

Comment projeter des individus ou des variables supplémentaires ?

D. Chessel & J.R. Lobry

La question de la projection des individus et des variables supplémentaires est souvent posée. On peut parler d'éléments supplémentaires au sens précis ou d'information supplémentaire au sens large.

Table des matières

1 La question	1
2 Approche du problème sur un exemple	2
3 Individus supplémentaires	4
3.1 Individus supplémentaires en ACP	5
3.2 Individus supplémentaires en ACP inter-classes	6
3.3 Classes supplémentaires en ACP intra-classes	9
4 Équivalents en analyse des correspondances	11
5 Lignes supplémentaires en co-inertie	13
6 Information supplémentaire	15
Références	19

1 La question

La question des individus et des variables supplémentaires revient régulièrement. Elle est posée par Raphaëlle Pin :

Je souhaiterais savoir comment on peut projeter des variables supplémentaires après avoir fait une ACM, sous R. J'ai bien trouvé la fonction `supcol`, mais il semble que ça ne marche que pour les données quantitatives, donc des AFC ou des ACP. Si quelqu'un peut m'aider...

par Hélène Bonnal :

Novice dans l'utilisation d'ADE, je souhaiterais obtenir des informations sur la façon de projeter des variables supplémentaires (qui n'interviennent pas dans la détermination des axes factoriels) dans

le cas d'une ACM. Est-on obligé de calculer à la main soi-même les coordonnées des projections sur les différents axes factoriels ou existe-t-il un module qui le fait ?

plus tard :

D'autre part, j'aimerais savoir si j'obtiens bien la projection de mes variables supplémentaires à la bonne échelle en utilisant column projection, ou dois-je faire supplementary columns (j'ai du mal à comprendre ce que fais exactement cette dernière procédure) ?

par Thierry Michels :

Nouvel utilisateur de ade4 version R, je suis à la recherche d'une méthode pour projeter des variables supplémentaires (non prises en compte dans le calcul des facteurs) sur le graphique d'une acm. Amicalement,

par Marc Déconchat :

Quelqu'un peut-il me donner des conseils pour réaliser des projections de colonnes supplémentaires dans le cas d'une CCA (pour ajouter des espèces supplémentaires, pas des variables explicatives), d'une analyse inter ou intra, dans une co-inertie ? Est-ce que le principe de variable supplémentaire est possible dans un K-tableaux ?

par Luis Tito de Morais :

Si je vois bien ce que représentent les individus supplémentaires dans une acp, je n'ai pas trouvé de texte (ni dans la doc ADE, ni dans les bouquins que j'ai pu consulter, Legendre & Legendre notamment) qui explique, ou simplement discute, les conditions dans lesquelles il est légitime ou non de mettre en oeuvre cette technique.

par un utilisateur resté anonyme :

J'ai une petite question sur les lignes supplémentaires dans une AFC inter. En plus des lignes du tableau initial (qu'on obtient dans le fichier .bels), je veux projeter de vrais lignes supplémentaires dans une AFC inter. Comment faut-il faire ? Je pense qu'il faut faire la transformation initiale $(p_{ij}/p_{i.})-1$ en utilisant les marges colonnes de l'AFC inter et les marges lignes des lignes supplémentaires, mais je me demandais si il n'y avait pas un truc ou une option spéciale dans un des modules qui permette de le faire automatiquement plutôt qu'à la main ?

On réunit ici quelques éléments de réponse mis à jour pour la version ade4 dans .

2 Approche du problème sur un exemple

Pour faire une réponse utile, on peut construire la situation expérimentale suivante, à partir de l'exemple meau [3].

```
library(ade4)
data(meau)
names(meau)
```

```
[1] "env"      "design" "spe"
```

La première composante est un tableau à 24 lignes et 10 colonnes (variables de milieu) :

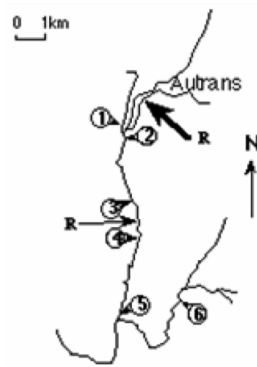
```
names(meau$env)
[1] "Temp" "Flow" "pH"   "Cond" "Oxyg" "Bdo5" "Oxyd" "Ammo" "Nitr" "Phos"
```

La seconde a 24 lignes et 2 colonnes (plan d'échantillonnage) :

```
names(meau$design)
[1] "season" "site"
```

La troisième a 24 lignes et 13 colonnes (espèces) :

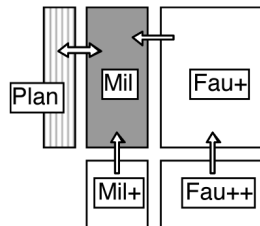
```
names(meau$spe)
[1] "Eda" "Bsp" "Brh" "Bni" "Bpu" "Cen" "Ecd" "Rhi" "Hla" "Hab" "Par" "Cae" "Eig"
```



On voit sur la carte des stations, que 5 d'entre elles sont dans un gradient amont-aval, la sixième servant de référence sur une autre rivière. C'est le cas typique de la présence d'individus supplémentaires. On sépare les deux types d'information :

```
env <- with(meau, env[design$site != "S6", ])
spe <- with(meau, spe[design$site != "S6", ])
design <- with(meau, design[design$site != "S6", ])
design$site <- as.factor(as.character(design$site)) # modalité non représentée
envsup <- with(meau, env[design$site == "S6", ])
spesup <- with(meau, spe[design$site == "S6", ])
```

