

Consultations statistiques avec le logiciel 

Peut-on interpréter un seul facteur ?

Résumé

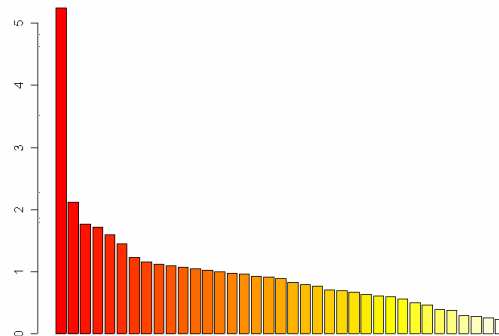
Il arrive souvent qu'une analyse multivariée conduite à l'interprétation d'un seul facteur. La fiche regroupe la description des outils graphiques qui sont utiles dans cette circonstance. Il n' y a aucune raison d'utiliser un plan factoriel pour exprimer un deuxième axe qui n'a aucune fonction. La fiche indique des outils d'expression d'un premier facteur dominant en ACP, AFC ou ACM.

Plan

1.	Introduction.....	2
2.	ACP normée et effet taille	2
	2.1. Interpréter le premier axe	3
	2.2. Quand le premier cache les autres	5
3.	Le cas de l'AFC	7
4.	Le cas de l'ACM	12
5.	Ordination réciproque sur un gradient	15
6.	Références	17

1. Introduction

On aborde ici une situation qui n'est pas rare et qu'on peut résumer simplement par l'apparition d'un graphe des valeurs propres du type :



La plupart des outils graphiques utilisés en analyse des données sont conçus pour étudier des plans factoriels. On est alors obligé de conserver un second facteur qu'on sait être d'intérêt nul uniquement pour représenter à deux dimensions une structure qui n'en n'a qu'une. Dans ade4, un effort a été fait pour posséder les outils graphiques associés. Ces outils renvoient aux théorèmes fondamentaux associés aux différents types d'analyse.

2. ACP normée et effet taille

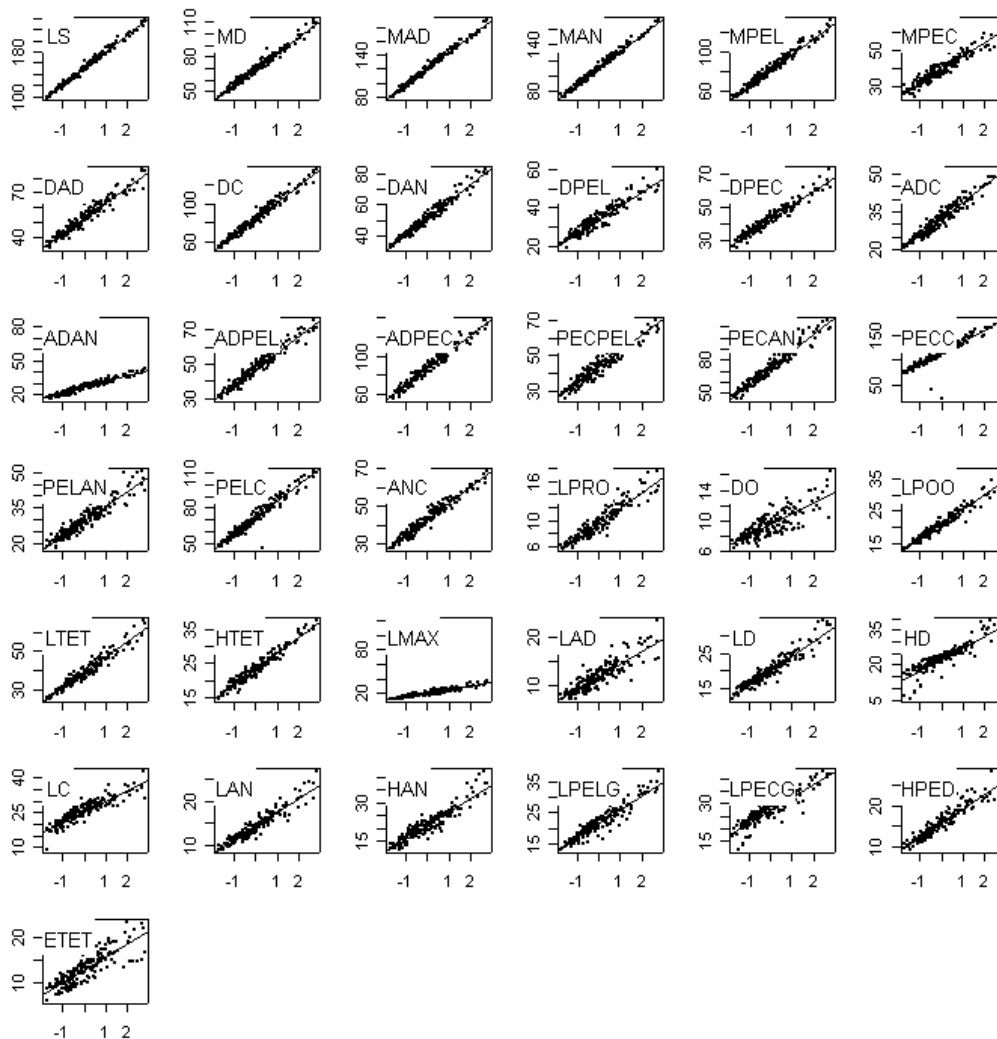
La question intervient souvent en cas d'*effet taille* : Cf. Lebart et al. (1995 p. 56). Toutes les variables sont corrélées positivement entre elles.

```
data(lascaux) # (Lascaux 1996)
pca1=dudi.pca(lascaux$morpho)
```



Le premier facteur suppose un traitement particulier. Utiliser le fait que la première coordonnée des lignes est une variable artificielle qui maximise (Hotelling 1933) la somme de ses carrés de corrélation avec les variables observées. Tracer donc les nuages coordonnée de rang 1 versus chacune des variables :

score (pcal)



La longueur standard (variable 1) est la meilleure indication de l'effet taille : elle mérite bien son nom. Les corrélations entre coordonnée des lignes de rang 1 et variables sont exactement les coordonnées des colonnes de rang 1 :

```
w=pcal$co[,1]
names(w)=row.names(pcal$co)
w
```

	LS	MD	MAD	MAN	MPEL	MPEC	DAD	DC	DAN	DPEL	DPEC	ADC	ADAN	ADPEL	ADPEC	PECPEL	PECAN	PECC	PELAN	PELC	ANC	LPRO	DO	LPOO	LTET	HTET	LMAX	LAD	LD	HD	LC	LAN	HAN	LPELG	LPECG	HPED	ETET
0.9938	0.9859	0.9930	0.9915	0.9826	0.9394	0.9676	0.9829	0.9798	0.9481	0.9643	0.9655	0.7083	0.9732	0.9791	0.9418	0.9684	0.9249	0.9414	0.9678	0.9734	0.9391	0.8082	0.9735	0.9690	0.9741	0.5149	0.8770	0.9507	0.8978	0.9038	0.9486	0.9335	0.9450	0.9110	0.9394	0.8408	

Noter pour les variables ADAN et LMAX la présence certaine de données aberrantes.

