Total inertia:  1.01509
----------------------
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +4.9546E-01 +0.4881 +0.4881   |02   +1.8411E-01 +0.1814 +0.6695 |
03   +1.0530E-01 +0.1037 +0.7732   |04   +4.8528E-02 +0.0478 +0.8210 |
05   +4.1277E-02 +0.0407 +0.8617   |06   +2.6060E-02 +0.0257 +0.8873 |
•••
19   +1.7797E-03 +0.0018 +0.9978   |20   +1.1714E-03 +0.0012 +0.9989 |
21   +7.5992E-04 +0.0007 +0.9997   |22   +2.7709E-04 +0.0003 +0.9999 |
23   +6.4894E-05 +0.0001 +1.0000   |24   +0.0000E+00 +0.0000 +1.0000 |
25   +0.0000E+00 +0.0000 +1.0000   |26   +0.0000E+00 +0.0000 +1.0000 |
27   +0.0000E+00 +0.0000 +1.0000   |28   +0.0000E+00 +0.0000 +1.0000 |
29   +0.0000E+00 +0.0000 +1.0000   |30   +0.0000E+00 +0.0000 +1.0000 |
31   +0.0000E+00 +0.0000 +1.0000   |32   +0.0000E+00 +0.0000 +1.0000 |
```

Observer que le nombre de valeurs propres nulles est égal au nombre de variables.

```
File ESF.flvp contains the eigenvalues and relative inertia for each axis
--- It has 32 rows and 2 columns

File ESF.flco contains the column scores
--- It has 32 rows and 2 columns
File :ESF.flco
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.853e+00| 6.759e-01|
|   2|-1.131e+00| 9.533e-01|
|----|----------|----------|

File ESF.flli contains the row scores
--- It has 110 rows and 2 columns
File :ESF.flli
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.442e+00| 6.127e-01|
|   2|-6.132e-01| 7.960e-01|
|----|----------|----------|
```

On a enfin une édition d'aides à l'interprétation spécifiques, reliée aux propriétés théoriques de cette analyse qui généralise l'analyse des correspondances multiples. Pour chaque variable, chaque modalités et chaque facteur, la variance des scores des modalités rapportée à la variance des scores des individus (rapport de corrélation) indique dans quelle mesure un axe prend en compte une variable :

```
CorRatioFCA: Correlation ratios after a FCA
Title of the analysis: ESF.fl
Number of rows: 110, columns: 9

Variable : 1
```

```
> Categ=   1 Weight=  0.468  -0.248  -0.108
> Categ=   2 Weight=  0.261   0.265   0.184
> Categ=   3 Weight=  0.152  -0.045  -0.061
> Categ=   4 Weight=  0.119   0.450   0.097
----------------------> r=    0.072   0.016
```

7 % de la variance entre position des points s'exprime dans la variance entre position des modalités positionnées par averaging sur la distribution associée : la variable n'est pas prise en compte sur le premier facteur.

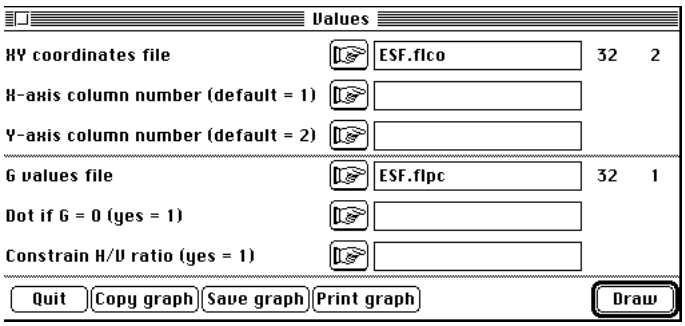
```
Variable : 2
> Categ=   1 Weight=  0.069  -1.853  -0.300
> Categ=   2 Weight=  0.159  -1.301   0.264
> Categ=   3 Weight=  0.091  -0.654   0.514
> Categ=   4 Weight=  0.094  -0.024   0.057
> Categ=   5 Weight=  0.587   0.676  -0.124
----------------------> r=    0.814   0.051
```

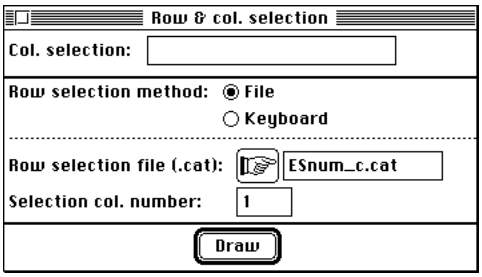
81 % de la variance entre position des points s'exprime dans la variance entre position des modalités positionnées par averaging sur la distribution associée : la variable est fortement prise en compte sur le premier facteur.

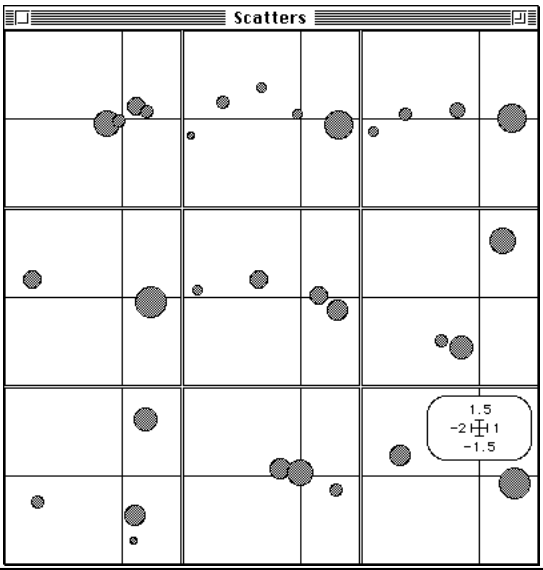
...

```
Variable : 9
> Categ=   1 Weight=  0.318  -1.320   0.334
> Categ=   2 Weight=  0.682   0.616  -0.156
----------------------> r=    0.813   0.052