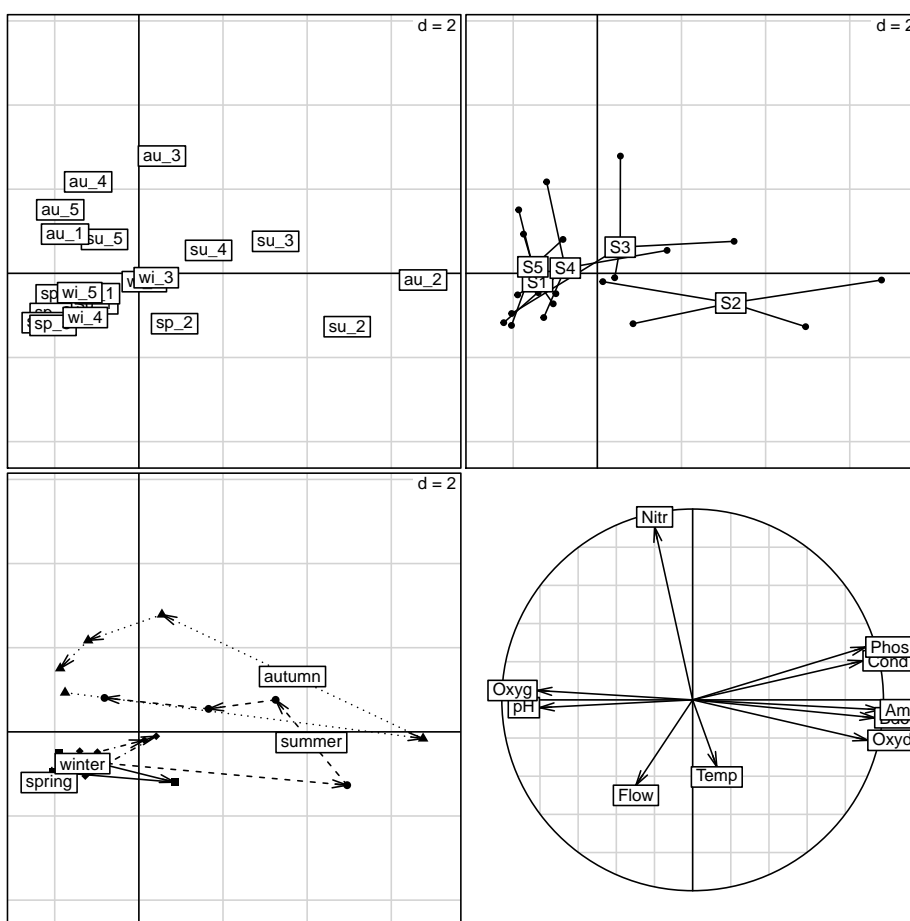
On se retrouve avec une situation fort simple qui pose la question du supplémentaire de multiples façons :



3 Individus supplémentaires

Plusieurs auteurs ont déjà souligné que le terme supplémentaire s'applique souvent de manière abusive à tout ce qui ne fait pas partie du tableau des données alors qu'on devrait bien réserver le terme projection en individus supplémentaires à une opération géométrique précise. Faisons l'ACP normée du tableau `env` de référence, en conservant 2 axes pour simplifier l'illustration.

```
env.pca <- dudi.pca(env, scan=F, nf = 3)
par(mfrow=c(2,2))
s.label(env.pca$li)
s.class(env.pca$li,design$site, cell=0)
s.traject(env.pca$li, design$season)
s.corcircle(env.pca$co)
```



La carte est tracée, puis on cherche à voir ce qu'elle ne contient pas explicitement, c'est à dire la disposition spatiale et temporelle des points. S1 à S5 sont les labels des stations (S1 est la référence amont, S2 la station polluée, S3, S4, et S5 les stations aval progressivement restaurées). Cela correspond à l'interprétation par les variables vues sur le cercle des corrélations. On a représentée sur la carte des stations de l'information extérieure. L'opération est légitime (on a une idée rapidement de la synergie spatio-temporelle) mais n'est en rien une projection en individus supplémentaires.

3.1 Individus supplémentaires en ACP

Pour représenter la station 6 on peut projeter les points correspondants sur le plan des axes principaux du nuage. Les lignes du tableau `envsup` sont des vecteurs de \mathbb{R}^{10} , l'espace dans lequel on trouve les 20 vecteurs lignes du tableau `env`. Les données dans le tableau `env` ont été normalisées, ce qui signifie qu'au vecteur

$$(x_{i1}, x_{i2}, \dots, x_{ip})$$

on a substitué le vecteur :

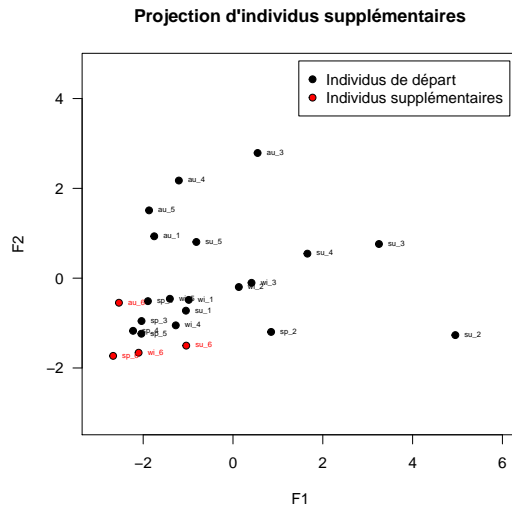
$$\left(\frac{x_{i1} - m(1)}{\sqrt{\text{var}(1)}}, \frac{x_{i2} - m(2)}{\sqrt{\text{var}(2)}}, \dots, \frac{x_{ip} - m(p)}{\sqrt{\text{var}(p)}} \right)$$

C'est le vecteur transformé qui est projeté et pour les individus supplémentaires la même transformation s'impose. Elle est implicite dans `suprow` :

```
envsup.pca <- suprow(env.pca, envsup)
envsup.pca
$tabsup
      Temp      Flow      pH      Cond      Oxyg      Bdo5      Oxyd
sp_6  0.6126890  1.03259433  0.9694584 -2.532274  0.4579486 -0.6775413 -0.8289959
su_6  0.9840156  1.08772904 -0.1938917 -1.413037 -0.1246009  0.1386289 -0.4174377
au_6 -0.6869543  0.07167512  1.3572418 -1.832750  0.6197679 -0.5455138 -1.0641721
wi_6 -0.8726177  2.42671483 -0.1938917 -1.273132  0.4579486 -0.7255514 -0.8877900
      Ammo      Nitr      Phos
sp_6 -0.7015063 -0.8724980 -0.9481148
su_6 -0.5472349 -0.5742081 -0.6928866
au_6 -0.6872219 -0.5861397 -0.7914975
wi_6 -0.7043632 -1.0216430 -0.9655167
$lisup
      Axis1      Axis2      Axis3
sp_6 -2.669469 -1.7303570  0.4537538
su_6 -1.041295 -1.5023124  0.5031495
au_6 -2.540048 -0.5472229 -0.1439222
wi_6 -2.101612 -1.6604253 -1.6333097
```

