

# Analyses préliminaires JD La Réunion

---

## Contents

<b>Objectif</b>	<b>2</b>
<b>Points principaux</b>	<b>2</b>
<b>Données</b>	<b>2</b>
table oiseaux . . . . .	2
table habitats . . . . .	3
Analyse en coordonnées principales sur la matrice d'habitats . . . . .	8
table Oiseaux . . . . .	13
table oiseaux : restreint aux espèces présentes sur >5% des points . . .	19
<b>AFC sans les espèces vues sur moins de 5% des points</b>	<b>20</b>
<b>Analyse fonctionnelle</b>	<b>22</b>
base optimale . . . . .	23
Analyse fonctionnelle . . . . .	24
ACP fonctionnelle . . . . .	26
<b>Conclusions</b>	<b>32</b>

## Objectif

Analyses préliminaires des données de la Réunion. Pour l'instant il s'agit surtout de comprendre comment elles sont structurées et ce qu'on peut en faire. Le document commence par des analyses multivariées sur les tables oiseaux et habitats, puis une FDA sur les variations d'abondance le long du gradient altitudinal.

## Points principaux

- deux analyses : une statique (PCOA sur données d'habitat et données oiseaux) et une par gradient (FDA sur les réponses des espèces à l'altitude)
- des résultats qui ne sont pas aberrants mais nécessitent un regard extérieur, je ne me sens pas à les analyser seul au regard de ma connaissance de l'avifaune de la Réunion. En tout cas on n'a aucune typologie de réponse claire et des résultats très multivariés.
- il faut qu'on discute de ce qu'on fait sur cette base : quelle approche on privilégie et quelles analyses supplémentaires

## Données

Deux tables de données :

```
ois <- read.table("data/OIS_modif29042020.txt", header = T, sep = "\t", row.names = 1)
mil <- read.table("data/MIL_modif29042020.txt", header = T, sep = "\t", row.names = 1)
```

Les colonnes sont décrites dans les fichiers excel initiaux.

### table oiseaux

```
dim(ois)
```

```
## [1] 410 23
```

```
summary(ois)
```

```
##          ACTR          CIMA          COCH          COCO          COFR
## Min.      : 0.000    Min.      :0.00000    Min.      :0.000000    Min.      :0.0000    Min.      : 0.0000
## 1st Qu.: 0.000    1st Qu.:0.00000    1st Qu.:0.000000    1st Qu.:0.0000    1st Qu.: 0.0000
## Median : 1.000    Median :0.00000    Median :0.000000    Median :0.0000    Median : 0.0000
## Mean      : 1.768    Mean      :0.02683    Mean      :0.004878    Mean      :0.3122    Mean      : 0.3829
## 3rd Qu.: 2.000    3rd Qu.:0.00000    3rd Qu.:0.000000    3rd Qu.:0.0000    3rd Qu.: 0.0000
## Max.      :34.000    Max.      :2.00000    Max.      :2.000000    Max.      :8.0000    Max.      :10.0000
##          GEST          HYBO          LOPU          MAMA          PADO
## Min.      : 0.0000    Min.      :0.0000    Min.      :0.000000    Min.      :0.00000    Min.      : 0.000
## 1st Qu.: 0.0000    1st Qu.:0.0000    1st Qu.:0.000000    1st Qu.:0.00000    1st Qu.: 0.000
## Median : 0.0000    Median :0.0000    Median :0.000000    Median :0.00000    Median : 0.000
## Mean      : 0.8049    Mean      :0.3488    Mean      :0.002439    Mean      :0.02683    Mean      : 1.066
## 3rd Qu.: 1.0000    3rd Qu.:0.0000    3rd Qu.:0.000000    3rd Qu.:0.00000    3rd Qu.: 0.000
## Max.      :31.0000    Max.      :6.0000    Max.      :1.000000    Max.      :5.00000    Max.      :17.000
##          PYJO          SATE          STPI          TEBO          TUNI
## Min.      : 0.0000    Min.      : 0.00    Min.      :0.0000    Min.      :0.00000    Min.      :0.0000    M
## 1st Qu.: 0.0000    1st Qu.: 0.00    1st Qu.:0.0000    1st Qu.:0.00000    1st Qu.:0.0000    1s
## Median : 0.0000    Median : 1.00    Median :0.0000    Median :0.00000    Median :0.0000    Me
## Mean      : 0.5561    Mean      : 1.59    Mean      :0.1902    Mean      :0.06829    Mean      :0.1878    Me
## 3rd Qu.: 0.0000    3rd Qu.: 3.00    3rd Qu.:0.0000    3rd Qu.:0.00000    3rd Qu.:0.0000    3r
## Max.      :10.0000    Max.      :10.00    Max.      :7.0000    Max.      :3.00000    Max.      :5.0000    Ma
```

Pas d'incohérence apparente. Beaucoup de 0, les distributions sont très asymétriques, sans surprise.