File ESF.flrc contains the correlation ratios between
the categorical variables and the factor scores
It has 9 rows and 2 columns
```

⚠ La carte des modalités, comme en ACM, a des propriétés particulières. Utiliser dans Scatters le multifenêtrage par la répartition des modalités entre les variables. Les cartes sont centrées par variables pour la pondération marginale.

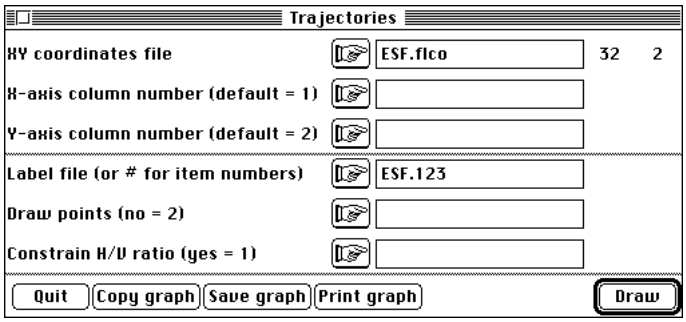


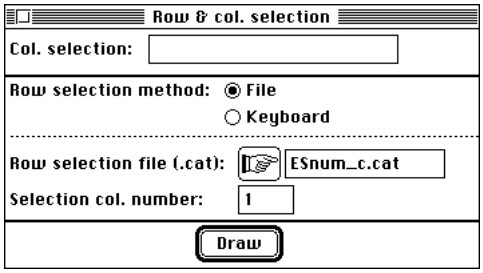


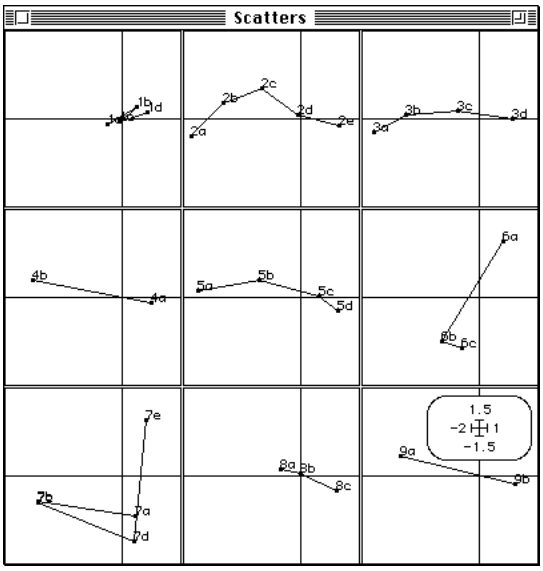


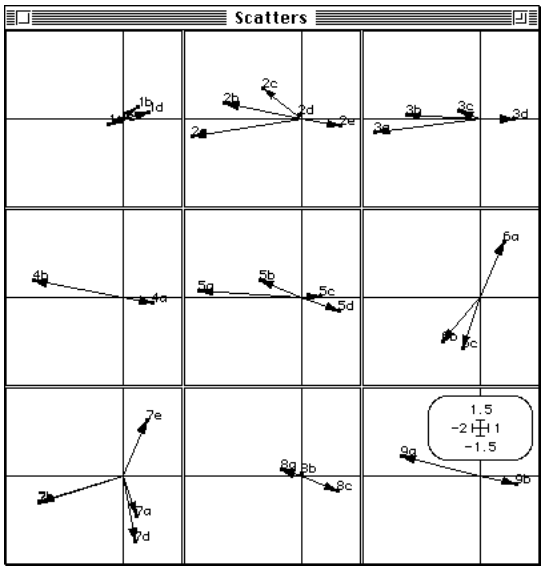
Noter la cohérence entre l'aide à l'interprétation numérique (en haut, à gauche, 7 % et 2 % de variance expliquée, à côté 81 % et 5 %) et les représentations graphiques ((en haut, à gauche, modalités non séparées, à côté modalités alignées sur l'axe 1) .

Trajectoires ou représentations vectorielles sont possibles ou utiles, selon les cas :

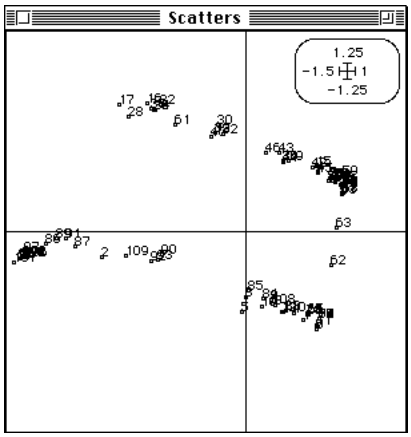


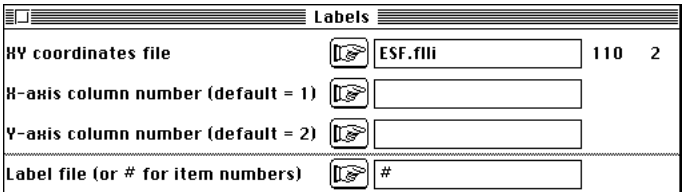






Les cartes lignes et colonnes sont liées, comme dans toute AFC, par des relations de double averaging déformé, qui supportent l'interprétation (les variables exprimées sur l'axe 1 ordonnent, les variables exprimées sur l'axe 2 séparent) :





L'analyse de cette option est une généralisation de l'ACM. Elle autorise une pondération des lignes quelconque. Appliquée à un tableau disjonctif complet avec la pondération uniforme, c'est strictement une ACM (MCA : Multiple Correspondence Analysis).

Appliquée à un tableau d'AFC décomposé par bloc de colonnes avec sa pondération marginale c'est strictement une AFC intra-classes (blocs de colonnes, COA : Internal COA).

On peut passer, pour ce type de données, dans la partie K-tableaux par KTabUtil : FuzzyToKTab.

Après cette option, les fonctions de DDUtil sont disponibles.

Cette analyse a été dupliquée sous le terme RCT-MCA en génétique (tableaux de fréquences alléliques) dans [4].

Les modules graphiques associées à cette analyse dans ADE 3.7 sont en cours de transfert.

[1] Bournaud, M., Richoux, P. & Usseglio-Polatera, P. (1992) An approach to the synthesis of qualitative ecological information from aquatic Coleoptera communities. Regulated rivers: *Research and Management* : 7, 165-180.

[2] Chevenet, F., Dolédec, S. & Chessel, D. (1994) A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* : 31, 295-309.

[3] Escoufier, Y. (1987) The duality diagramm : a means of better practical applications. In : *Development in numerical ecology*. Legendre, P. & Legendre, L. (Eds.) NATO advanced Institute , Serie G .Springer Verlag, Berlin. 139-156.

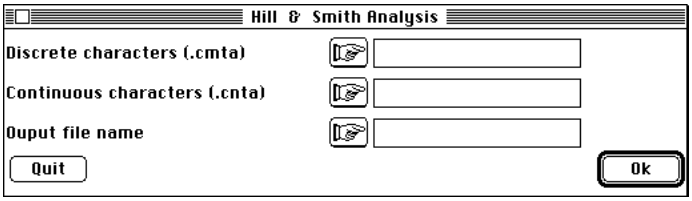
[4] Guinand, B. (1996) Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biological Journal of the Linnean Society* : 58, 173-195.

MCA : Hill & Smith Analysis

Méthode d'analyse multivariée due à Hill & Smith (1976)[1] qui permet de mélanger variables qualitatives et quantitatives.

Le codage des variables, c'est-à-dire le mode d'enregistrement de l'information est un problème permanent de l'analyse des données écologiques. C'est sans doute pour cela que l'ouvrage de Jongman & Coll. [2] s'ouvre sur un problème de ce type. Les variables sont quantitatives et conduisent à une ACP (PCA : Correlation matrix PCA), qualitatives et demandent une ACM (MCA : Multiple Correspondence Analysis) ou floues et supportent une ACF (MCA : Fuzzy Correspondence Analysis). Quand plusieurs types sont ensemble, on peut chercher soit à transformer certaines variables pour les amener dans le type dominant soit à mélanger les types. Une solution générale d'analyse des mélanges est l'analyse canonique généralisée [3] qui sera prochainement disponible dans ADE-4. L'analyse de Hill & Smith est un cas particulier qui autorise le mélange d'une ACP normée et d'une ACM. Simple et efficace, cette méthode est une excellente introduction à la stratégie de l'analyse canonique généralisée.

L'option utilise une seule fenêtre de dialogue :



Nom du fichier d'entrée d'une analyse des correspondances multiples à *n* lignes, *v* variables et *m* modalités.

Nom du fichier d'entrée d'une analyse en composantes principales normées à *n* lignes et *p variables*. Les deux analyses portent sur les mêmes individus-lignes et utilisent la même pondération des lignes.

Nom générique des fichiers de sortie (création).

Utiliser la carte Dune de la pile ADE-4•Data. Envoyer le tableau de milieu dans un tableur :

	1	2	3	4	5	6	7	8
1	2.8	1	2	4	1	0	0	0
2	3.5	1	2	2	0	1	0	0
3	4.3	2	2	4	1	0	0	0
4	4.2	2	2	4	1	0	0	0
5	6.3	1	1	2	0	0	1	0
6	4.3	1	2	2	0	0	1	0
7	2.8	1	3	3	0	0	1	0
8	4.2	5	3	3	0	0	1	0
9	3.7	4	1	1	0	0	1	0
10	3.3	2	1	1	0	1	0	0
11	3.5	1	3	1	0	1	0	0
12	5.8	4	2	2	1	0	0	0
13	6	5	2	3	1	0	0	0
14	9.3	5	3	0	0	0	0	1
15	11.5	5	2	0	0	0	0	1
16	5.7	5	3	3	1	0	0	0
17	4	2	1	0	0	0	0	1
18	4.6	1	1	0	0	0	0	1
19	3.7	5	1	0	0	0	0	1
20	3.5	5	1	0	0	0	0	1

En lignes, on a 20 sites (description de l'ouvrage cité p. XVII). La première colonne est l'épaisseur de l'horizon 1 en cm. C'est une variable quantitative. La colonne 2 est l'humidité du sol exprimée en 4 classes ordonnées. La variable est dite ordinale et rangée dans la classe quantitative. La colonne 3 est le type d'utilisation de la prairie (1-prairie de fauche, 2-utilisation mixte, 3-pâturage). On peut considérer qu'il s'agit d'une variable qualitative (dans l'ouvrage elle est traitée comme une variable ordinale donc quantitative). La colonne 4 est le type d'apport d'engrais codé en niveaux ordonnés de 0 à 4. La variable est quantitative. Les colonnes 5 à 8 sont des indicatrices des classes qui décrivent une variable qualitative (type d'aménagement) à 4 modalités (1- agriculture traditionnelle, 2- agriculture biologique, 3 - culture de loisir et 4 - conservation de la nature).

Extraire les trois variables quantitatives dans un fichier Quan.txt et les deux variables qualitatives dans un fichier Qual.Txt après recodage à partir des indicatrices :

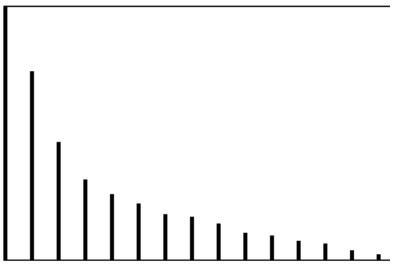
Quan.txt

	1	2	3
1	2.8	1	4
2	3.5	1	2
3	4.3	2	4
4	4.2	2	4
5	6.3	1	2
6	4.3	1	2
7	2.8	1	3
8	4.2	5	3
9	3.7	4	1
10	3.3	2	1
11	3.5	1	1
12	5.8	4	2
13	6	5	3
14	9.3	5	0
15	11.5	5	0
16	5.7	5	3
17	4	2	0
18	4.6	1	0
19	3.7	5	0
20	3.5	5	0

Qual.txt

	1	2
1	2	1
2	2	2
3	2	1
4	2	1
5	1	3
6	2	3
7	3	3
8	3	3
9	1	3
10	1	2
11	3	2
12	2	1
13	2	1
14	3	4
15	2	4
16	3	1
17	1	4
18	1	4
19	1	4
20	1	4

Passer ces deux fichiers en binaire (TextToBin : Text->Binary) dans Quan (20-3) et Qual (20-2). Ces données de milieu sont destinée à expliquer la répartition des espèces du tableau floristique DuneVeg disponible sur la même carte (op. cit. p. VIII). DuneVeg a 30 lignes-espèces et 20 colonnes-sites. Le transposer (FilesUtil : Transpose) en Flore (20-30) pour avoir les relevés en lignes. Faire son analyse des correspondances (COA : COrrespondence Analysis) :



COrrespondence Analysis

Data file	Flore	20	30	Number of axes ?	3

Entreprendre alors l'ACP normée (PCA : Correlation matrix PCA) du tableau Quan avec la pondération de l'AFC (stratégie d'analyse canonique des correspondances [4], voir CCA : Initialize explanatory variables) :

Correlation matrix PCA

Matrix input file	Quan	20	3
Row weights (default=1/n)	3		
Column weights (default=1)			
Option: file for row weighting	Flore.fcpl	20	1
Option: file for column weighting			
1 = Save correlation matrix	1		

Lire le fichier Qual (CategVar : Read Categ File) :

Read Categ File

Input file	Qual	20	2

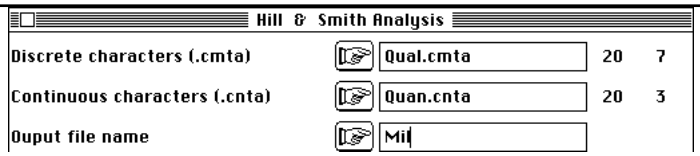
Entreprendre alors l'analyse des correspondances multiples pondérées (MCA : Multiple Correspondence Analysis) avec la même pondération :

Multiple Correspondence Analysis

.cat type file	Qual.cat		
Option: Row weight file	Flore.fcpl	20	1

Dans ces deux analyses préliminaires le nombre d'axes conservés n'a pas grande signification et peut être quelconque.

Réunir les deux analyses par la présente option :