On reconnaît le tableau transformé et les coordonnées qui dérivent de la projections des lignes.

```
x <- env.pca$li[, 1]
y <- env.pca$li[, 2]
plot(x, y, asp = 1, pch = 19,
      main = "Projection d'individus supplémentaires",
      xlab = "F1", ylab = "F2", las = 1,
      xlim = c(-3, 6))
text(x, y, rownames(env.pca$li), pos = 4, cex = 0.5)
xsup <- envsup.pca$lisup[, 1]
ysup <- envsup.pca$lisup[, 2]
points(xsup, ysup, pch = 21, bg = "red")
text(xsup, ysup, rownames(envsup.pca$lisup), pos = 4, cex = 0.5,
      col = "red")
legend("topright", inset = 0.02, legend = c("Individus de départ",
      "Individus supplémentaires"), pch = c(19, 21), pt.bg = "red")
```



Cette opération est bien connue mais la question posée est beaucoup plus subtile : peut-on l'étendre en ACP inter ou intra-classes ?

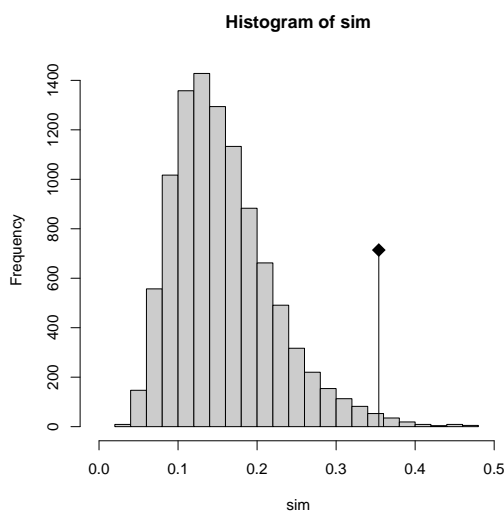
3.2 Individus supplémentaires en ACP inter-classes

Exécuter l'ACP normée inter-dates et tester sa pertinence :

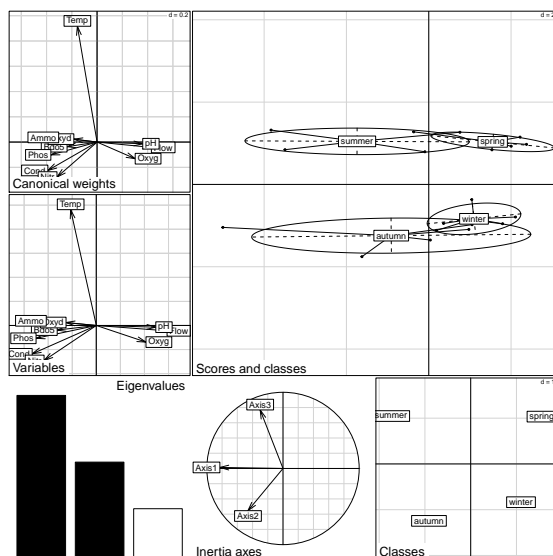
```
bet1 <- bca(env.pca, design$season, scannf=FALSE)
w <- randtest(bet1, nrepet = 9999)
w
Monte-Carlo test
Call: randtest.between(xtest = bet1, nrepet = 9999)
Observation: 0.3540029

Based on 9999 replicates
Simulated p-value: 0.0099
Alternative hypothesis: greater

Std.Obs Expectation Variance
3.102907848 0.157486146 0.004011082
plot(w,nclass=20)
```



`plot(bet1)`



Sur l'axe 1, on voit la composante saisonnière de la pollution. Sur l'axe 2, une composante saisonnière indépendante est représentée par la seule variable température. Que faire ici de la station 6 ? Répondre suppose qu'on ait d'abord identifié les opérations exécutées dans l'interclasse. Le tableau normalisé est moyenné par classe (date) ce qui donne un tableau de centre de gravité 4-10. Les 4 points de \mathbb{R}^{10} définissent un plan principal, sur lequel ils sont projetés (composante `bet1$li`). L'analyse elle-même contient une projection en individus supplémentaires des lignes du tableau normalisés (composante `bet1.ls`). Il serait donc logique de projeter en lignes supplémentaires les mêmes points que les lignes supplémentaires de l'analyse initiale. C'est là qu'il faut se méfier :

```
suprow(bet1, envsup)
$tabsup
  Temp Flow  pH Cond Oxyg Bdo5 Oxyd Ammo Nitr Phos
```

```

sp_6  11  303 8.5  245  100  1.7  0.9 0.05 2.70 0.16
su_6  13  310 8.2  285   82  8.5  1.6 0.59 3.70 0.60
au_6   4  181 8.6  270  105  2.8  0.5 0.10 3.66 0.43
wi_6   3  480 8.2  290  100  1.3  0.8 0.04 2.20 0.13
$lisup
      Axis1      Axis2
sp_6 64.75165 -71.75189
su_6 43.93421 -77.35057
au_6  4.61635 -79.63510
wi_6 124.69918 -96.48828

```

```

methods(suprow)

```

```

[1] suprow.acm* suprow.coa* suprow.dudi* suprow.mix* suprow.pca*
see '?methods' for accessing help and source code