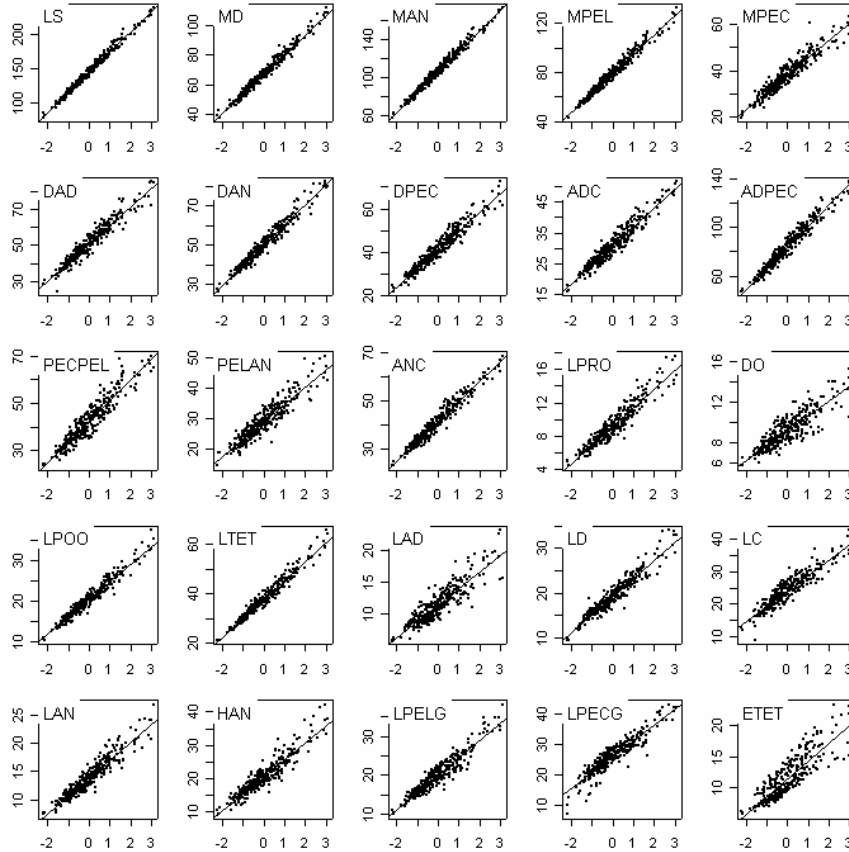
2.1. Interpréter le premier axe

Enlever les variables à données manquantes et ADAN puis refaire cette analyse :

```

non=unlist(apply(lascaux$morpho,2,function(x) any(is.na(x))))
tab=tab[, -10]
pca2=dudi.pca(tab)
Select the number of axes: 3
score(pca2)

```



Dans cet exemple, manipuler le plan 1-2, comme il est fait dans la plupart des cas, est une erreur totale. La redondance des variables est un fait qui perturbe profondément l'analyse. Ce fait (la variation de taille) existe, il doit être étudié seul. Le code génétique (Lascaux, op. cit. p.20) est basé sur 6 marqueurs allozymiques diagnostiques des populations de truites méditerranéenne (sauvages) et atlantiques (domestiques). La variable donne le nombre d'allèles méditerranéens (ancestraux) possédés par le poisson pour les trois systèmes enzymatiques. 0 indique un poisson assimilé à un homozygote atlantique (domestique) et 6 indique un poisson assimilé à un homozygote méditerranéen (sauvage). Cette variable est typiquement un facteur à modalité ordonnée :

```
gen=as.ordered(lascaux$gen)
```

La rivière définit l'environnement, c'est une autre variable importante

```
riv=lascaux$riv
```

```
anova(lm(pca2$li[,1]~riv*gen))
```

Analysis of Variance Table

```

Response: pca2$li[, 1]
      Df Sum Sq Mean Sq F value Pr(>F)
riv    11  2391    217  16.26 <2e-16
gen     6   580     97   7.23  4e-07
riv:gen 42   634     15   1.13  0.28
Residuals 246  3289     13

```

Les deux variables jouent sur la taille sans interaction :

```
summary(lm(pca2$li[,1]~riv+gen))
```

Call:

```
lm(formula = pca2$li[, 1] ~ riv + gen)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-9.7095 -2.4182 -0.0345  1.9138 17.0364
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.9165	1.1231	5.27	2.7e-07
rivAlemany	-7.5320	1.2676	-5.94	8.1e-09
rivAude	-7.5812	1.4262	-5.32	2.1e-07
rivBoutadiol	-9.3615	1.3445	-6.96	2.3e-11
rivCampcardos	-4.2467	1.3174	-3.22	0.0014
rivCampeilles	2.7189	1.8265	1.49	0.1377
rivEyne	-7.0610	1.3281	-5.32	2.1e-07
rivLladure	-8.6187	1.3145	-6.56	2.5e-10
rivNohedes	-3.7652	1.2680	-2.97	0.0032
rivorbieu	-1.3198	1.3525	-0.98	0.3300
rivRiu_Ferrer	-8.1418	1.3821	-5.89	1.1e-08
rivTech	-7.9905	1.3262	-6.03	5.2e-09
gen.L	4.0343	0.6907	5.84	1.4e-08
gen.Q	0.7351	0.6386	1.15	0.2506
gen.C	0.0466	0.6481	0.07	0.9427
gen^4	0.7243	0.6429	1.13	0.2609
gen^5	-0.4071	0.7000	-0.58	0.5613
gen^6	-0.2221	0.7419	-0.30	0.7649

Residual standard error: 3.69 on 288 degrees of freedom
 Multiple R-Squared: 0.431, Adjusted R-squared: 0.397
 F-statistic: 12.8 on 17 and 288 DF, p-value: <2e-16

L'ennui (pourquoi ?) :

```
table(riv,gen)
```

	gen						
riv	0	1	2	3	4	5	6
Agly	8	3	1	0	0	0	0
Alemany	16	6	1	2	3	0	2
Aude	0	0	0	0	0	5	25
Boutadiol	5	0	1	1	8	5	10
Campcardos	5	4	3	6	9	2	1
Campeilles	0	0	0	0	0	0	8
Eyne	4	3	2	3	2	4	12
Lladure	8	2	0	3	4	5	8
Nohedes	14	12	2	2	0	0	0
orbieu	3	7	7	5	1	0	1
Riu_Ferrer	2	1	0	2	4	8	10
Tech	8	3	4	0	1	3	6

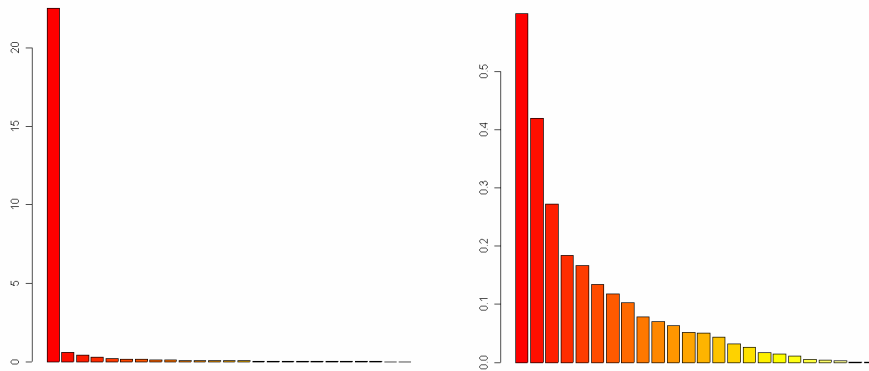
En cas d'ACP à un seul facteur de signification sure, on évitera donc les cartes factorielles abusives.

2.2. Quand le premier cache les autres

Mais derrière le premier, peut aussi se cacher d'autres informations. Il convient d'éviter systématiquement les conseils de ceux qui savent ce qu'il faut faire ou connaître sérieusement les domaines de validité des méthodes proposées, car les données réelles peuvent prendre en défaut les meilleures intentions.

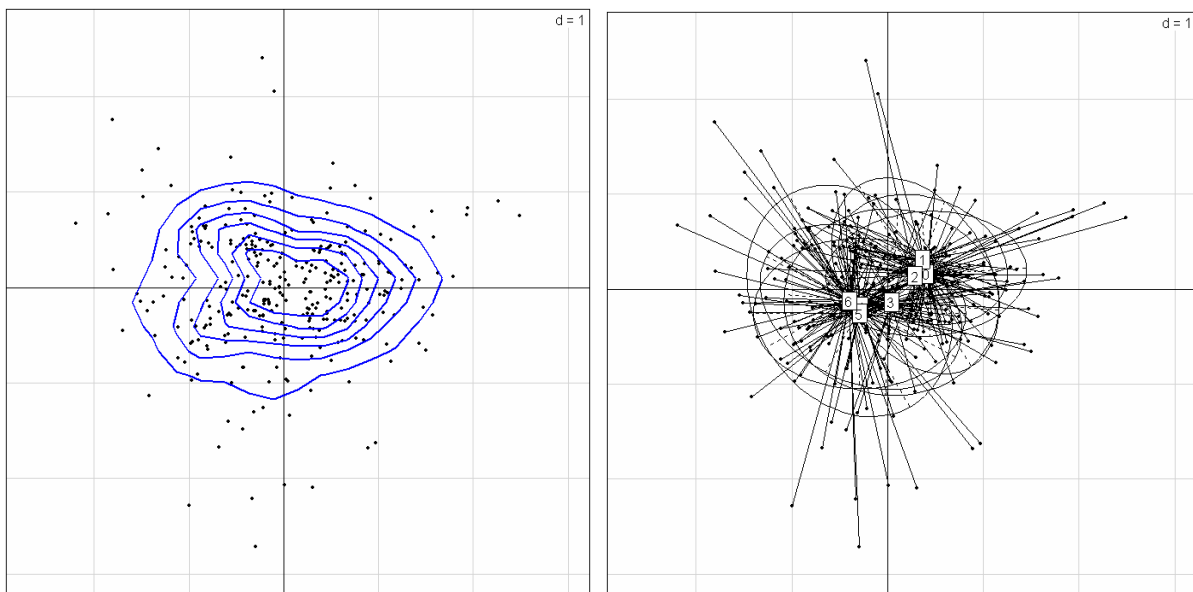
```
barplot(pca2$eig)
```