```
Hill, M.O. & Smith, A.J.E. (1976)
Principal component analysis of taxonomic data with multi-state discrete characters
Taxon : 25, 249-255

First input file (Multiple correspondence analysis): Qual.cmta
Second input file (Normalized principal component analysis): Quan.cnta
Output table (Mixed): Mil.hita
File Mil.hita has 20 rows and 10 columns (7 categories + 3 variables)
```

On constitue un nouveau triplet statistique en accolant purement et simplement les tableaux des deux analyses de base. Le programme contrôle que la compatibilité est assurée en vérifiant que les pondérations des analyses de départ sont identiques. Cette pondération commune des lignes des tableaux de départ est conservée comme pondération des lignes du nouveau tableau :

```
File Mil.hipl contains the row weights
It has 20 rows and 1 column
```

Les pondération des colonnes sont par contre modifiées. Dans l'ACM chaque colonne est une modalité. On somme les poids des porteurs de cette modalité et on divise par le nombre de variables. Ceci attribue des poids aux modalités tels que chaque variable totalise (par ses modalités) un poids égal à un sur le nombre de variables. On fait ici la même chose mais en divisant par le nombre de variables total (qualitatives et quantitatives) et au attribue aux variables quantitatives le poids uniforme de un sur le nombre total de variables. Ici, il y a 2 variables qualitatives et 3 variables quantitatives. Chacune des dernières a un poids de 1/5 et les poids des modalités des premières sommés par variable redonne 1/5. La somme totale des poids des colonnes du tableau vaut ainsi 1.

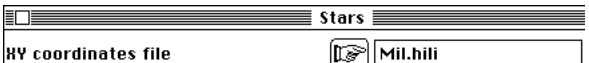
```
File Mil.hipc contains the column weights (V/nvartot)*DM and (1/nvartot)Ip
It has 10 rows  (7 categories + 3 variables) and 1 column
```

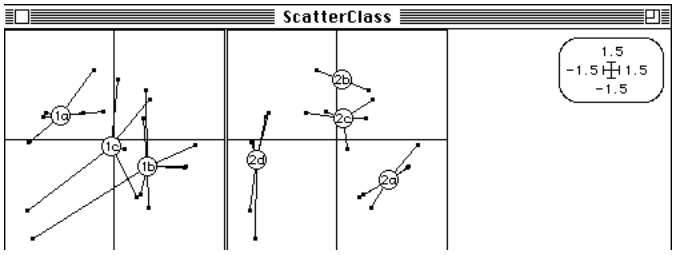
On obtient ainsi un nouveau triplet statistique qui donne une analyse d'inertie standard. L'inertie de l'ACM initiale (nombre de modalités/nombre de variables -1) vaut 7/2 -1, soit 2.5. L'inertie de l'ACP initiale vaut (nombre de variables) 3. L'inertie de l'analyse conjointe vaut 2.5(2/5) +3/5 = 1.6. Comme en ACP normée ou en ACM cette valeur ne dépend pas des observations mais de la structure des variables.

```
DiagoRC: General program for two diagonal inner product analysis
Input file: Mil.hita
--- Number of rows: 20, columns: 10
-----------------------
Total inertia:      1.6
-----------------------
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +4.9856E-01 +0.3116 +0.3116   |02   +3.7237E-01 +0.2327 +0.5443 |
03   +2.4724E-01 +0.1545 +0.6989   |04   +1.8705E-01 +0.1169 +0.8158 |
05   +1.4411E-01 +0.0901 +0.9058   |06   +8.1055E-02 +0.0507 +0.9565 |
07   +5.5711E-02 +0.0348 +0.9913   |08   +1.3909E-02 +0.0087 +1.0000 |
09   +0.0000E+00 +0.0000 +1.0000   |10   +0.0000E+00 +0.0000 +1.0000 |

File Mil.hivp contains the eigenvalues and relative inertia for each axis
--- It has 10 rows and 2 columns

File Mil.hico contains the column scores
--- It has 10 rows and 2 columns
File :Mil.hico
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.552e+00| 9.846e-01|
|   2|-9.149e-01| 1.328e+00|
|----|----------|----------|
```

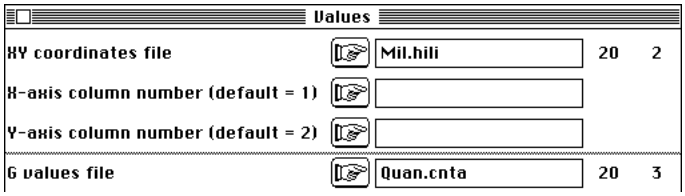
Les coordonnées des lignes sont centrées, de variance égales aux valeurs propres. On s'en sert comme dans une ACM ou une ACP :

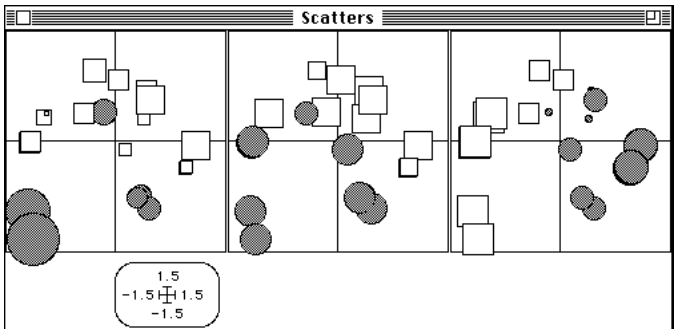
```
File Mil.hili contains the row scores
--- It has 20 rows and 2 columns
File :Mil.hili
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.182e+00| 1.114e+00|
|   2|-1.347e+00| 9.427e-01|
|----|----------|----------|
```





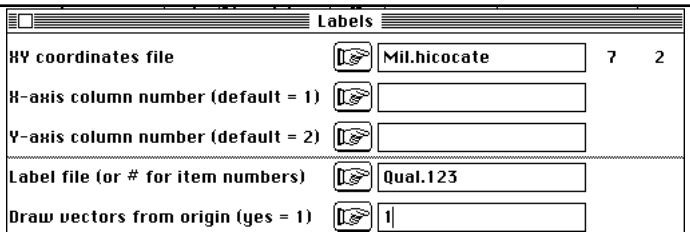
La variable usage est plutôt interprétée comme ordination sur F1 (1-fauche, 2-pâturage, 3-les deux) et la variable aménagement comme partition en trois groupes.

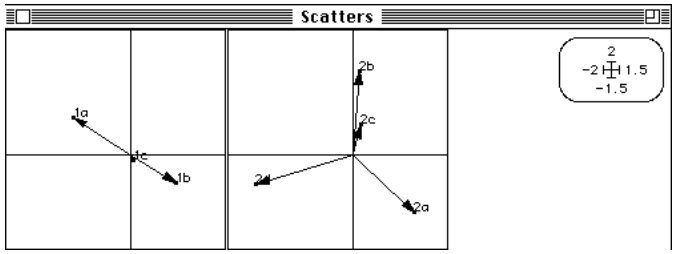




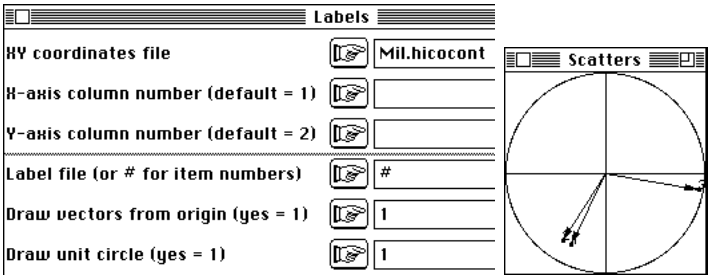
Les coordonnées des colonnes sont séparées en deux groupes. On s'en sert comme dans une ACM ou une ACP.