```

Pour une projection en lignes supplémentaire, pour une AFC (coa) ou une ACP (pca) les données supplémentaires sont traitées comme les données initiales. Sinon les lignes supplémentaires ne sont pas modifiées et projetées dans leur état brut. Manifestement le coup n'était pas prévu. Il est donc nécessaire d'avoir fait l'opération de normalisation préalable à l'interclasse au préalable, donc d'utiliser :

```

suprow(bet1,envsup.pca$tabsup)
$tabsup
      Temp      Flow      pH      Cond      Oxyg      Bdo5      Oxyd
sp_6 0.6126890 1.03259433 0.9694584 -2.532274 0.4579486 -0.6775413 -0.8289959
su_6 0.9840156 1.08772904 -0.1938917 -1.413037 -0.1246009 0.1386289 -0.4174377
au_6 -0.6869543 0.07167512 1.3572418 -1.832750 0.6197679 -0.5455138 -1.0641721
wi_6 -0.8726177 2.42671483 -0.1938917 -1.273132 0.4579486 -0.7255514 -0.8877900
      Ammo      Nitr      Phos
sp_6 -0.7015063 -0.8724980 -0.9481148
su_6 -0.5472349 -0.5742081 -0.6928866
au_6 -0.6872219 -0.5861397 -0.7914975
wi_6 -0.7043632 -1.0216430 -0.9655167
$lisup
      Axis1      Axis2
sp_6 2.974279 1.3303478
su_6 1.403085 1.3863951
au_6 2.539760 -0.1087551
wi_6 2.970903 -0.3136127

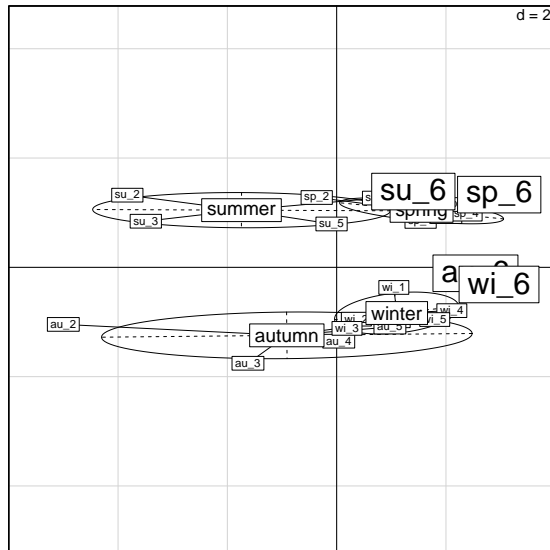
```

Sur l'analyse de type between on peut alors projeter des vecteurs lignes modifiés par ailleurs :

```

s.class(bet1$ls, design$season, clab = 0, xlim = c(-6, 4))
s.label(bet1$ls, clab = 0.75, add.p = T)
s.label(bet1$li, clab = 1.25, add.p = T)
s.label(suprow(bet1, envsup.pca$tabsup)$lisup, clab=2,add.p=T,cpoi=2)

```

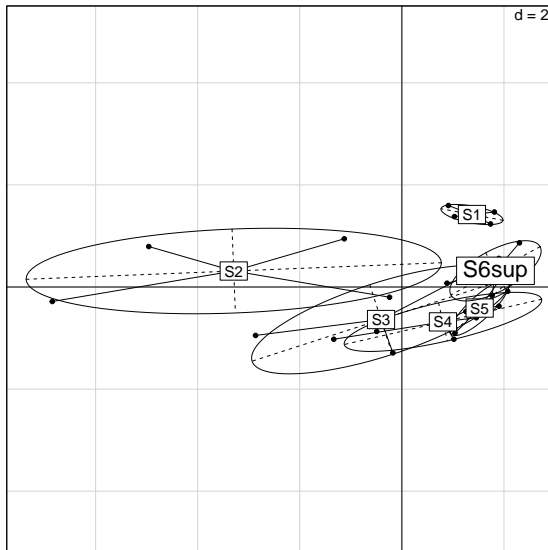



On voit la superposition des centres de gravité qui définissent les axes, des points qui définissent les centres de gravité et d'un point supplémentaire par classe qui, sans être aberrant est en marge de la classe à laquelle il aurait pu appartenir.

3.3 Classes supplémentaires en ACP intra-classes

Continuons la réflexion ouverte par la question posée. Nous venons de faire une ACP inter-dates et la station supplémentaire ajoutait un point supplémentaire dans chaque classe. Mais si nous faisons l'analyse inter-stations la station supplémentaire ajoutera une classe. Pour savoir si la projection en individus supplémentaires est possible et valide, là encore il convient de revenir à la procédure interne. Pour l'inter-classe, il n'y a pas de difficulté nouvelle :

```
bet2 <- bca(env.pca, design$site, scannf = FALSE)
s.class(bet2$ls, design$site)
lisup <- suprow(bet2, envsup.pca$stabsup)$lisup
facsup <- as.factor(rep("S6sup", 4))
s.class(lisup, facsup, add.p=T, clab=1.5)
```

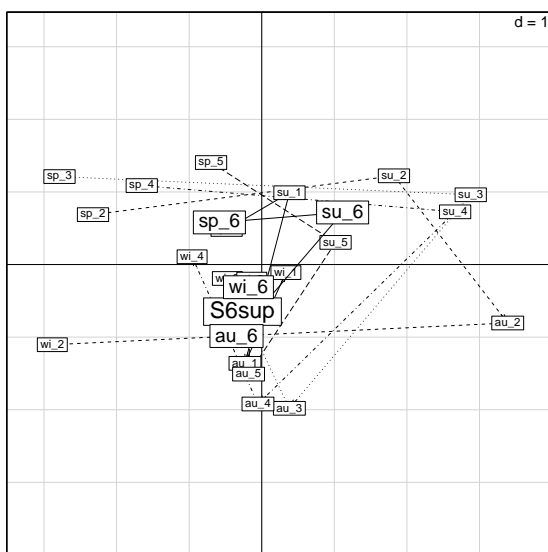


Associer la partition par stations et exécuter l'intra-classes stations :

```
wit1 <- wca(env.pca, design$site, scannf = FALSE)
```

Dans l'intra-classe, par contre, le tableau normalisé est à nouveau centré par classes et le nouveau nuage de 20 points de \mathbb{R}^{10} définit des axes principaux sur lequel sont projetés les lignes recentrées (`wit1$li`) mais aussi les lignes du tableau normalisé (`wit1$ls`). Donc, il est concevable de projeter les lignes supplémentaires de l'ACP initiale, ou le nuage recentré si on veut compléter la carte des valeurs recentrées (`wit1$li`) :

```
s.traject(wit1$li, design$site, clab = 0)
s.label(wit1$li, clab=0.75, add.plot = T)
w1 <- scalewt(envsup.pca$tabsup, scale=F) # tableau recentré
w1sup <- suprow(wit1, w1)$lisup
s.traject(w1sup, add.p=T, clab=1.5, lab="S6sup")
s.label(w1sup, clab=1.25, add.p=T)
```



On retiendra donc qu'en inter et intra-classes, la projection des individus supplémentaires est aisée. Seul un minimum d'attention est requis pour associer les nuages de lignes actifs et passifs de manière à ce que les superpositions associent des tableaux ayant subi des traitements cohérents.