```
barplot(pca2$eig[-1])
```



Les valeurs propres suivantes sont inférieures à 1. On conseille parfois de les éliminer (Kaiser-Guttman criterion dans Jackson 1993).

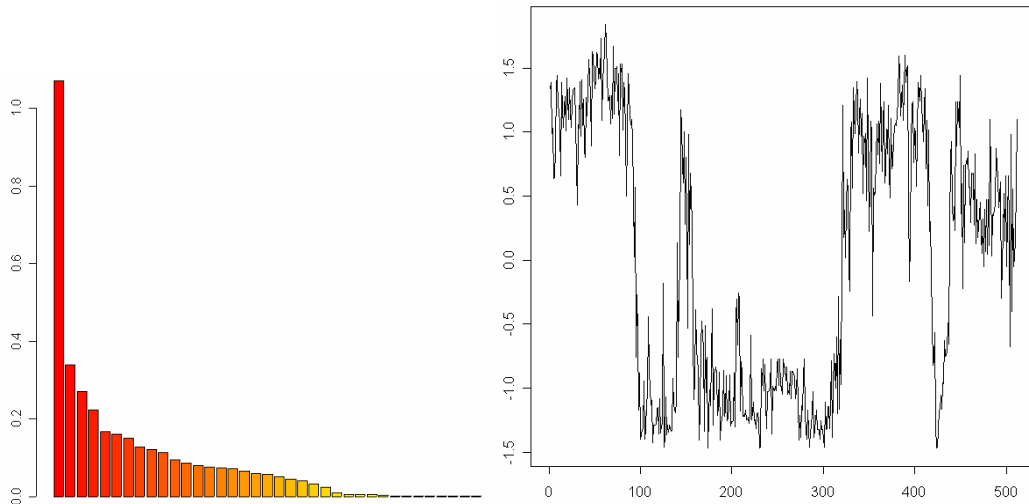
```
s.kde2d(pca2$li, 2, 3)
s.class(pca2$li, lascaux$gen, yax=2, yax=3)
```



La forme a aussi une composante génétique et/ou environnementale à étudier.

```
data(steppe) # 512 placettes sur un transect
pca3=dudi.pca(steppe$tab, scale=F)
```

```
Select the number of axes: 1
```



Quand il y a un axe très marqué, il faut l'analyser seul. Par contre, quand deux valeurs propres sont voisines, il convient de comprendre que les axes définissent un plan qui est à un sens, mais que les axes eux-mêmes peuvent très bien ne pas en avoir. En effet, si le nuage initial est proche d'un plan, ce plan sera celui des deux premiers axes principaux. Mais si la projection sur ce plan a une forme plutôt circulaire, un couple arbitraire d'axes servira seulement à définir le plan mais pas à identifier, dans ce plan, des directions privilégiées. *Autant un axe isolé doit être interprété seul, autant deux axes à inertie voisine doivent être interprétés ensemble.*

```
pca4=dudi.pca(lascaux$colo)
```

```
Select the number of axes: 2
```

```
s.corcircle(pca4$co,clab=0.75)
```

```
s.corcircle(as.data.frame(unclass(varimax(as.matrix(pca4$co),nor=F)$loadings)),c  
lab=0.75)
```



3. Le cas de l'AFC

Le graphe canonique de l'AFC est tout aussi explicite.

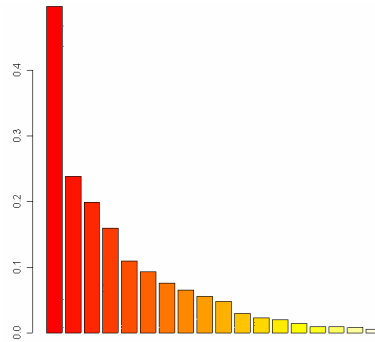
```
data(jv73) # (Verneaux 1973)
```

```
names(jv73)
```

```
[1] "morpho" "phychi" "poi" "xy" "contour" "fac.riv"
```

```
coal=dudi.coa(jv73$poi)
```

```
Select the number of axes: 1
```

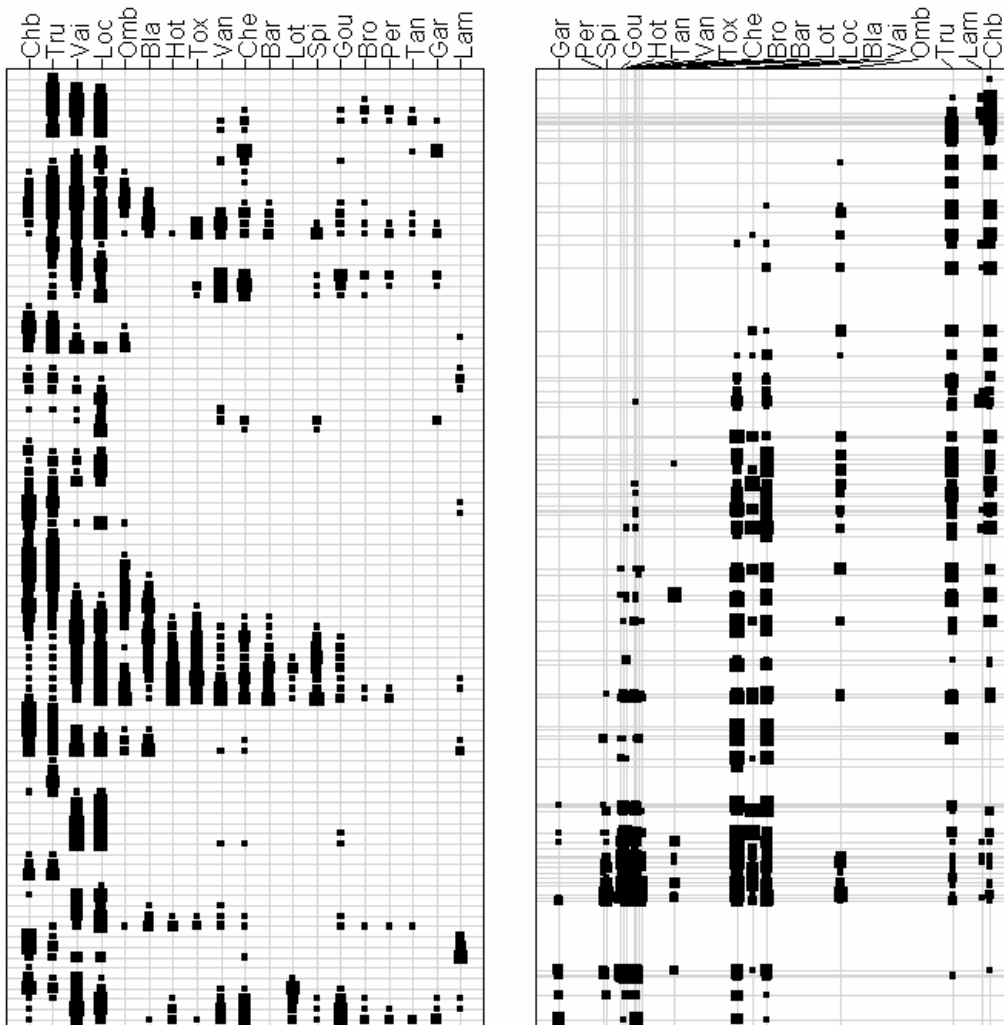


Un facteur contient l'essentiel de la structure. Représenter les données :