```
File Mil.hicocate contains the column scores of categories (discrete parameters)
--- It has 7 rows and 2 columns
--- It contains 7 first lines from file Mil.hico
--- It is to be used with Qual.cat and associated files
File :Mil.hicocate
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.552e+00| 9.846e-01|
|   2|-9.149e-01| 1.328e+00|
|----|----------|----------|
```






```
File Mil.hicocont contains the column scores of continuous parameters
--- It has 3 rows and 2 columns
--- It contains 3 last lines from file Mil.hico
--- It can be used for drawing correlation circles
File :Mil.hicocont
|Col.|   Mini    |   Maxi    |
|----|-----------|-----------|
|   1|-4.239e-01| 9.375e-01|
|   2|-7.126e-01|-1.683e-01|
|----|-----------|-----------|
```



La théorie des cosinus carrés des angles donne à cette analyse des propriétés numériques extrêmement efficaces. La coordonnée des lignes (axe 1) en ACP normée maximise la somme des carrés des corrélations entre les variables et un score (propriété rappelée dans [1] et illustrée dans [5]). La coordonnée des lignes (axe 1) en ACM maximise la moyenne des rapport de corrélation entre les variables qualitatives et un score (propriété due à [6] exprimée en terme d'analyse de variance dans [1] et illustrée dans [7]). La présente analyse réunit les deux propriétés et maximise la moyenne sur l'ensemble des variables des R^2 entre variables et un score, ce R^2 étant un carré de corrélation pour les quantitatives et un rapport de corrélation pour les qualitatives. Cette moyenne sur l'ensemble des descripteurs vaut la valeur propre.

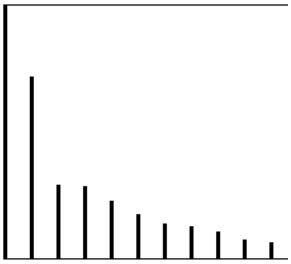
Le listing donne ces quantités pour chaque axe et on y retrouve la relation des deux ensembles de variables sur l'axe 1 et la seule prise en compte d'un sous-ensemble de quantitatives sur l'axe 2, ce qui vient d'être exprimée sur les graphiques précédents.

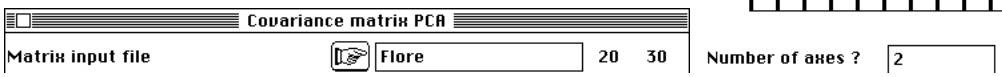
```
R2 (x1000) Column = axes
First bloc: discrete parameters
Second bloc: continuous parameters
Third bloc: overall mean = eigenvalue/2
Variable :   1          486         203
Variable :   2          833         666

Variable :   1          115         508
Variable :   2          180         457
Variable :   3          879          28
```

```
Overall mean            499        372
File Mil.hiR2 contains the R2 coefficients
--- It has 5 rows and 2 columns
```

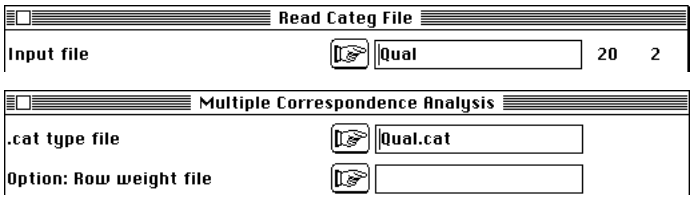
⚠ Les options ACP normée pondérée et ACM pondérée ont conduit à une version de l'analyse de Hill et Smith pondérée qui est une extension de l'article d'origine. La pondération uniforme semble cependant ici préférable, car on disposera de plusieurs tests de signification non disponibles pour des pondérations quelconques.

Les données recodées sont directement disponibles sur la carte Dune+1 de la pile ADE-4•Data. Faire l'ACP centrée du tableau floristique :



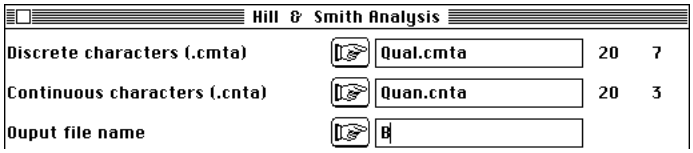


Lire les variables qualitatives et faire l'ACM :



Faire l'ACP normée des variables quantitatives :

Associer les deux tableaux :



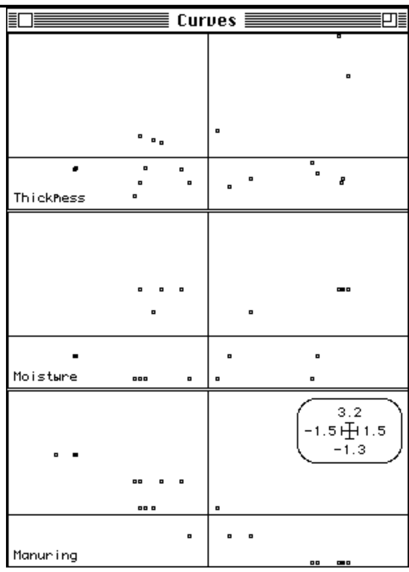

```
Hill, M.O. & Smith, A.J.E. (1976)
Principal component analysis of taxonomic data with multi-state discrete
characters
Taxon : 25, 249-255

First Input (Multiple correspondence analysis): Qual.cmta
Second Input (Normed principal component analysis): Quan.cnta
Output table (Mixed): B.hita
File B.hita has 20 rows and 10 columns (7 categories + 3 variables)

File B.hipl contains the row weights
It has 20 rows and 1 column

File B.hipc contains the column weights (V/nvartot)*DM and (1/nvartot)Ip
It has 10 rows  (7 categories + 3 variables) and 1 column

DiagoRC: General program for two diagonal inner product analysis
Input file: B.hita
--- Number of rows: 20, columns: 10
-----------------------
Total inertia:      1.6
-----------------------
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
```

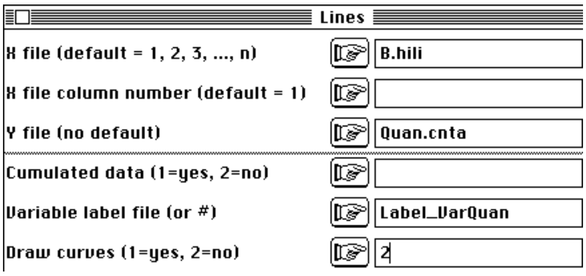
```
01   +5.0842E-01 +0.3178 +0.3178   |02   +3.7156E-01 +0.2322 +0.5500 |
03   +2.4612E-01 +0.1538 +0.7038   |04   +1.9799E-01 +0.1237 +0.8276 |
05   +1.3853E-01 +0.0866 +0.9141   |06   +8.2159E-02 +0.0513 +0.9655 |
07   +4.3773E-02 +0.0274 +0.9928   |08   +1.1451E-02 +0.0072 +1.0000 |
09   +0.0000E+00 +0.0000 +1.0000   |10   +0.0000E+00 +0.0000 +1.0000 |

File B.hivp contains the eigenvalues and relative inertia for each axis
--- It has 10 rows and 2 columns

File B.hico contains the column scores
--- It has 10 rows and 2 columns
File :B.hico
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.022e+00| 1.324e+00|
|   2|-8.492e-01| 1.338e+00|
|----|----------|----------|

File B.hili contains the row scores
--- It has 20 rows and 2 columns
File :B.hili
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.135e+00| 1.056e+00|
|   2|-1.211e+00| 9.691e-01|
|----|----------|----------|

File B.hicocate contains the column scores of categories (discrete
parameters)
--- It has 7 rows and 2 columns
--- It contains 7 first lines from file B.hico
--- It is to be used with Qual.cat and associated files
File :B.hicocate
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.022e+00| 1.324e+00|
|   2|-8.492e-01| 1.338e+00|
|----|----------|----------|

File B.hicocont contains the column scores of continuous parameters
--- It has 3 rows and 2 columns
--- It contains 3 last lines from file B.hico
--- It can be used for drawing correlation circles
File :B.hicocont
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-9.466e-01| 4.312e-01|
|   2|-7.168e-01|-1.604e-01|
|----|----------|----------|
```

Noter l'association aménagement (variable qualitative 2) et engrais (variable quantitative 3) donnant respectivement un R2 de 0.86 et 0.90 sur l'axe 1 :

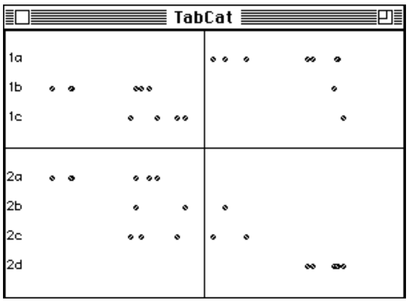
```
R2 (x1000) Column = axes
First bloc: discrete parameters
Second bloc: continuous parameters
Third bloc: overall mean = eigenvalue/2
Variable :   1          445        276
Variable :   2          860        586

Variable :   1          155        514
Variable :   2          186        456
Variable :   3          896         26