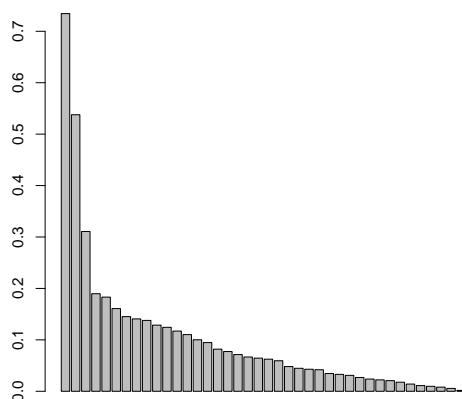
4 Équivalents en analyse des correspondances

Pour illustrer le fonctionnement de pratiques identiques en AFC, utilisons l'objet `avimedi`. Ce tableau faunistique [1] a 302 lignes et 60 colonnes. Séparer le tableau en deux parties :

```
data(avimedi)
names(avimedi)
[1] "fau" "plan" "nomesp"
summary(avimedi$plan)
reg      str
Pr:178   S1:40
Co:124   S2:40
         S3:92
         S4:46
         S5:40
         S6:44
Prov <- avimedi$fau[avimedi$plan$reg=="Pr",]
Cors <- avimedi$fau[avimedi$plan$reg=="Co",]
```

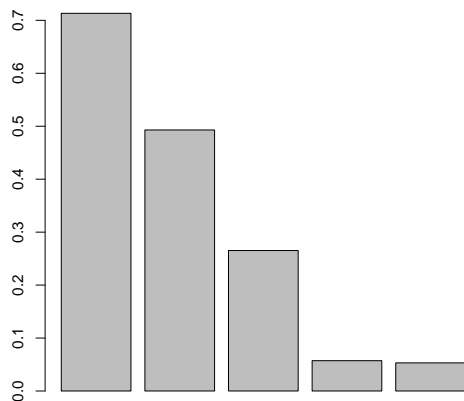
On veut projeter en individus supplémentaires les relevés du second tableau sur l'inter-classe du premier en analyse des correspondances. Faire l'analyse initiale :

```
afc1 <- dudi.coa(Prov, scannf = FALSE, nf = 3)
barplot(afc1$eig)
```

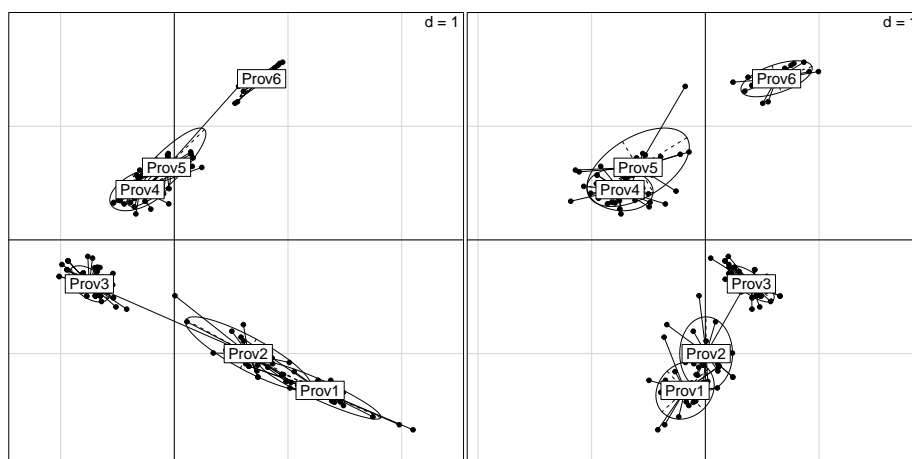


Coupler avec la variable de stratification (architecture de la végétation) pour exécuter l'analyse inter-classe :

```
straProv <- avimedi$plan$str[avimedi$plan$reg=="Pr"]
levels(straProv) <- paste("Prov", 1:6, sep="")
straCors <- avimedi$plan$str[avimedi$plan$reg=="Co"]
levels(straCors) <- paste("Cors", 1:6, sep="")
afcbet <- bca(afc1, straProv, scannf = FALSE, nf = 3)
barplot(afcbet$eig)
```

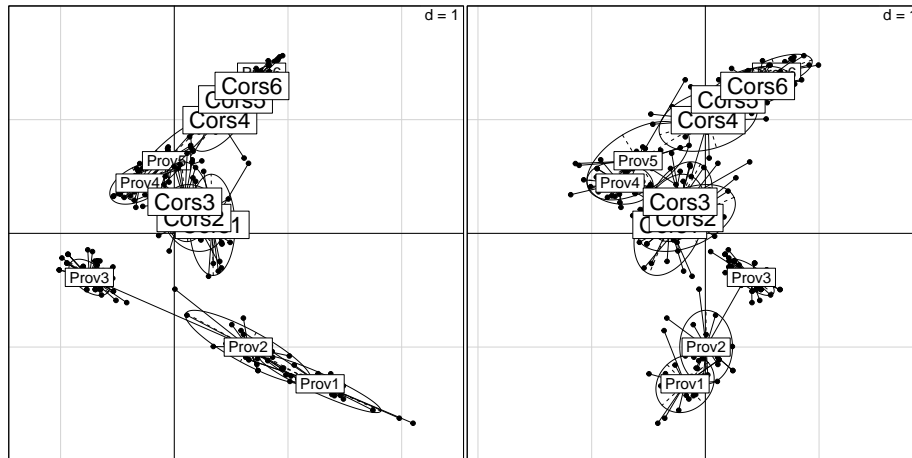


```
par(mfrow = c(1,2))
s.class(afcbet$ls,straProv,xax=2,yax=1,ylim=c(-2,2))
s.class(afcbet$ls,straProv,xax=3,yax=1,ylim=c(-2,2))
```



Projeter en individus supplémentaires les lignes de l'autre tableau :

```
afcsup <- suprow(afc1,Cors)
betsup <- suprow(afcbet,afcsup$tabsup)
par(mfrow = c(1,2))
s.class(afcbet$ls,straProv,xax=2,yax=1,ylim=c(-2,2))
s.class(betsup$lisup,straCors,xax=2,yax=1,add.p=T,clab=1.5)
s.class(afcbet$ls,straProv,xax=3,yax=1,ylim=c(-2,2))
s.class(betsup$lisup,straCors,xax=3,yax=1,add.p=T,clab=1.5)
```



On refait ainsi les cartes inter-strates des relevés de Corse en individus supplémentaires sur la carte inter-strates des relevés de Provence. L'opération est fort significative au plan biologique. La convergence des cortèges faunistiques en milieu forestier [2] positionne les relevés forestiers des deux zones au même endroit.