```
par(mfrow=c(1,2))
```

```
table.value(jv73$poi,csi=0.5,clabel.r=0,cleg=0)
```

```
table.value(jv73$poi,csi=0.5,clabel.r=0,  
cleg=0,y=coal$li[,1],x=coal$co[,1])
```



L'analyse donne un score des lignes et un score des colonnes qui maximise la corrélation mesurée à travers le tableau (Williams 1952). Cette propriété est exprimée par le tableau dont les lignes et les colonnes sont positionnées par la première coordonnée (Hill 1974). En 1952, Williams écrit en introduction :

A number of writers have discussed the analysis of data in the form of contingency tables, in which association is known to exist. A useful way of interpreting such association is to regard it as being due to the correlation between a pair of variates, corresponding to the classes of the two classifications of the table. According to one method, the contingency table is regarded as a frequency table for a sample from a bivariate normal population, classified into intervals of the two variates, the limits of which are to be estimated ...

La deuxième dimension graphique est utilisée pour les données. Il n'y a pas d'effet Guttman parce qu'il n'y a pas de taxon caractéristique du milieu du gradient.

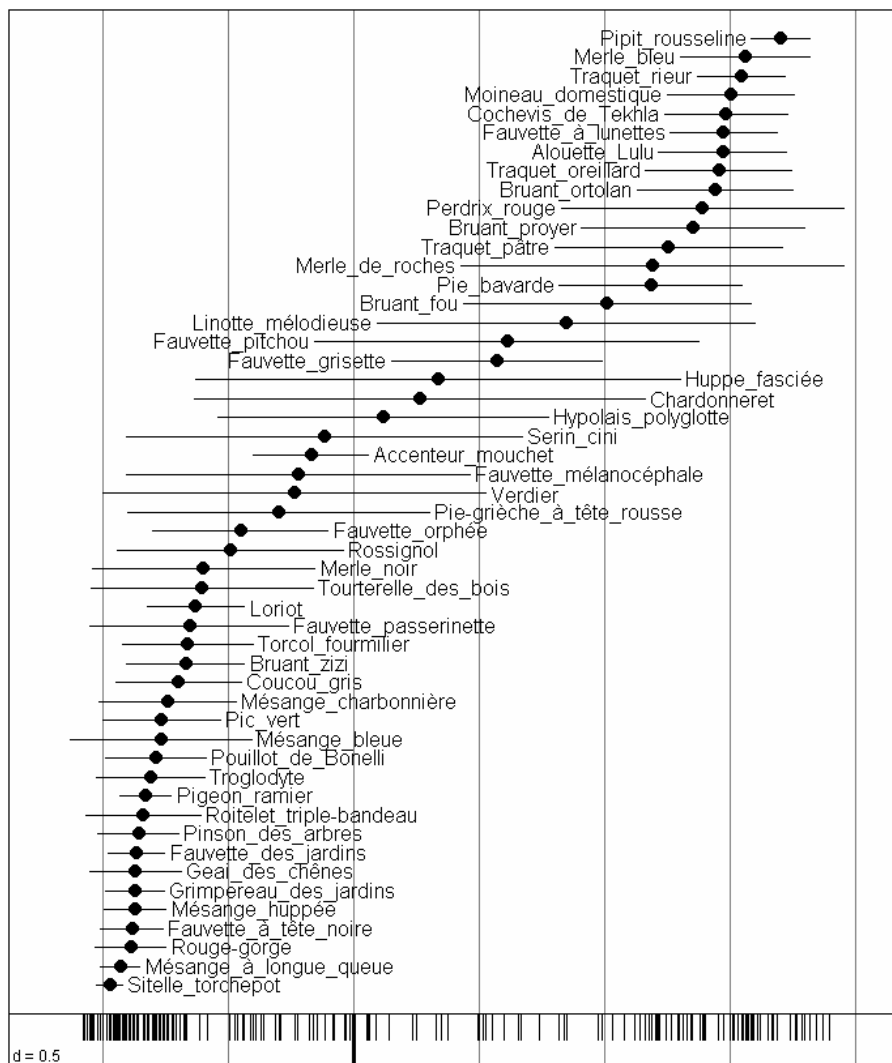
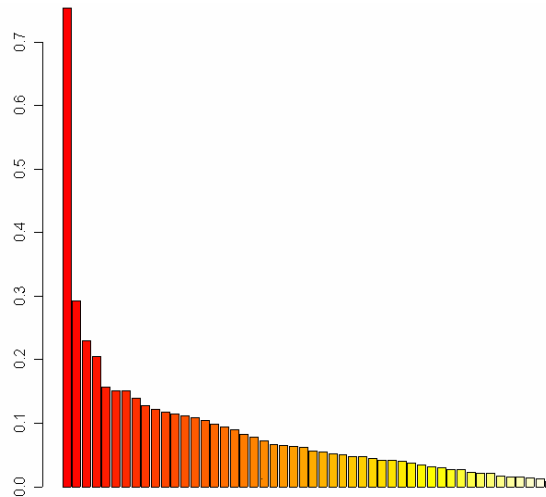


Figure 1 : Analyse des correspondances et graphe canonique du modèle de discrimination. Position de 182 relevés par la première coordonnée. Position des espèces par averaging. Optimisation de la variance entre les positions moyennes. Voir Lebreton and Yoccoz (1987) pour une vision globale du rôle de l'AFC en ornithologie.

```

data(rpjdl) # (Prodon and Lebreton 1981)
names(rpjdl)
[1] "fau" "mil" "frlab" "lab" "lalab"
coa2=dudi.coa(rpjdl$fau)
Select the number of axes: 4

```



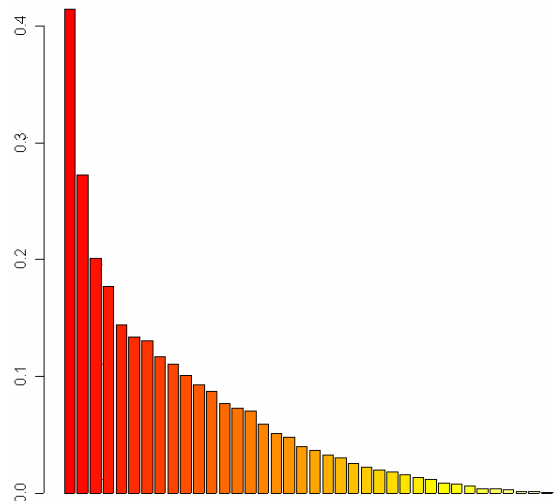
```
sco.distri(coa2$li[,1], rpjdl$fau, labe=rpjdl$frlab)
```

La fonction **sco.distri** permet de simplifier la représentation d'une ligne ou d'une colonne par la seule expression de la position moyenne et de l'écart-type, représentation introduite dans Chessel et al. (1982). Le premier facteur est examiné de manière particulière mais n'est pas, ici, le seul interprétable.