Overall mean            508        372
File B.hiR2 contains the R2 coefficients
--- It has 5 rows and 2 columns
```

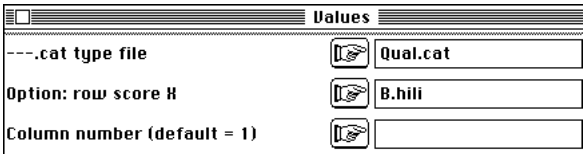
Exprimer directement l'optimisation sous-jacente à l'analyse en utilisant les deux modules graphiques adaptés Curves et Tabcat. Pour le premier :



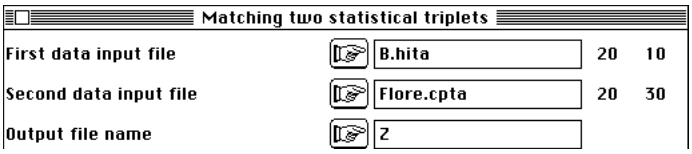


Pour le second :

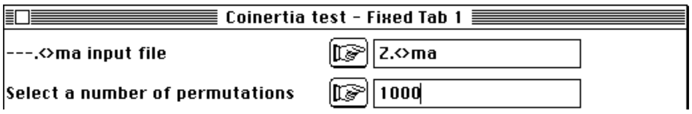


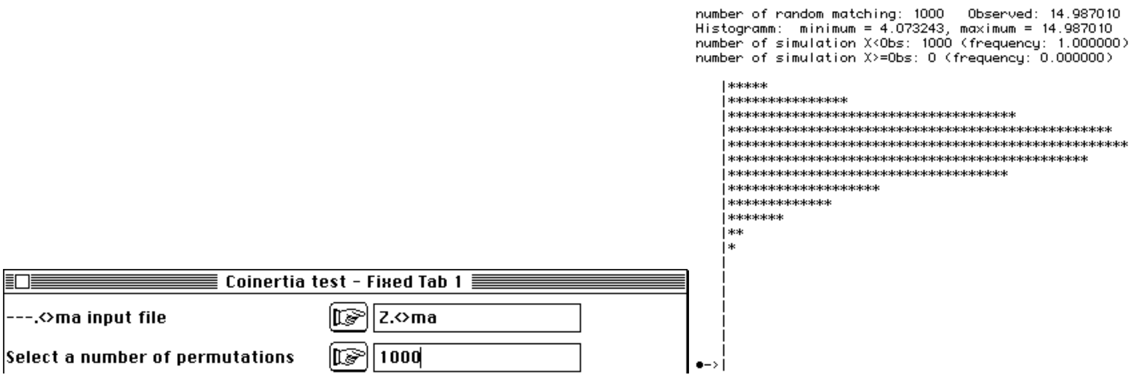


Coupler enfin cette analyse mixte avec l'analyse du tableau floristique :

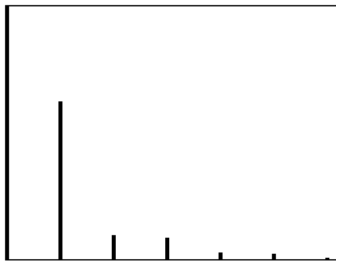


Vérifier la forte signification de la co-structure :



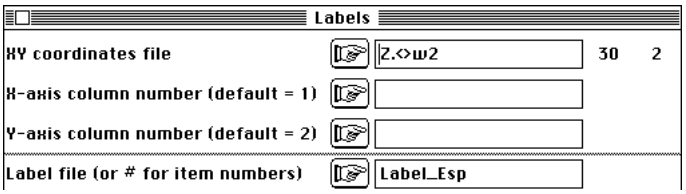


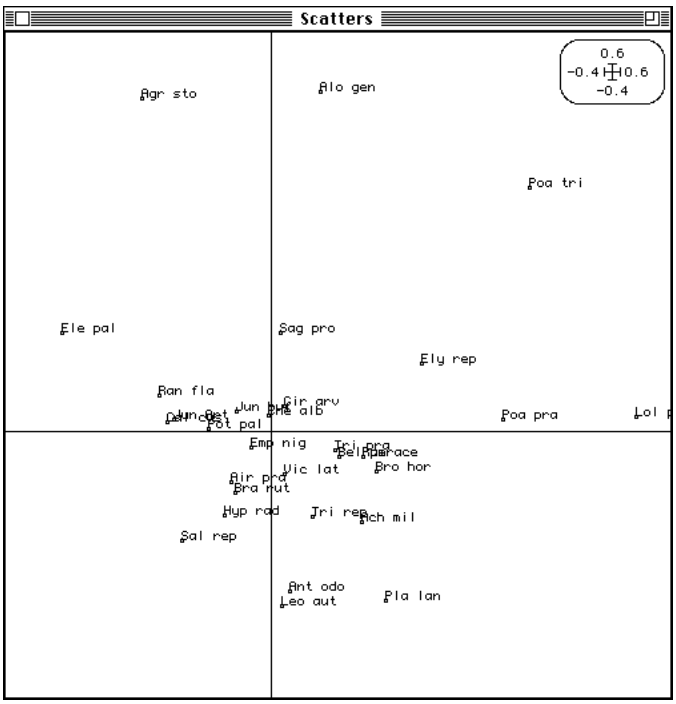
Exécuter l'analyse :



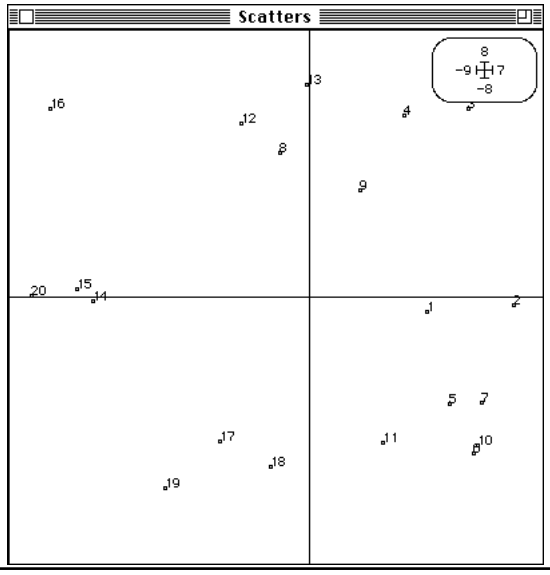
Number of axes ? 2

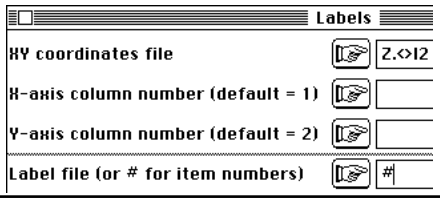
Pour interpréter, utiliser le code des taxons (scores de somme de carrés égaux à 1) :





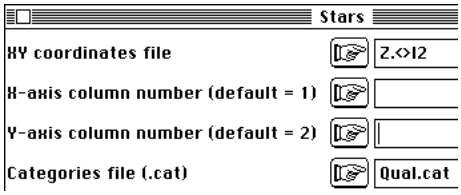
Les espèces positionnent les relevés par combinaisons linéaires :

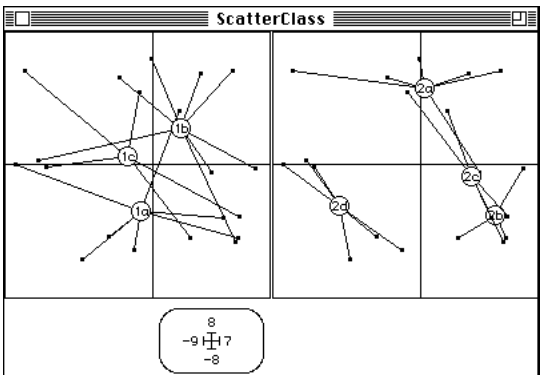




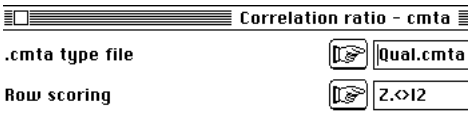
La position des relevés ainsi définie optimise ses relations avec les variables, donc

- partiellement avec les variables qualitatives, ce qui s'exprime soit graphiquement :



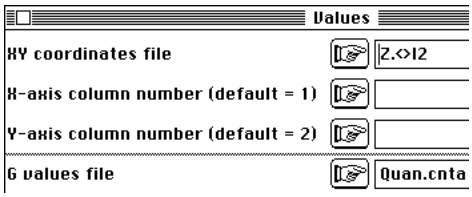


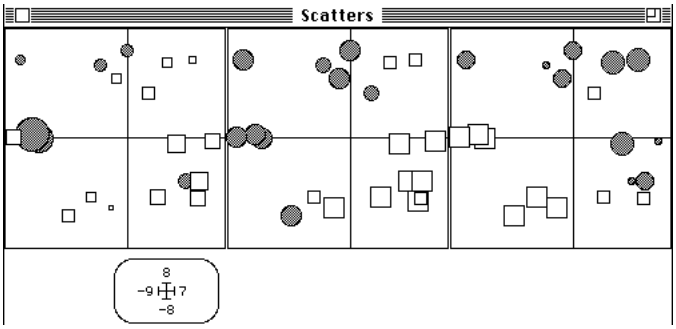
soit numériquement :



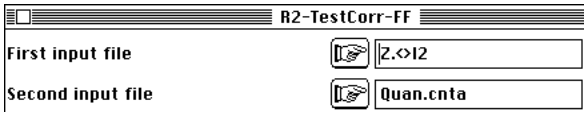

```
Caregorical variables: file Qual.cmta
Number of rows: 20, variables: 2, categories: 7

Row scoring: file Z.<>l2
Number of rows: 20, columns: 2

Variable :   1--> r=    0.087   0.270
Variable :   2--> r=    0.585   0.564

Mean     :    --> r=    0.336   0.417
```

- partiellement avec les variables quantitatives, ce qui s'exprime soit graphiquement :





soit numériquement :



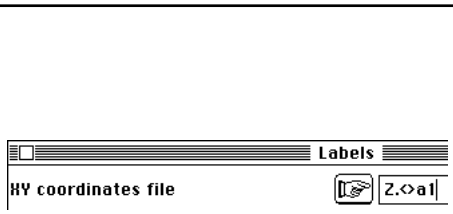

```
Note : Two sides test t for correlation
Proba <X>|Obs| or X<-|Obs|)x10000 - R2x1000
i = number of variable from file Quan.cnta
j = number of variable from file Z.<>l2

[i =   1][j =   1] R =    -0.5259 | Proba =   166 | R2 =  277
[i =   1][j =   2] R =    -0.7939 | Proba =     0 | R2 =  630
[i =   1][j =   3] R =     0.517 | Proba =   188 | R2 =  267