Mais, en n'utilisant que les espèces des milieux ouverts de Provence d'une part et en intégrant les différences de structure des communautés en milieu insulaire [1] d'autre part, les milieux ouverts de Corse ne prennent pas la place qu'on attend. On peut donc retenir de cet exemple que la représentation d'individus supplémentaires en inter-classe après une ACP (centrée ou normée) et une AFC inter-classe ne pose pas de problème.

5 Lignes supplémentaires en co-inertie

Profitons de l'occasion pour étendre la notion d'individus supplémentaires à l'analyse de co-inertie. On repart des données de `meau`. Préparons le tableau des individus supplémentaires de cette analyse :

```
env.pca.sup <- suprow(env.pca,envsup)
```

Faisons l'ACP centrée du second tableau :

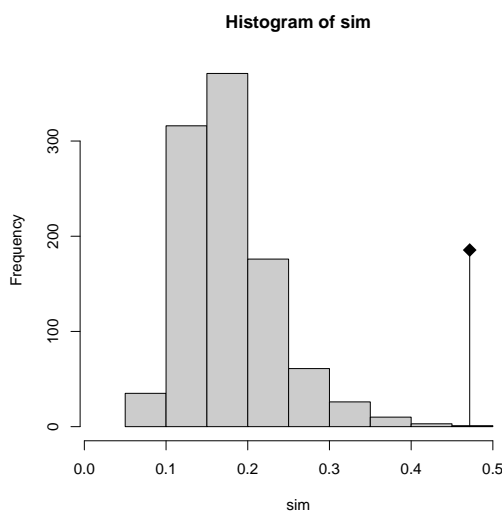
```
spe.pca <- dudi.pca(spe,scannf=F, scale=FALSE)
```

Préparons le tableau des individus supplémentaires associés :

```
spe.pca.sup <- suprow(spe.pca,spesup)
```

Exécutons l'analyse de co-inertie et le test associé :

```
coi1 <- coinertia(env.pca,spe.pca,scannf=F)
coi1.randtest <- randtest(coi1)
plot(coi1.randtest)
```



Pour les individus supplémentaires, il faut s'attendre à une petite difficulté. Le tableau croisé est décrit dans l'objet `coi1` :

```

coi1
Coinertia analysis
call: coinertia(dudiX = env.pca, dudiY = spe.pca, scannf = F)
class: coinertia dudi
$rank (rank)      : 10
$nf (axis saved) : 2
$RV (RV coeff)   : 0.4716674

eigenvalues: 82.6 9.547 4.919 2.544 1 ...

  vector length mode  content
1 $eig    10      numeric Eigenvalues
2 $lw    13      numeric Row weights (for spe.pca cols)
3 $cw    10      numeric Col weights (for env.pca cols)

  data.frame nrow ncol content
1 $tab     13    10  Crossed Table (CT): cols(spe.pca) x cols(env.pca)
2 $li     13     2   CT row scores (cols of spe.pca)
3 $l1     13     2   Principal components (loadings for spe.pca cols)
4 $co     10     2   CT col scores (cols of env.pca)
5 $c1     10     2   Principal axes (loadings for env.pca)
6 $lX     20     2   Row scores (rows of env.pca cols)
7 $mX     20     2   Normed row scores (rows of env.pca)
8 $lY     20     2   Row scores (rows of spe.pca)
9 $mY     20     2   Normed row scores (rows of spe.pca)
10 $aX     3     2   Corr env.pca axes / coinertia axes
11 $aY     2     2   Corr spe.pca axes / coinertia axes

CT rows = cols of spe.pca (13) / CT cols = cols of env.pca (10)

```

Le tableau croisé est du type Y^tDX . Les lignes de ce tableau sont les colonnes de Y , donc les espèces. Les colonnes de ce tableau sont celle de X , donc les variables. Les relevés supplémentaires associés à X sont des lignes supplémentaires pour ce tableau et les opérations sont simples :

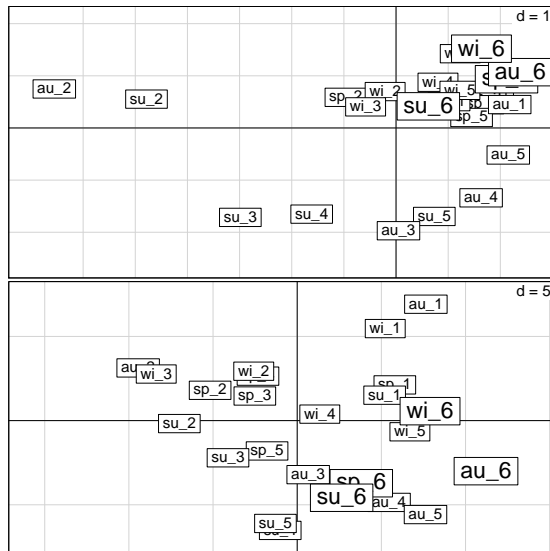
```
env1 <- suprow(coi1,env.pca.sup$tabsup)
```

Par contre les relevés supplémentaires du tableau faunistique sont apparentés aux colonnes du tableau croisé : ce sont des colonnes supplémentaires. Il faut donc transposer au préalable :

```
spe1 <- supcol(coi1,t(spe.pca.sup$tabsup))
```

On peut alors représenter les relevés supplémentaires dans les deux espaces de référence. Pour les relevés de milieu et pour la station supplémentaires, puis de même dans le tableau faunistique et pour la station supplémentaire :

```
par(mfrow=c(2,1))
s.label(coi1$IX)
s.label(env1$liisup,add.p=T,clab=1.5)
s.label(coi1$IY)
s.label(spe1$cosup,add.p=T,clab=1.5)
```



L'écart entre les deux systèmes est sensible. Sur le premier jeu de variables (milieu) la station 6 prend position à proximité de la station 1 (témoin non pollué). Sur le deuxième jeu de variables (faune) elle se place au delà de la station 5 (restaurée). La position dans le gradient amont-aval est donc un élément supplémentaire à prendre en compte.