```

data(aviurba)
coa3=dudi.coa(aviurba$fau)
Select the number of axes: 2

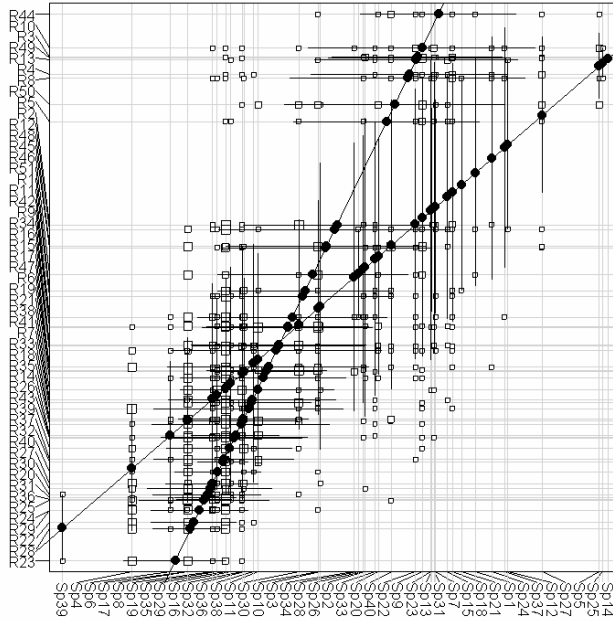
```



```

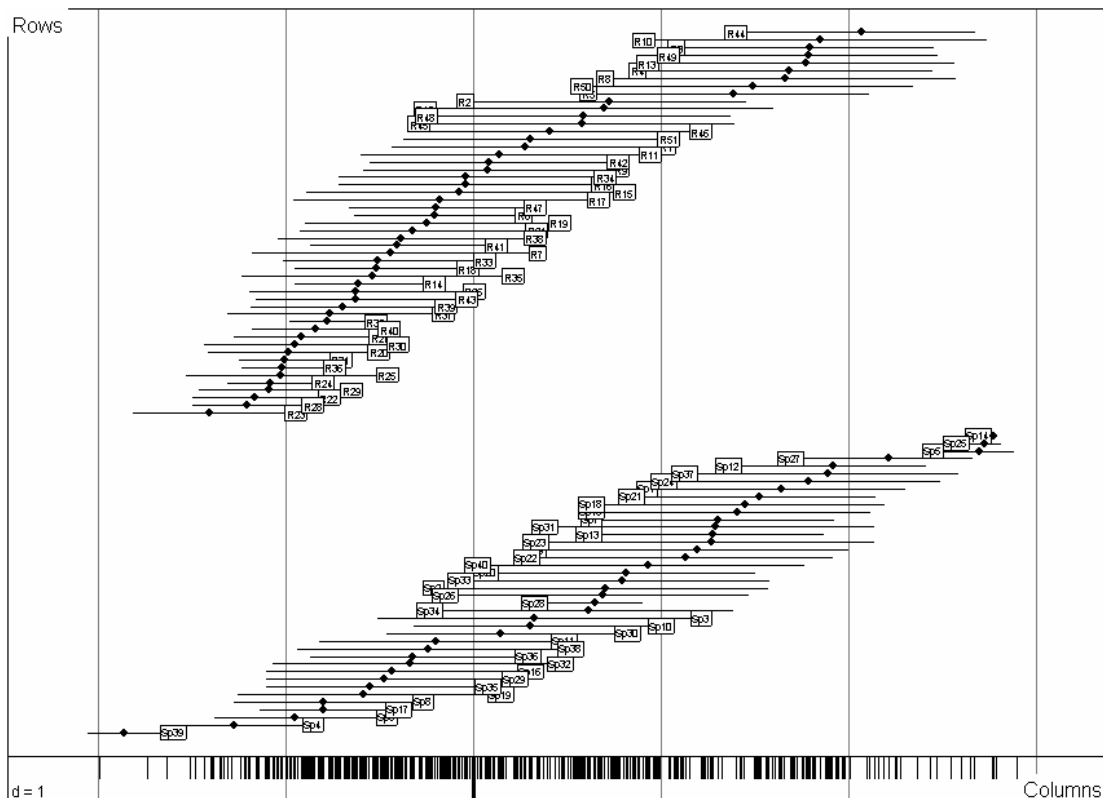
table.cont(aviurba$fau, coa3$co[,1], coa3$li[,1], csi=0.25, abmean.x=T,
  abline.x=T, abmean.y=T, abline.y=T)

```



On voit le résultat du célèbre théorème d'Hirschfeld (1935) qui dit que pour obtenir une double régression linéaire à travers une table de contingence il faut et il suffit de prendre un couple de coordonnées de l'analyse des correspondances. La procédure ne porte évidemment pas ce nom mais nul ne peut contester qu'elle est décrite dans cet article sans aucune trace d'ambiguïté. Que donnera l'ordre suivant **abline(c(0,1), lwd=3, col="red")**

Quand on n'utilise qu'un des côtés de cette démarche symétrique (code numérique des relevés et les espèces à la moyenne de leur distribution), on optimise la description de la séparation des niches (Hill 1977). En fait, même sans le savoir la procédure inverse (code numérique des espèces et relevés à la moyenne de leur distribution) est en oeuvre.

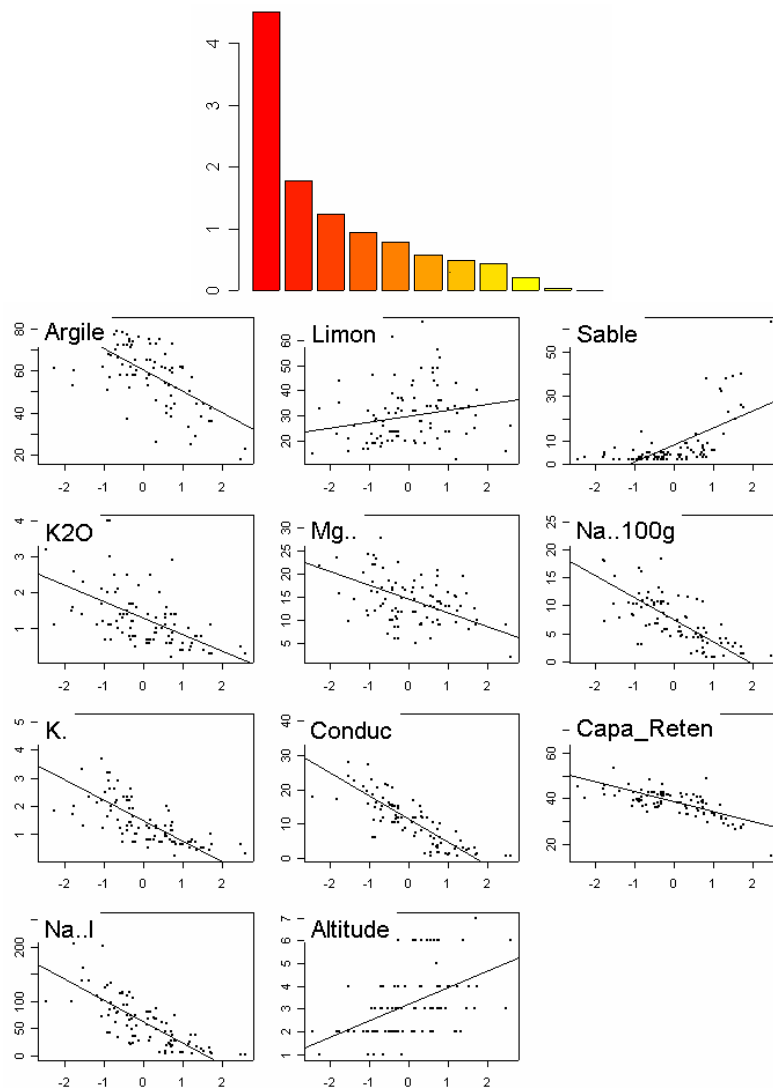