[i =   2][j =   1] R =     0.1959 | Proba =  4123 | R2 =   38
[i =   2][j =   2] R =     0.4768 | Proba =   320 | R2 =  227
[i =   2][j =   3] R =     0.5816 | Proba =    70 | R2 =  338
```

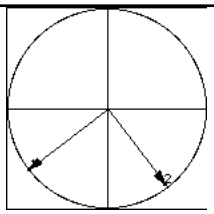
On prévoit que le critère optimisé par l'analyse (premier facteur) est le produit de la variance du code relevé (première colonne de Z.<>l2) soit 21.69 (un peu moins que l'optimum de 23.58):

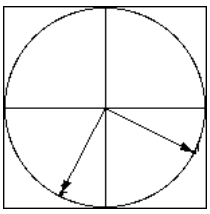
Num	Covaria.	Varian1	varian2	Correla.	INER1	INER2
1	2.831	0.4437	21.69	0.9124	0.5084	23.58
2	2.236	0.4056	17.39	0.842	0.3716	16.94

par la moyenne des rapports de corrélation (0.087, 0.585) et des carrés de corrélation (0.277, 0.630, 0.267) soit 0.369 (franchement moins que l'optimum 0.508). On retrouve la valeur 0.369*21.69 = 8.01 comme première valeur propre de l'analyse de co-inertie :

```
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +8.0123E+00 +0.5346 +0.5346   |02   +4.9997E+00 +0.3336 +0.8682 |
03   +7.8546E-01 +0.0524 +0.9206   |04   +6.8986E-01 +0.0460 +0.9667 |
•••
```







Noter que dans toutes les analyses, il faut manipuler les plans 1-2 car les structures sont à deux dimensions sans ambiguïté.

Après cette option, les fonctions de DDUtil sont disponibles.

La forte cohérence mathématique du logiciel fait que l'introduction d'une variante (ici le mélange quantitatif-qualitatif) augmente en conséquence les analyses possibles. Les modules d'inter-intra, d'analyse discriminante, de co-inertie et de variables instrumentales et les modules de K-tableaux acceptent cette nouvelle option. L'utilisateur peut ainsi faire des analyses auxquelles personne jusqu'à présent n'a pensé. C'est évidemment sous sa responsabilité.

[1] Hill, M.O. & Smith, A.J.E. (1976) Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon* : 25, 249-255.

[2] Jongman, R.H., ter Braak, C.J.F. & van Tongeren, O.F.R. (1987) *Data analysis in community and landscape ecology*. Pudoc, Wageningen. 1-298.

[3] Tenenhaus, M. (1984) L'analyse canonique généralisée de variables numériques, nominales ou ordinales par des méthodes de codage optimal. In : *Data Analysis and Informatics, III*. Diday, E. & Coll. (Eds.) Elsevier Science Publishers B.V., North-Holland. 71-84.

[4] Ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* : 67, 1167-1179.

[5] Carrel, G., Barthelemy, D., Auda, Y. & Chessel, D. (1986) Approche graphique de l'analyse en composantes principales normée : utilisation en hydrobiologie. *Acta Œcologica, Œcologia Generalis* : 7, 2, 189-203.

[6] Saporta, G. (1975) *Liaisons entre plusieurs ensembles de variables et codage de données qualitatives*. Thèse de 3° cycle, Université Pierre et Marie Curie, Paris VI. 1-102.

[7] Pialot, D., Chessel, D. & Auda, Y. (1984) Description de milieu et analyse factorielle des correspondances multiples. *Compte rendu hebdomadaire des séances de l'Académie des sciences*. Paris, D : 298, Série III, 11, 309-314.

MCA : Multiple Correspondence Analysis

Méthode d'analyse multivariée pour tableaux de variables qualitatives.

On peut considérer que l'ACM est l'équivalent de l'ACP normée pour variables qualitatives. L'article de Tenenhaus et Young (1985) fait définitivement le tour des approches théoriques nombreuses de cette méthode[1].

L'option utilise un fichier acceptable par le module CategVar : Read Categ File. Soit **A** un tableau de variables qualitatives avec *n* lignes et *v* colonnes (*v* est le nombre de variables). La variable *j* a *m*(*j*) modalités et *m* est le nombre de modalités total :

$$m = \sum_{j=1}^{v} m(j)$$

Soit **X** le tableau disjonctif complet associé (*n lignes* et *m* colonnes). Son terme général x_{ik} vaut 1 si l'individu *i* porte la modalité *k* et 0 sinon. L'option autorise une pondération *a priori* arbitraire des lignes. Soit r_i, le poids de la ligne *i* avec

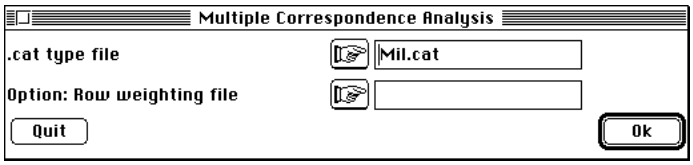
$$\sum_{i=1,n} r_i = 1.$$

Par défaut chaque ligne compte pour 1/*n*. Notons $i \quad P(k)$ le fait que l'individu *i* porte la modalité *k* La somme des poids des porteurs d'une modalité est :

$$q_k = \sum_{i \ P(k)} r_i$$

L'ACM de **A** est l'AFC pondérée de **X**, c'est à dire l'analyse du schéma de dualité défini par le tableau $\left[x_{ik} / \ q_k - 1\right]$, les poids des colonnes q_k / v et les poids des lignes r_i.

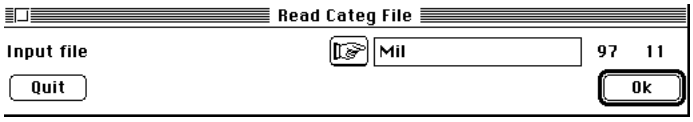
L'option utilise une seule fenêtre de dialogue :



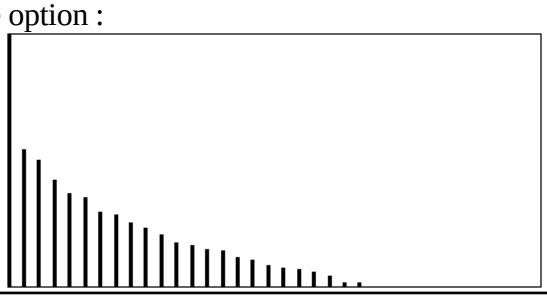
Nom de fichier du type ---.cat créé par CategVar : Read Categ File qui permet l'accès à un fichier de variables qualitatives.

Nom de fichier à une colonne de nombres positifs contenant les poids des lignes (par défaut, pondération uniforme).

Utiliser la carte Mafragh+2 de la pile ADE-4•Data pour obtenir le fichier Mil (97-11) par TextToBin : Char->Binary. Lire ce fichier par CategVar : Read Categ File :



Exécuter la présente option :



```
Uniform row weights

File Mil.cmpl contains the row weights
It has 97 rows and 1 column

File Mil.cmpc contains the column weights (1/V)*DM
It has 35 rows and 1 column
```

Le listing donne d'abord les poids des modalités (inchangés par rapport à l'édition par CategVar : Read Categ File si on a utilisé une pondération imposée uniforme, ce qui est le cas ici) :

```
Marginal distributions by variable:
----------------------------------------
Variable number 1 has 4 categories
----------------------------------------
[1]  Category:  1    Weight:    0.216
[2]  Category:  2    Weight:    0.186
[3]  Category:  3    Weight:    0.227
[4]  Category:  4    Weight:    0.371

•••

Variable number 11 has 5 categories
----------------------------------------
[31] Category:  1    Weight:    0.0515
[32] Category:  2    Weight:    0.371
[33] Category:  3    Weight:    0.247
[34] Category:  4    Weight:    0.155
[35] Category:  5    Weight:    0.175
----------------------------------------

File Mil.cmta contains the tabled processed by MCA
It has 97 rows and 35 columns (categories)
```

On a ensuite les résultats généraux de l'analyse d'un triplet :

```
DiagoRC: General program for two diagonal inner product analysis
Input file: Mil.cmta
--- Number of rows: 97, columns: 35
-----------------------
Total inertia:  2.18182
-----------------------
```

Observer que l'inertie totale est égale au quotient du nombre de modalité (35) par le nombre de variables (11) diminué de 1.

```
Num. Eigenval.   R.Iner.  R.Sum    |Num. Eigenval.   R.Iner.  R.Sum  |
01   +3.8159E-01 +0.1749 +0.1749   |02   +2.0715E-01 +0.0949 +0.2698 |
03   +1.9180E-01 +0.0879 +0.3577   |04   +1.6194E-01 +0.0742 +0.4320 |
•••
23   +7.6411E-03 +0.0035 +0.9968   |24   +6.9656E-03 +0.0032 +1.0000 |
25   +0.0000E+00 +0.0000 +1.0000   |26   +0.0000E+00 +0.0000 +1.0000 |
27   +0.0000E+00 +0.0000 +1.0000   |28   +0.0000E+00 +0.0000 +1.0000 |
29   +0.0000E+00 +0.0000 +1.0000   |30   +0.0000E+00 +0.0000 +1.0000 |
31   +0.0000E+00 +0.0000 +1.0000   |32   +0.0000E+00 +0.0000 +1.0000 |
33   +0.0000E+00 +0.0000 +1.0000   |34   +0.0000E+00 +0.0000 +1.0000 |
35   +0.0000E+00 +0.0000 +1.0000
```

Observer que le nombre de valeurs propres nulles est égal au nombre de variables.