## table habitats

```
dim(mil)
```

```
## [1] 410 16
```

```
summary(mil)
```

```
##          ALTI          ACDE          ACMI          BOCO          BRAN
## Min.      : 20.0    Min.      :0.0000    Min.      :0.000000    Min.      :0.0000    Min.      :0.0000    M
## 1st Qu.: 803.8    1st Qu.:0.0000    1st Qu.:0.000000    1st Qu.:0.0000    1st Qu.:0.0000    1
## Median :1290.0    Median :0.0000    Median :0.000000    Median :0.0000    Median :0.0000    M
## Mean      :1204.2    Mean      :0.1243    Mean      :0.003585    Mean      :0.1222    Mean      :0.0841    M
## 3rd Qu.:1598.8    3rd Qu.:0.0575    3rd Qu.:0.000000    3rd Qu.:0.0000    3rd Qu.:0.0000    3
## Max.      :2880.0    Max.      :1.0000    Max.      :0.700000    Max.      :1.0000    Max.      :1.0000    M
##          FOEX          FRAR          FRBA          PATU          SAAR
```

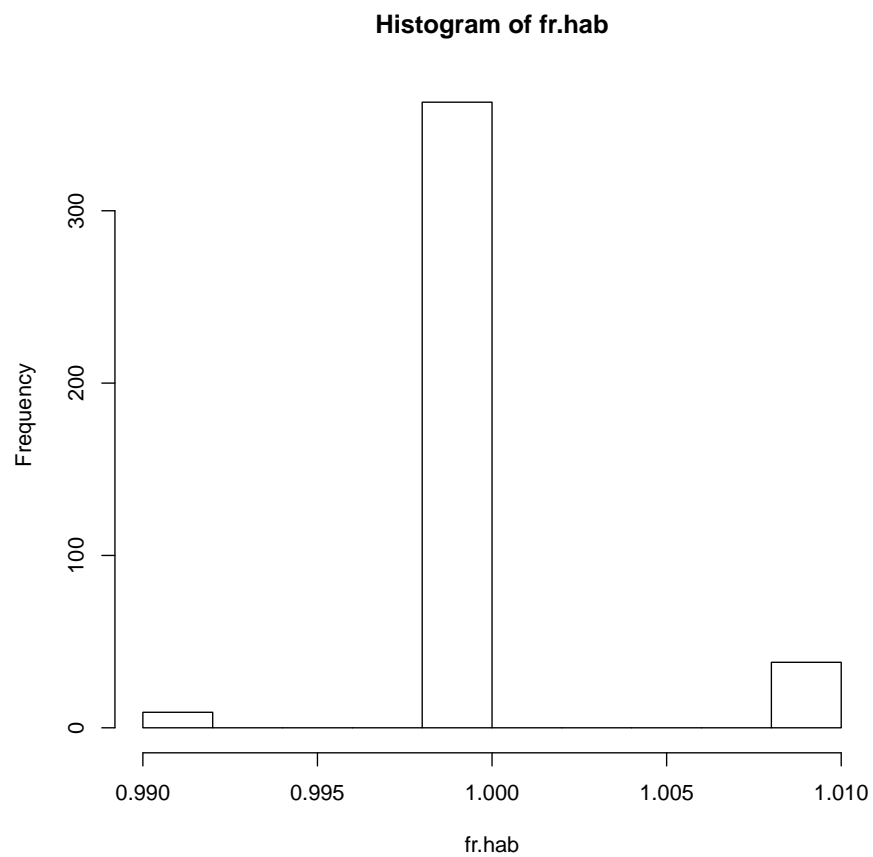
```
## Min.      :0.00000 Min.      :0.000000 Min.      :0.0000 Min.      :0.00000 Min.      :0.00000
## 1st Qu.:0.00000 1st Qu.:0.000000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.00000
## Median :0.00000 Median :0.000000 Median :0.0000 Median :0.00000 Median :0.00000
## Mean    :0.06646 Mean    :0.003756 Mean    :0.1195 Mean    :0.05895 Mean    :0.03398
## 3rd Qu.:0.06000 3rd Qu.:0.000000 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.00000
## Max.     :0.72000 Max.     :0.390000 Max.     :1.0000 Max.     :1.00000 Max.     :1.00000
```