```
score(coa3,clab.r=0.5,clab.c=0.5)
```

Rappelons enfin que les lignes, les colonnes et les cases d'un tableau relevant de l'AFC sont justiciables d'une représentation cohérente sur une seule échelle (Thioulouse and Chessel 1992). C'est un cas particulier de l'analyse des correspondances multiples, dont nous discutons maintenant.

4. Le cas de l'ACM

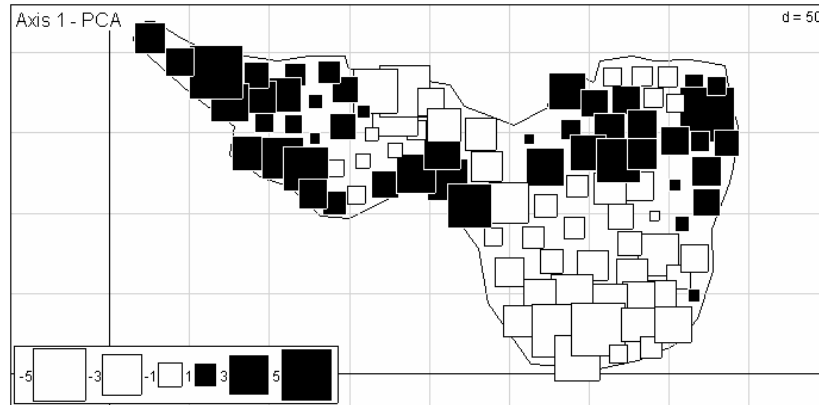
```
data(mafragh) # (de Belair and Bencheikh-Lehocine 1987)
pca5=dudi.pca(mafragh$mil)
Select the number of axes: 1
```



Quand l'altitude augmente, la proportion d'argile diminue, la salinité et la capacité de rétention aussi.

L'interprétation indépendante du facteur 1 s'impose.

```
contour=area.util.contour(mafragh$area)
s.value(mafragh$xy, pca5$li[,1], sub = "Axis 1 - PCA", contour=contour)
```

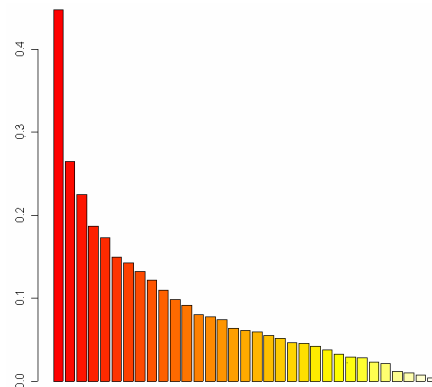


Passons toutes les variables dans le mode qualitatif :

```
mil0=as.data.frame(lapply(mafragh$mil,
  function(x) factor(cut(x,br=quantile(x,seq(0,1,le=5)),inc=T))))
summary(mil0)
```

Argile	Limon	Sable	K2O	Mg..
[18,52]:25	[13,22]:26	[2,3]:44	[0.2,0.7]:26	[2,10]:25
(52,62]:27	(22,28]:24	(3,4):11	(0.7,1.1]:24	(10,13.5]:24
(62,73]:24	(28,35]:23	(4,7):19	(1.1,1.6]:24	(13.5,17.1]:24
(73,83]:21	(35,67]:24	(7,63]:23	(1.6,4]:23	(17.1,31.4]:24
Na..100g	K.	Conduc	Capa_Reten	
[0.7,4.1]:26	[0.2,0.7]:25	[0.6,4.3]:25	[14.6,35.3]:25	
(4.1,6.8]:25	(0.7,1.2]:25	(4.3,11.3]:24	(35.3,38.8]:25	
(6.8,10.6]:22	(1.2,2]:23	(11.3,15.5]:25	(38.8,41.6]:23	
(10.6,25.2]:24	(2,5.1]:24	(15.5,40.5]:23	(41.6,74.8]:24	
Na..1	Altitude			
[2.2,24]:26	[1,2]:41			
(24,57.6]:23	(2,3):24			
(57.6,94]:26	(3,4):15			
(94,254]:22	(4,7):17			

```
apply(mafragh$mil,2,function(x) cut(x,br=quantile(x,seq(0,1,by=0.2))))
```



```
mca1=dudi.acm(mil0)
```

Select the number of axes: 1

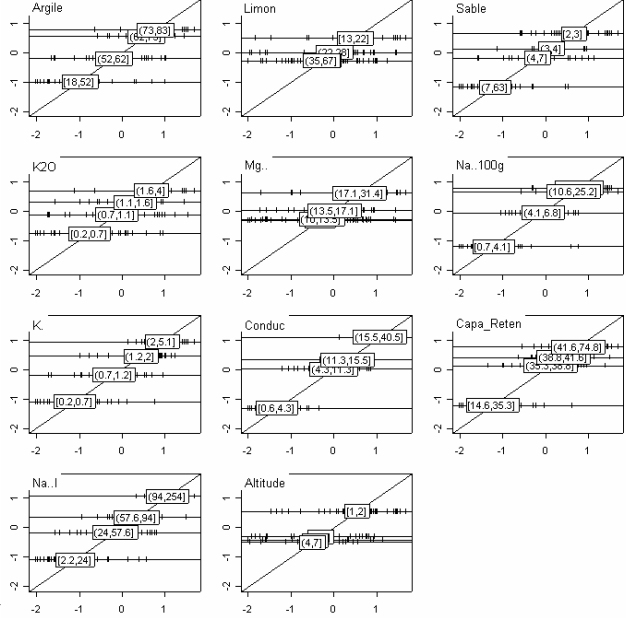
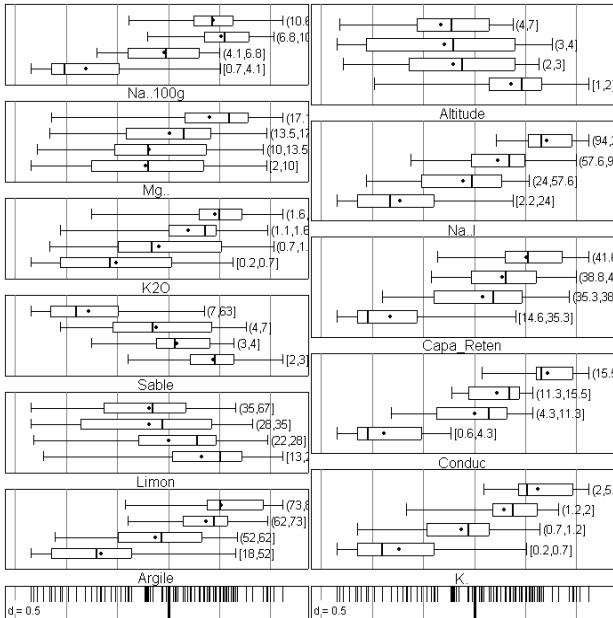
```
score(mca1)
```

On cherche l'illustration graphique du théorème (Saporta 1975) : l'ACM définit un code des lignes qui maximise la moyenne des rapports de corrélation pour chaque variable. C'est l'expression des résultats numériques contenus dans :

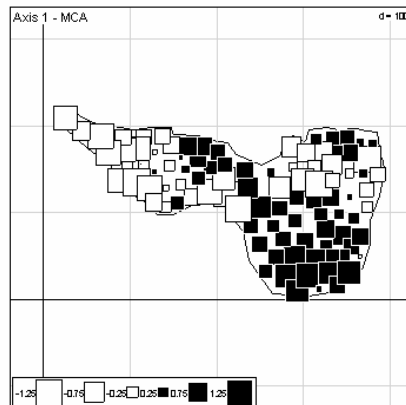
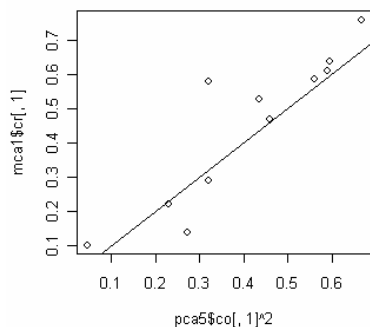
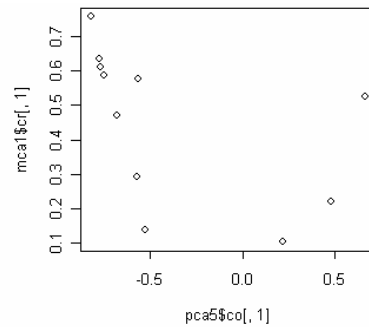
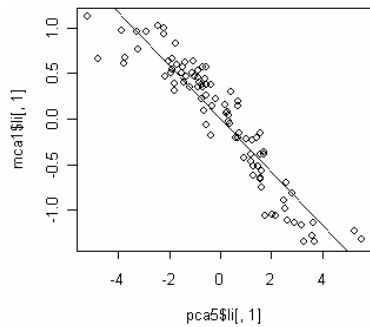
mca1\$scr	RS1	RS1	RS1
Argile	0.4689	Mg..	0.1389
Limon	0.1026	Na..100g	0.6359
Sable	0.5270	K.	0.5870
K2O	0.2909	Conduc	0.7576
		Capa_Reten	0.5776
		Na..1	0.6099
		Altitude	0.2200

Dans Pialot et al. (1984), on utilise des courbes de Gauss pour décrire les variances. Ici, on se contente d'un boxplot à gauche ou de la transposition du rapport de corrélation en carré de corrélation à droite. La représentation de la totalité des points garantit qu'on ne fera pas un usage abusif du modèle très simplifié.