```
File Mil.cmvp contains the eigenvalues and relative inertia for each axis
--- It has 35 rows and 2 columns

File Mil.cmco contains the column scores
--- It has 35 rows and 2 columns
File :Mil.cmco
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.343e+00| 1.189e+00|
|   2|-1.075e+00| 2.094e+00|
|----|----------|----------|
```

```
File Mil.cmli contains the row scores
--- It has 97 rows and 2 columns
File :Mil.cmli
|Col.|   Mini   |   Maxi   |
|----|----------|----------|
|   1|-1.241e+00| 1.116e+00|
|   2|-8.241e-01| 1.208e+00|
|----|----------|----------|
```

On a enfin une édition d'aides à l'interprétation spécifiques, reliée aux propriétés fondamentale de cette analyse. Pour chaque variable, chaque modalités et chaque facteur, la variance des scores des modalités rapportée à la variance des scores des individus (rapport de corrélation) indique dans quelle mesure un axe prend en compte une variable :

```
CorRatioMCA: Correlation ratios after a MCA
Title of the analysis: Mil.cm
Number of rows: 97, columns: 11

Variable : 1
> Categ=   1 Weight=  0.216  -1.202  -0.261
> Categ=   2 Weight=  0.186  -0.095   0.557
> Categ=   3 Weight=  0.227   0.075   1.094
> Categ=   4 Weight=  0.371   0.703  -0.795
----------------------> r=    0.500   0.578
```

50 % de la variance entre position des points s'exprime dans la variance entre position des modalités sur l'axe 1. Ce taux vaut 58 % sur l'axe 2. La variable joue un rôle dans l'interprétation des deux premiers facteurs.

```
Variable : 2
> Categ=   1 Weight=  0.237   0.484  -1.075
> Categ=   2 Weight=  0.289   0.172  -0.299
> Categ=   3 Weight=  0.299  -0.249   1.075
> Categ=   4 Weight=  0.175  -0.513   0.112
----------------------> r=    0.129   0.647

Variable : 3
> Categ=   1 Weight=  0.814   0.292   0.083
> Categ=   2 Weight=  0.186  -1.281  -0.364
----------------------> r=    0.374   0.030
```

•••

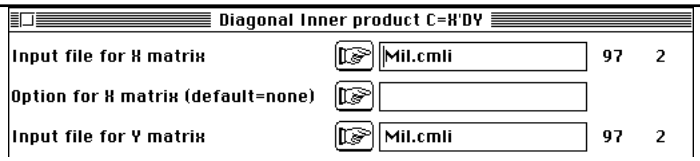
```
Variable : 11
> Categ=   1 Weight=  0.052   0.895   2.094
> Categ=   2 Weight=  0.371   0.483   0.018
> Categ=   3 Weight=  0.247  -0.305  -0.551
> Categ=   4 Weight=  0.155  -0.265  -0.341
> Categ=   5 Weight=  0.175  -0.622   0.425
----------------------> r=    0.229   0.351

File Mil.cmrc contains the correlation ratios between
the categorical variables and the factor scores
It has 11 rows and 2 columns
```

⚠ L'ACM code les lignes (paquets d'individus identiques) pour maximiser la moyenne des rapports de corrélation. Les modules graphiques permettent d'exprimer complètement ces propriétés[2]. On s'intéresse d'abord au premier axe. Calculer les moyennes des coordonnées factorielles des lignes, dans MatAlg :

Elles sont centrées :

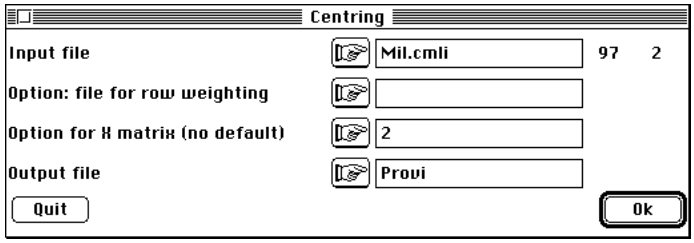
```
-----------------------------------------------------------------
[  1] -3.6485e-10
[  2]  4.1381e-09
-----------------------------------------------------------------
```

Calculer variances et covariances :

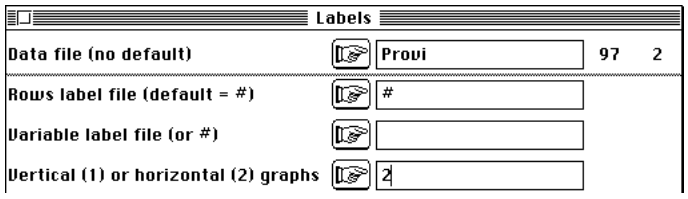


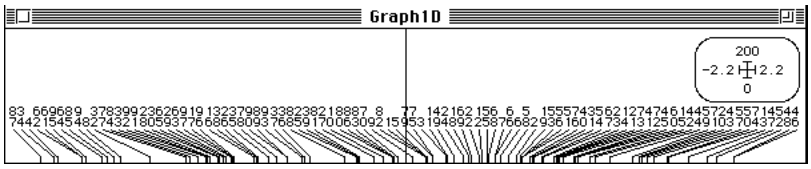