De cet exercice, on retiendra que la notion de projections d'individus supplémentaires est toujours en jeu dès qu'il y a projection d'une part et individus supplémentaires d'autre part. Une double analyse d'inertie est cependant la condition impérative qui rend possible l'opération. L'inter-classe (analyse d'inertie du nuage des centre de gravités `bca()`), la co-inertie (analyse d'inertie du tableau croisé `coinertia`), STATIS sur les X (analyse d'inertie d'un compromis de tableau `pta`) et même STATICO (analyse d'inertie d'un opérateur de co-inertie compromis) tolèrent les individus supplémentaires.

Après une analyse d'inertie on peut toujours projeter sur les sous-espaces propres (engendrés par les axes ou les composantes principales) tout vecteur de cet espace. La projection est toujours possible mais elle ne prend de sens qu'en comparaison d'autres projections. Ce qui est comparé doit donc être comparable, c'est la seule contrainte.

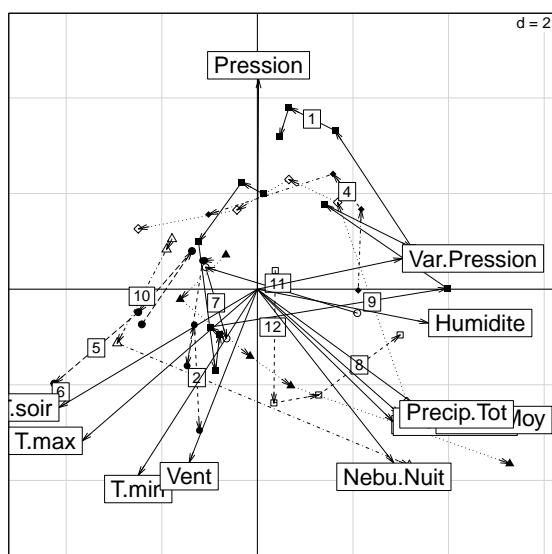
6 Information supplémentaire

Il n'est pas toujours pertinent de privilégier la projection en individus supplémentaire au sens euclidien par rapport à l'expression directe par un procédé

quelconque d'information supplémentaire. C'est vrai pour les variables qualitatives et les variables faunistiques (abondance taxonomique). En effet, l'averaging (représentation des moyennes) est en général plus explicite et plus simple à expliquer.

Prenons par exemple, le jeu de données `trichometeo` [4]. Les variables météorologiques (11) donnent une ACP de dépouillement facile :

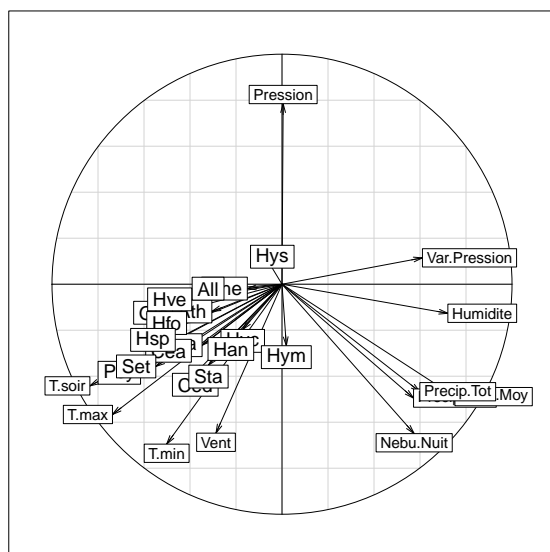
```
data(trichometeo)
pca1 <- dudi.pca(trichometeo$meteo, scannf=FALSE)
s.traject(pca1$li, trichometeo$cla)
s.arrow(10*pca1$c1, add.p=T, clab=1.5)
```



On note la succession haute pression (beau temps) puis fortes températures puis précipitations (orages d'été) caractéristiques du temps estival de la région. La question porte sur l'influence des variables météorologique sur l'abondance des piégeages lumineux. Le tableau faunistique a 17 espèces (variables).

On peut chercher à représenter l'information sur l'abondance des taxons dans le cadre de ce plan factoriel très explicite. Mais la notion d'objets supplémentaires porte cette fois sur les variables. On ne peut faire de projections supplémentaires que dans l'espace où se fait une projection réelle, ici \mathbb{R}^2 . Comme on peut projeter ce qu'on veut, le soin est laissé à l'utilisateur de choisir. En projetant des variables normalisées (de moyenne nulle et de variance unité), on se trouve dans la situation exacte de l'analyse de départ (ACP normée). Projeter des variables normalisées est compatible avec un cercle de corrélation et permet d'ajouter les variables faunistiques :

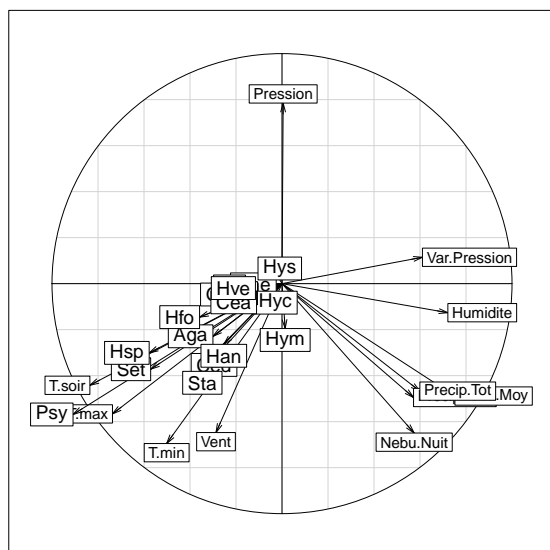
```
w1 <- supcol(pca1, scale(log(trichometeo$fau+1)))$cosup
s.corcircle(pca1$co)
s.arrow(w1, add.p=T, clab=1.25)
```

Les projections des variables supplémentaires normalisées (vecteurs de norme 1) donnent des coordonnées qui sont des coefficients de corrélation avec les coordonnées factorielles. Ces corrélations sont presque toutes de même signe et indiquent clairement la période météorologique qui favorise l'émergence des larves de toutes les espèces.