```
par(mfrow=c(1,2))
sco.boxplot(mca1$li[,1],mi10[,1:6])
sco.boxplot(mca1$li[,1],mi10[,7:11])
```



```
par(mfrow=c(2,2))
plot(pca5$li[,1],mca1$li[,1]) ; abline(lm(mca1$li[,1]~pca5$li[,1]))
plot(pca5$co[,1],mca1$cr[,1])
plot(pca5$co[,1]^2,mca1$cr[,1]) ; abline(c(0,1))
s.value(mafragh$xy, mca1$li[,1], sub = "Axis 1 - MCA",contour=contour)
```



On peut comparer point par point les deux analyses. Les scores des individus sont très corrélés, les carrés de corrélations très liés aux rapports de corrélation, la structure spatiale est identique. En ACM, comme en ACP ou en AFC, le dépouillement d'un axe ne pose pas de problèmes particuliers. Ceci invite à ne plus faire l'impasse sur une structure multivariée clairement unidimensionnelle. Le plus étonnant, c'est que le detrending (Hill and Gauch 1980), défendu par les uns (Peet et al. 1988) et rejeté par d'autres (Wartenberg et al. 1987), discuté en tout cas (Jackson and Somers 1991), qui vise à l'élimination de l'effet Guttman, s'il existe, rend non interprétable le facteur 2 et donc impose une lecture unidimensionnelle des résultats. Pourtant les cartes factorielles en AFC détendancée sont fréquentes.

5. Ordination réciproque sur un gradient

En dernier ressort, la plus grande attention est de rigueur : les résultats obtenus, même d'apparence simples, n'imposent pas une interprétation. La plasticité de la relation entre la donnée et la méthode est considérable : le plus difficile est de se laisser faire, c'est-à-dire laisser la structure émerger de l'analyse (Benzécri 1969).

```
coa2=dudi.coa(rpjd1$fau)
```

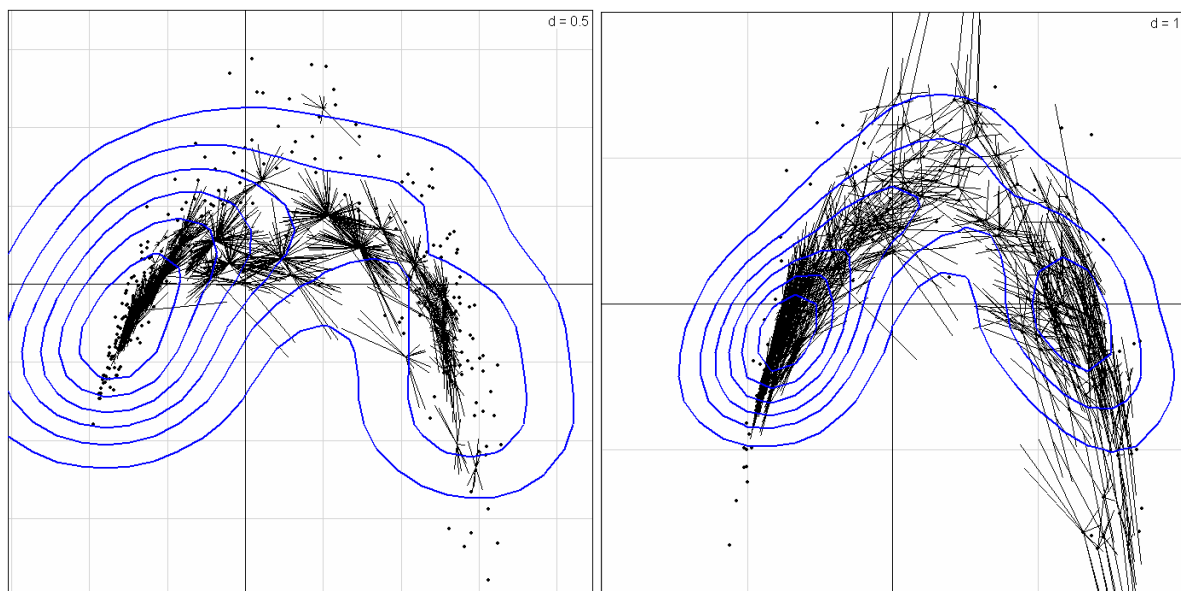
```
Select the number of axes: 2
```

```
s.distri(coa2$li, rpjd1$fau, csta=0.3, cell=0)
```

```
s.distri(coa2$co, t(rpjd1$fau), csta=0.3, cell=0)
```

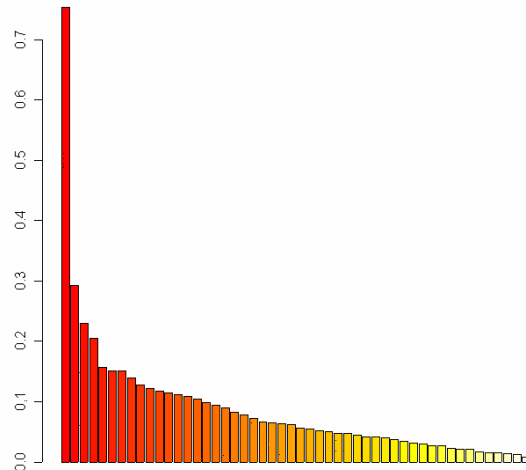
```
s.distri(coa2$c1, t(rpjd1$fau), csta=0.3, cell=0, xlim=c(-2,2), ylim=c(-2,2))
```