```
------------------------------------------------------------------
[  1]  3.8159e-01 -5.8374e-10
[  2] -5.8374e-10  2.0715e-01
------------------------------------------------------------------
```

Les variances sont égales aux valeurs propres et les covariances sont nulles. Normaliser ces coordonnées factorielles (Bin->Bin : Centring, ou DDUtil : Add normed scores) :

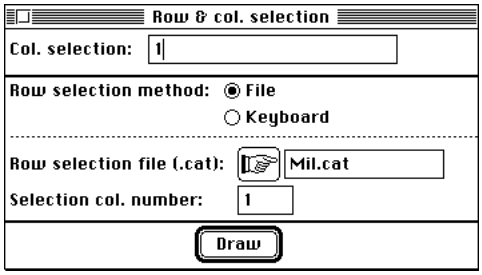


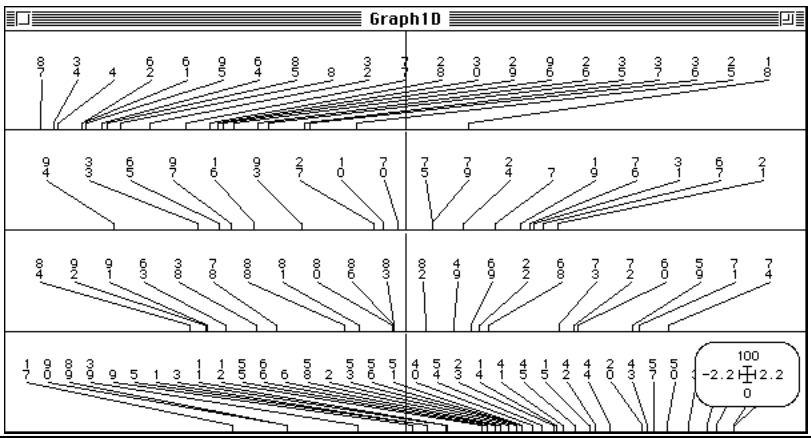
Représenter l'ordination des lignes du tableau par la première colonne de Provi (Graph1D) :



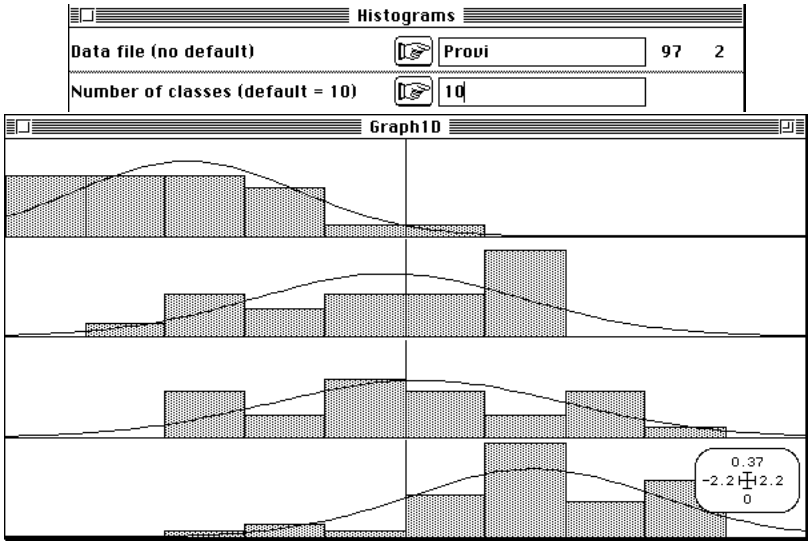


Les 97 lignes sont ordonnées sur un axe avec un code numérique centré de variance unité. Séparer ces lignes par paquets de porteurs d'une même modalité :

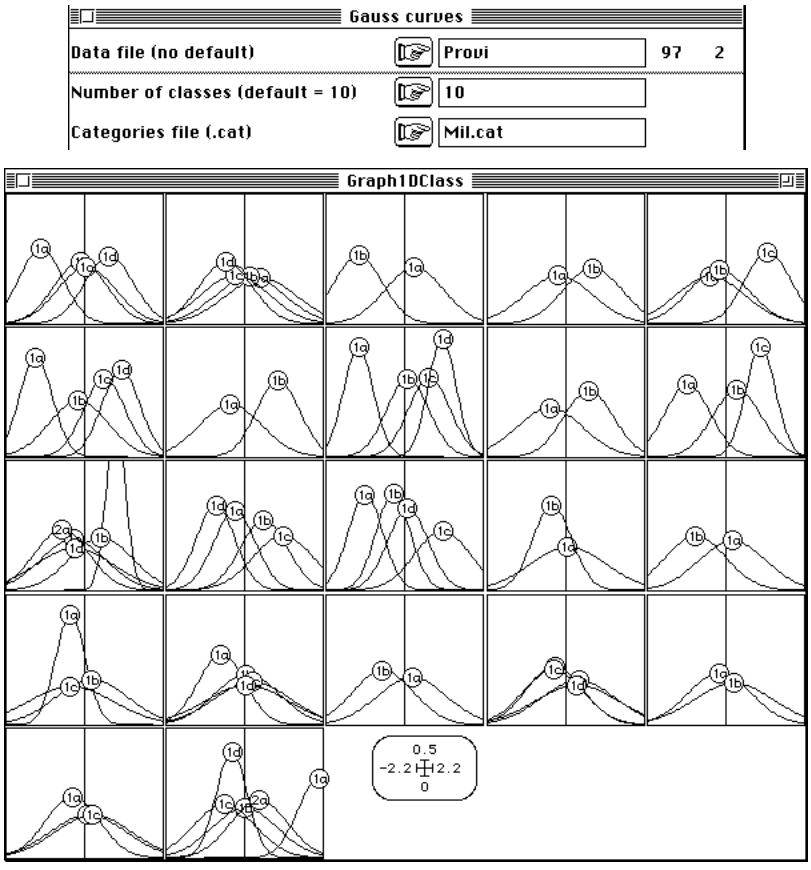




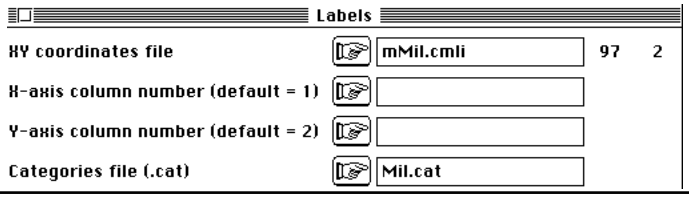
Chaque paquet définit une moyenne et la variance de ces moyennes est la part expliquée de l'ordination de départ par la première variable (50 %, dans le listing ci-dessus). Simplifier avec l'option Histograms :

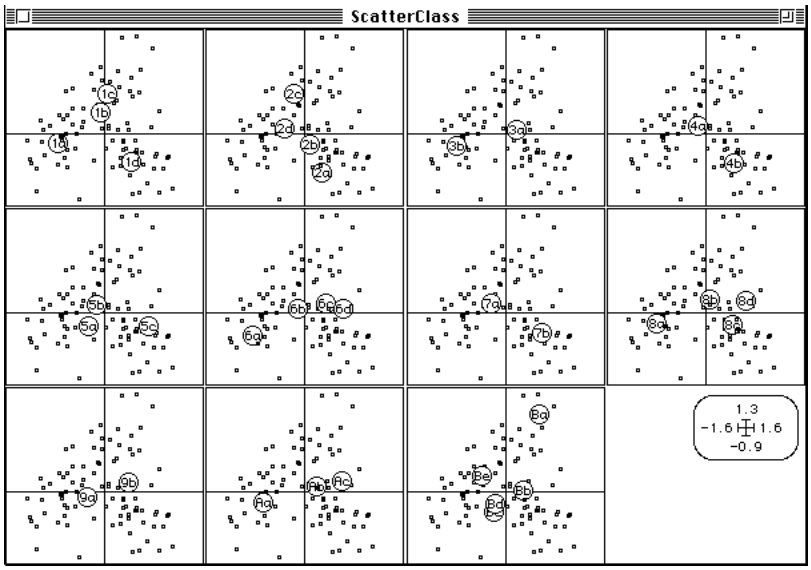


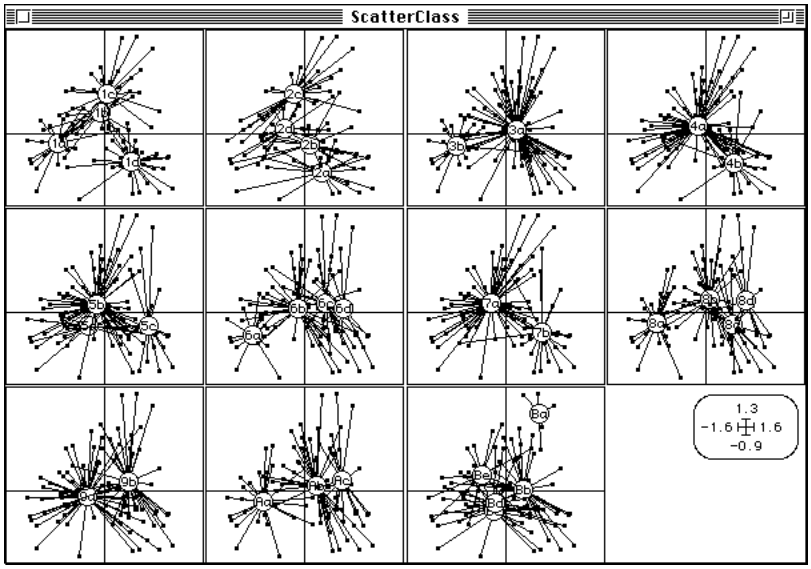
Généraliser avec le module Graph1DClass :

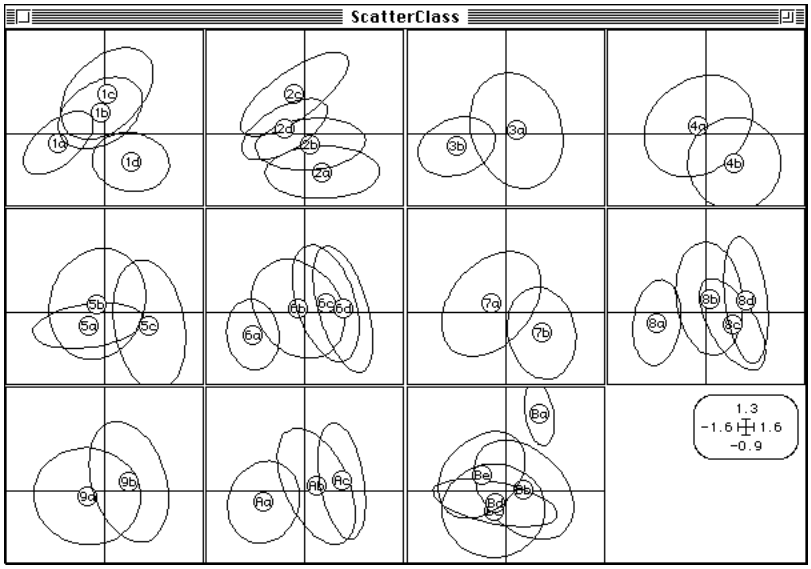


Sur une même ordination des lignes, l'analyse a optimisé la moyenne des variances des positions des modalités par averaging. On peut donc lire les associations entre modalités par le biais d'une ordination de référence des lignes du tableau. Le module ScatterClass généralise la représentation à deux dimensions :

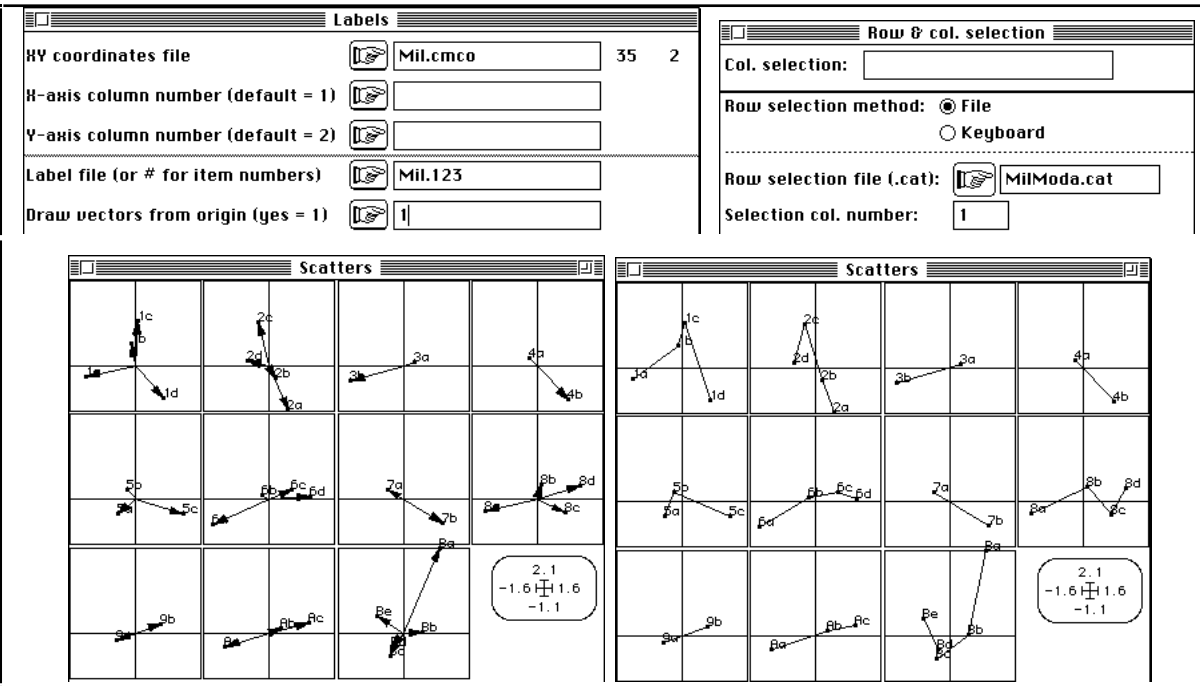








Les cartes factorielles des modalités résument l'information par les positions moyennes :



Après cette option, les fonctions de DDUtil sont disponibles.

[1] Tenenhaus, M. & Young, F.W. (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis ans other methods for quantifying categorical multivariate data. *Psychometrika* : 50, 91-119.

[2] Pialot, D., Chessel, D. & Auda, Y. (1984) Description de milieu et analyse factorielle des correspondances multiples. *Compte rendu hebdomadaire des séances de l'Académie des sciences. Paris*, D : 298, Série III, 11, 309-314.