On peut aussi centrer le tableau faunistique et projeter ces colonnes. Les coordonnées des projections sont des covariances et l'image intègre la variabilité des abondances. La présence du cercle unité sert alors de repère mais la longueur des vecteurs projetés n'étant plus égale à 1, les points ne tombent pas forcément dans ce cercle.

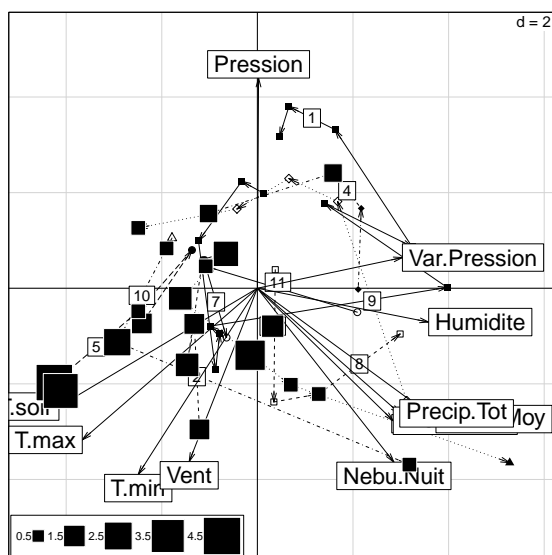
```
w2 <- supcol(pca1,scale(log(trichometeo$fau+1), scale=FALSE))$cosup
s.corcircle(pca1$co)
s.arrow(w2, add.p=T,clab=1.25)
```



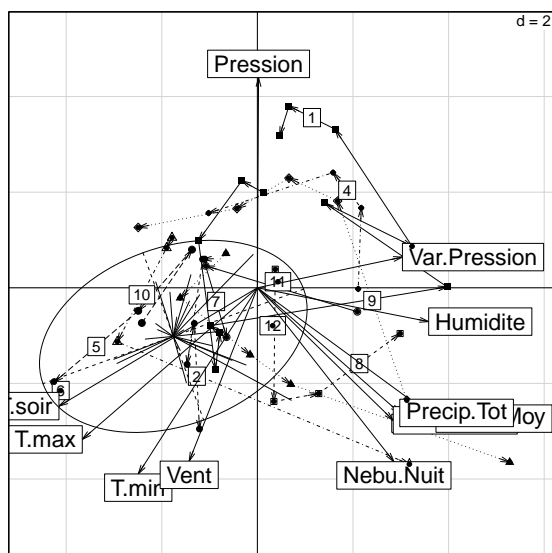
L'abondance d'un taxon peut être considérée comme une distribution de fréquence sur l'ensemble des points du plan factoriel. On passe de la représentation

de variables supplémentaire à celle d'information supplémentaire :

```
w3 <- log(trichometeo$fau + 1)
s.traject(pca1$li, trichometeo$cla)
s.arrow(10*pca1$c1,add.p=T,clab=1.5)
s.value (pca1$li, w3$Set,add.p=T)
```

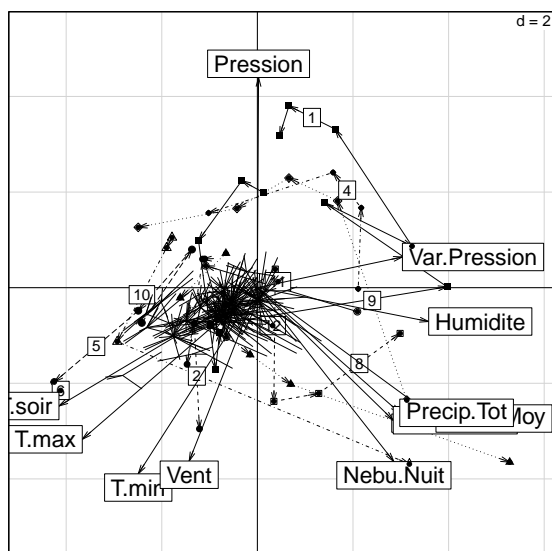


```
s.traject(pca1$li, trichometeo$cla)
s.arrow(10*pca1$c1,add.p=T,clab=1.5)
s.distri (pca1$li, w3$Set,add.p=T, cstar=0.5)
```



L'expression des données sur la variable supplémentaire dans le plan des individus de l'analyse de départ est le dual de la projection de la variable supplémentaire dans le plan des variables de l'analyse de départ. Il n'y a évidemment aucune contradiction entre ces points de vue, simplement des moyens de s'exprimer.

```
s.traject(pca1$li, trichometeo$cla)
s.arrow(10*pca1$c1, add.p=T, clab=1.5)
s.distri (pca1$li, w3, cstar=0.25, cell=0, add.p=T)
```



On pourra aussi superposer les moyennes des positions des espèces, représenter l'abondance des espèces sur les plans de co-inertie, ... Ici domine l'idée d'une combinaison de variables météorologiques ayant une influence commune sur les émergences de tous les taxons. Notons enfin qu'il arrive souvent que de véritables projections euclidiennes soient également des représentation par moyennes de distribution et que les notions d'individus supplémentaires et d'information supplémentaire se confondent.

Quoiqu'il en soit la graphique appliquée à la statistique multidimensionnelle est un moyen d'expression. Cela suppose quelques libertés dans les choix et la référence à un comportement *conforme à la règle* peut être le signe d'une certaine absence d'imagination. Ce n'est évidemment pas une raison pour faire n'importe quoi.

Références

- [1] J. Blondel, D. Chessel, and B. Frochot. Bird species impoverishment, niche expansion, and density inflation in mediterranean island habitats. *Ecology*, 69 :1899–1917, 1988.
- [2] J. Blondel and H. Farré. The convergent trajectories of bird communities along ecological successions in european forests. *Ecologia (Berlin)*, 75 :83–93, 1988.
- [3] D. Pegaz-Maucet. *Impact d'une perturbation d'origine organique sur la dérive des macro-invertébrés benthiques d'un cours d'eau. Comparaison avec le benthos*. PhD thesis, University of Lyon 1, 1980.
- [4] P. Usseglio-Polatera and Y. Auda. Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie*, 23 :65–79, 1987.