Attention : il y a des fréquences cumulées d'habitats par point >100% !

```
# fréquence cumulée des habitats par point
fr.hab=apply(mil[,-1],1,"sum")
summary(fr.hab)
```

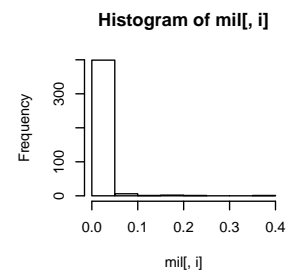
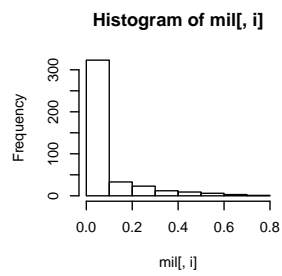
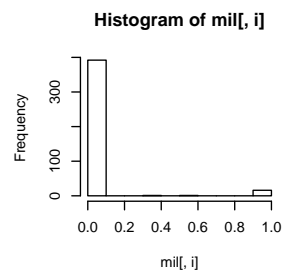
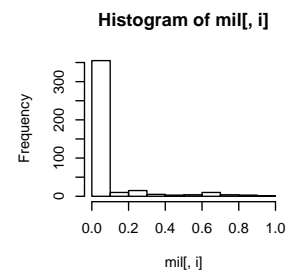
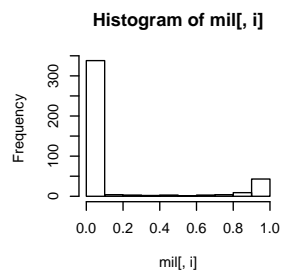
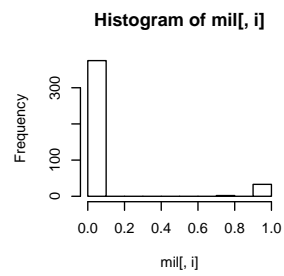
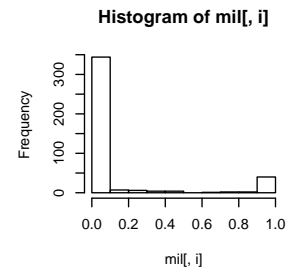
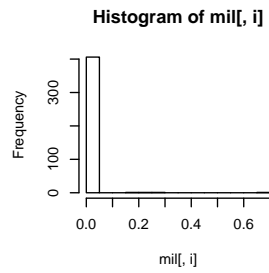
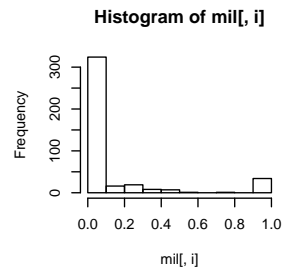
```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    0.990   1.000   1.000   1.001   1.000   1.010
```

```
hist(fr.hab)
```

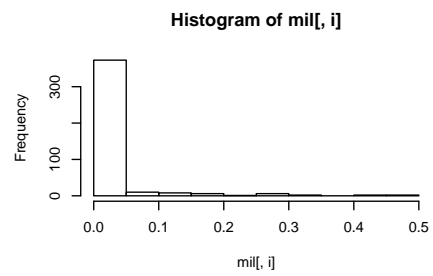
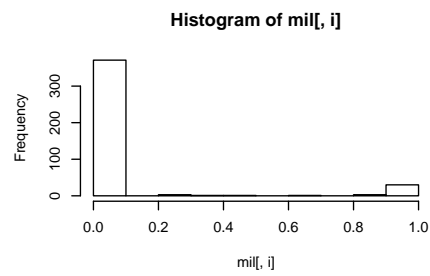
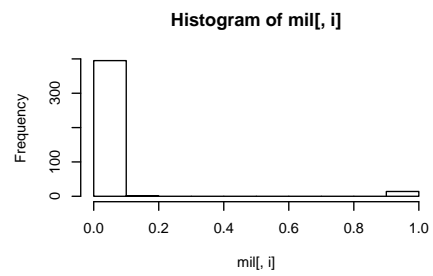
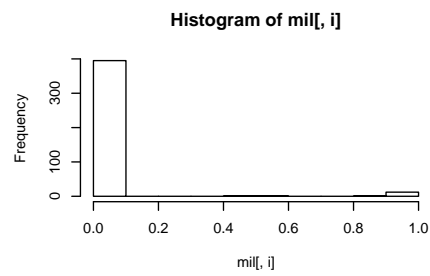
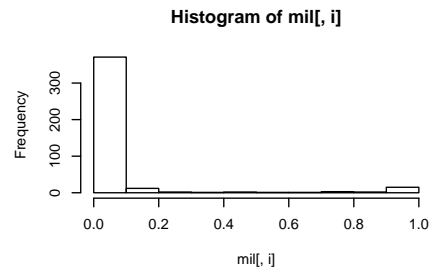
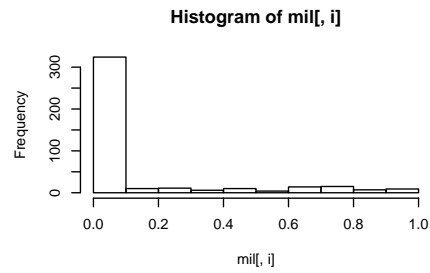