```
s.kde2d(coa2$li, add.p=T)
```



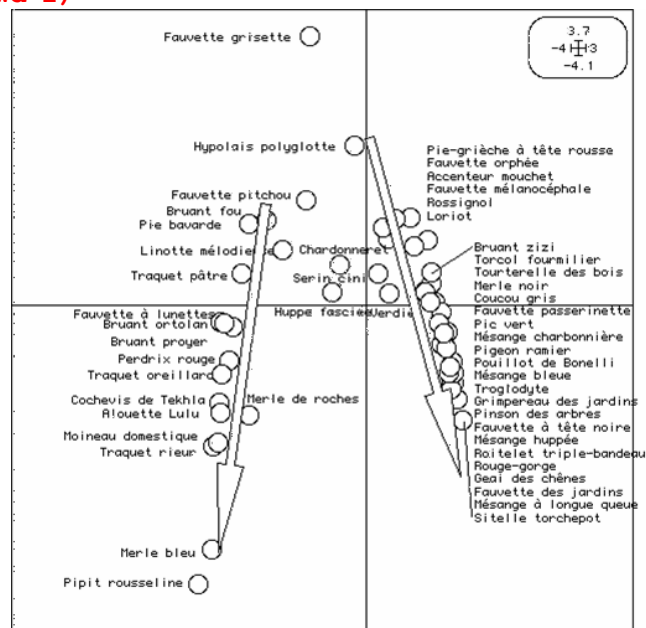
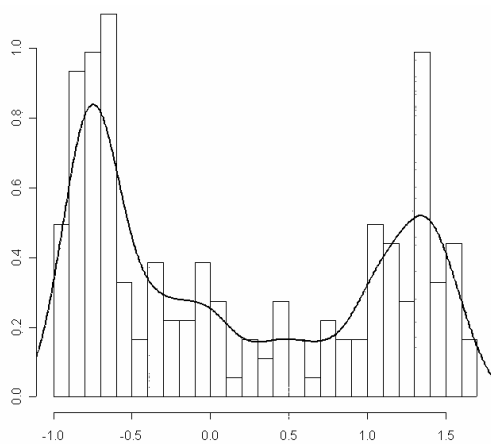
A gauche, relevés avec des scores de variance 1, espèces à la moyenne, densité des espèces dans le plan. A droite, espèces avec des scores de variance 1, relevés à la moyenn, densité de points relevés dans le plan.

Cette représentation laisse présente la possibilité que les deux extrémités du gradient sont compressées. C'est l'un des principaux défauts que Hill et Gauch (1980) attribuent à l'AFC (Voir aussi Gauch (1982 p. 151). On pourra aussi invoquer l'effet Guttman et son empreinte défavorable qui conduit au detrending. On pourra arguer que le graphe des valeurs propres définit sans ambiguïté un gradient :



L'arch effect n'est pas une conséquence directe de la présence d'un gradient. Il faut se souvenir que la propriété fondamentale de l'AFC est d'être une **analyse canonique** (Estève 1978). La baisse radicale de la valeur propre (carré de corrélation canonique) indique que le premier axe est radicalement plus important que les autres. Il faut se faire alors à l'idée que la structure principale n'est pas un gradient :

```
hist(coa2$li[,1],proba=T,nclass=20)
lines(density(coa2$li[,1],a=0.5),lwd=2)
```



Il y a deux gradients et une articulation par les espèces à spectre large.

On retiendra de ces exemples que le modèle canonique de l'AFC s'il est d'une efficacité certaine ne doit pas être employée sans discernement. Le graphe canonique montre le mélange des types de structure (partition et ordination). On dispose d'outil facilitant la construction de modèles à partir des données dans des situations variées aussi bien unidimensionnelles que complexes. On peut dire qu'ils permettent de voir l'intérieur de l'organisation des objets étudiés.

6. Références

- Benzécri, J. P. 1969. Statistical analysis as a tool to make patterns emerge from data. Pages 35-60 in S. Watanabe, editor. Methodologies of pattern recognition. Academic Press, New York.
- Chessel, D., J. D. Lebreton, and R. Prodon. 1982. Mesures symétriques d'amplitude d'habitat et de diversité intra-échantillon dans un tableau espèces-relevés: cas d'un gradient simple. *Compte rendu hebdomadaire des séances de l'Académie des sciences. Paris, D* **295**:83-88.
- de Belair, G., and M. Bencheikh-Lehocine. 1987. Composition et déterminisme de la végétation d'une plaine côtière marécageuse : La Mafragh (Annaba, Algérie). *Bulletin d'Ecologie* **18**:393-407.
- Estève, J. 1978. Les méthodes d'ordination : éléments pour une discussion. Pages 223-250 in J. M. Legay and R. Tomassone, editors. *Biométrie et Ecologie*. Société Française de Biométrie, Paris.
- Gauch, H. G. J. 1982. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge.
- Hill, M. O. 1974. Correspondence analysis : A neglected multivariate method. *Applied Statistics* **23**:340-354.
- Hill, M. O. 1977. Use of simple discriminant functions to classify quantitative phytosociological data. Pages 181-199 in E. Diday, editor. *Proceedings of the First International Symposium on Data Analysis and Informatics*. INRIA Rocquencourt, France.
- Hill, M. O., and H. G. Gauch. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* **42**:47-58.
- Hirschfeld, H. O. 1935. A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences* **31**:520-524.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**:498-520.
- Jackson, D. A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**:2204-2214.
- Jackson, D. A., and K. M. Somers. 1991. Putting things in order: the ups and downs of detrended correspondence analysis. *The American Naturalist* **137**:707-712.
- Lascaux, J. M. 1996. Analyse de la variabilité morphologique de la truite commune (*Salmo trutta* L.) dans les cours d'eau du bassin pyrénéen méditerranéen. Thèse de doctorat en sciences agronomiques, INP Toulouse.
- Lebart, L., A. Morineau, and M. Piron. 1995. *Statistique exploratoire multidimensionnelle*. Dunod, Paris.
- Lebreton, J. D., and N. Yoccoz. 1987. Multivariate analysis of bird count data. *Acta Oecologica, Oecologia Generalis* **8**:125-144.
- Peet, R. K., R. G. Knox, J. S. Case, and R. B. Allen. 1988. Putting things in order : the advantages of detrended correspondence analysis. *The American Naturalist* **131**:924-934.
- Pialot, D., D. Chessel, and Y. Auda. 1984. Description de milieu et analyse factorielle des correspondances multiples. *Comptes rendus hebdomadaires des Séances de l'Académie des Sciences, D* **298**:309-314.
- Prodon, R., and J. D. Lebreton. 1981. Breeding avifauna of a Mediterranean succession : the holm oak and cork oak series in the eastern Pyrénées. 1 : Analysis and modelling of the structure gradient. *Oikos* **37**:21-38.

- Saporta, G. 1975. Liaisons entre plusieurs ensembles de variables et codage de données qualitatives. Thèse de 3^o cycle, Université Pierre et Marie Curie, Paris VI.
- Thioulouse, J., and D. Chessel. 1992. A method for reciprocal scaling of species tolerance and sample diversity. *Ecology* **73**:670-680.
- Verneaux, J. 1973. Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon.
- Wartenberg, D., S. Ferson, and F. J. Ohlf. 1987. Putting things in order : a critique of detrended correspondence analysis. *The American Naturalist* **129**:434-448.
- Williams, E. J. 1952. Use of scores for the analysis of association in contingency tables. *Biometrika* **39**:274-289.