### Distribution des habitats

```
par(mfrow=c(3,3))
for(i in 2:10){
  hist(mil[,i])
}
```



```
par(mfrow=c(3,2))
for(i in 11:16){
  hist(mil[,i])
}
```



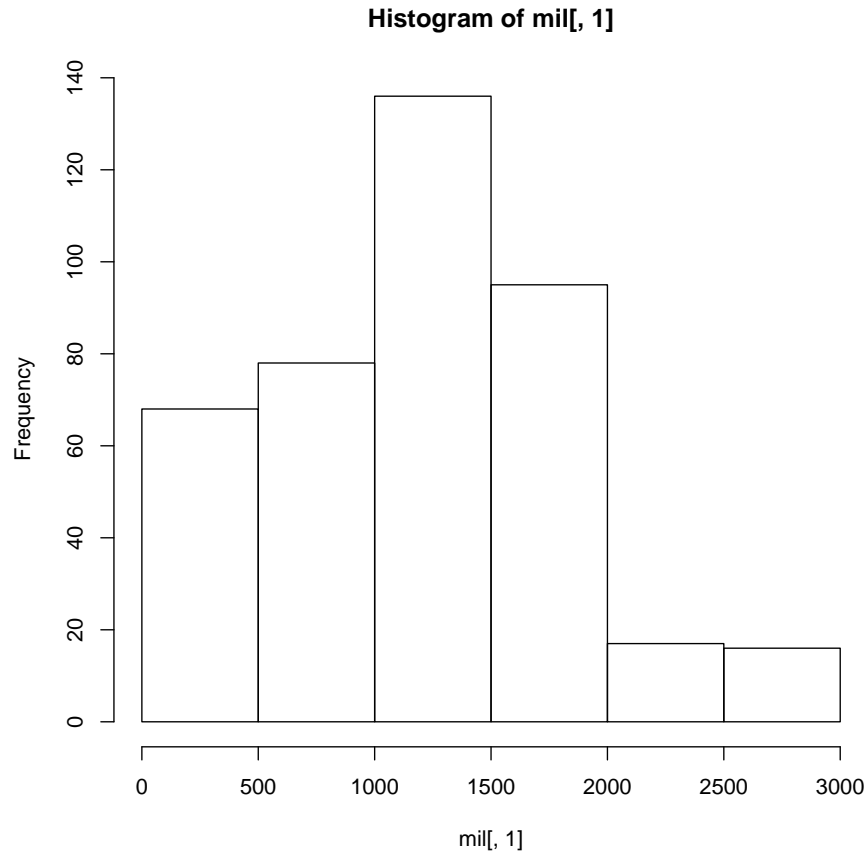
Les distris des habitats posent deux problèmes :

- la somme à 100 (en théorie)
- leur asymétrie

## Altitude

Pour l'altitude, variation assez classique :

```
hist(mil[,1])
```



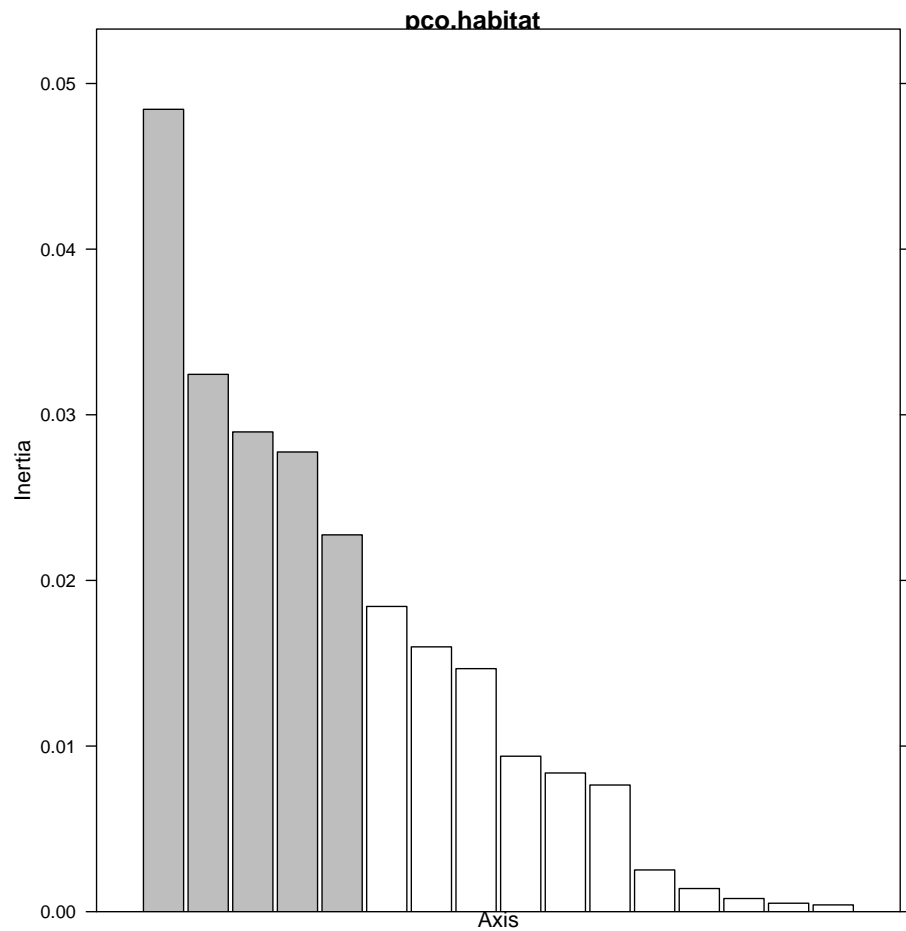
## Analyse en coordonnées principales sur la matrice d'habitats

```
# PCOA  
pco.habitat <- dudi.pco(distmat, scannf = F, nf = 5) # avec 5 axes environ 70% de variation
```

L'analyse en coordonnées principale permet de résoudre les deux problèmes mentionnés ci dessus et d'ajouter l'altitude aux variables d'habitat.

```
screeplot(pco.habitat)
```





```
100*cumsum(pco.habitat$eig)/sum(pco.habitat$eig)
```

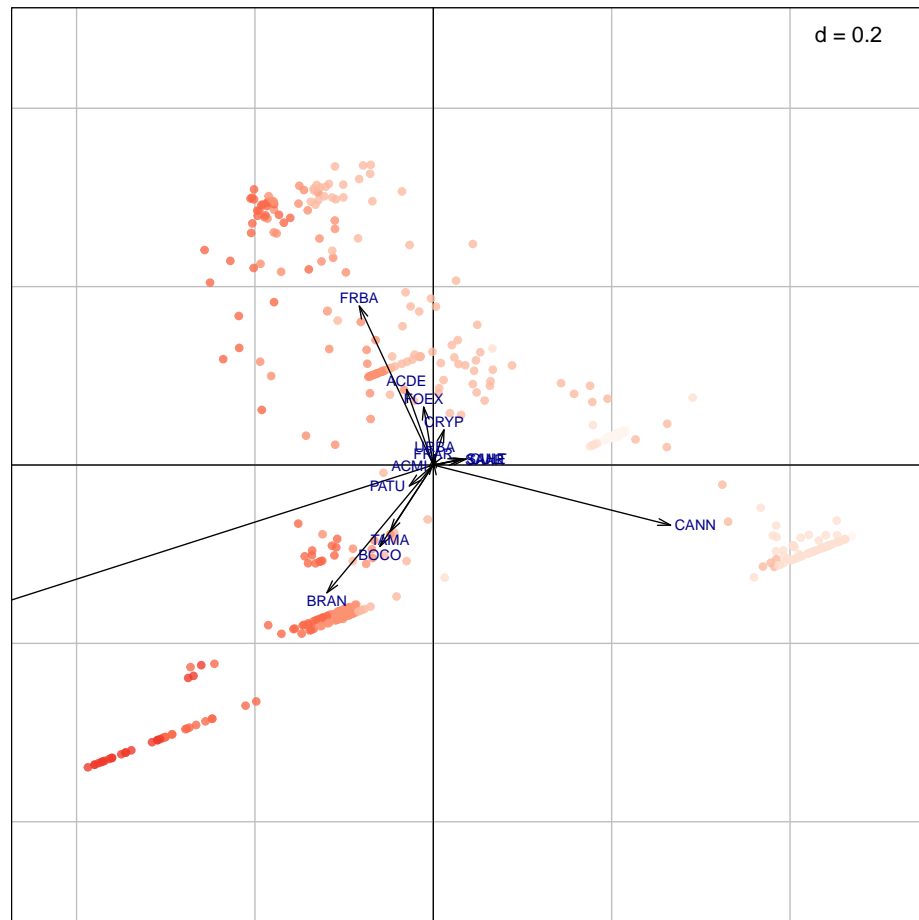
```
## [1] 20.14260 33.63155 45.67581 57.21639 66.67637 74.33942 80.98807 87.08968 90.
## [15] 99.82905 100.00000
```

On est sur des % de variance expliquée assez classiques pour ce genre de données. En première intention je retiens 5 composantes.

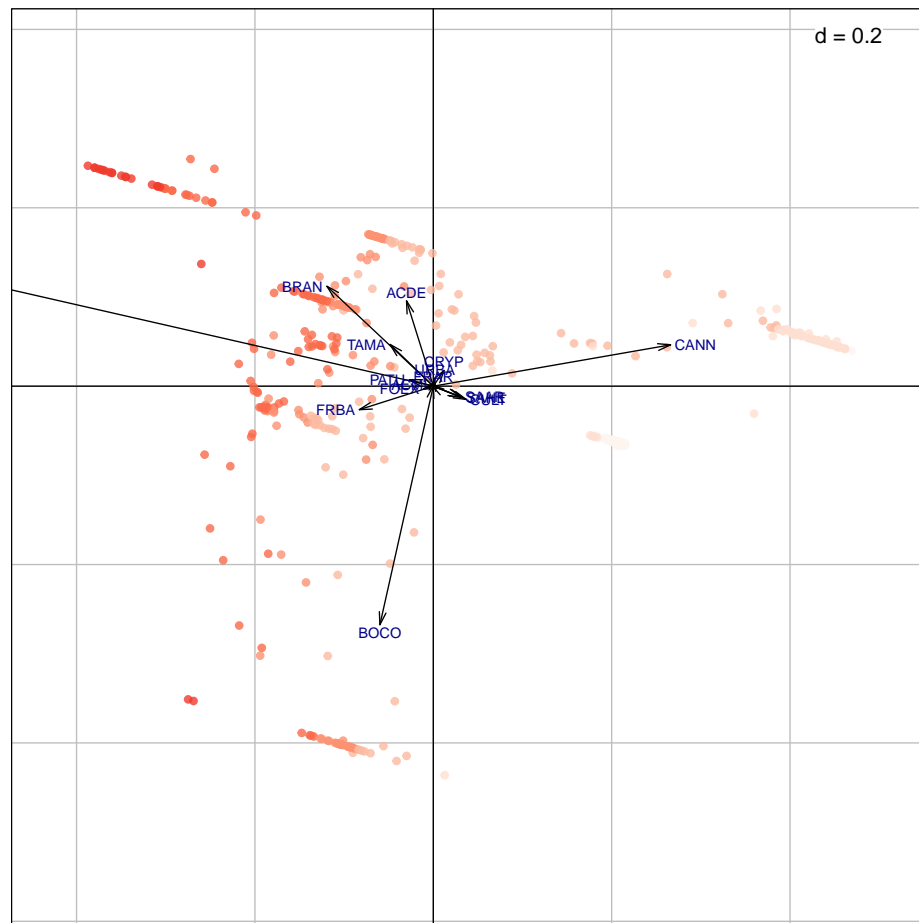
**Projection des points et des variables (axe 1 vs chacun des 4 autres)**

```
# axes 1 vs 2
s.label(pco.habitat$li,ppoints=list(alpha=0.8,col=c2),plabels=list(alpha=0,boxes=list(draw=F
```

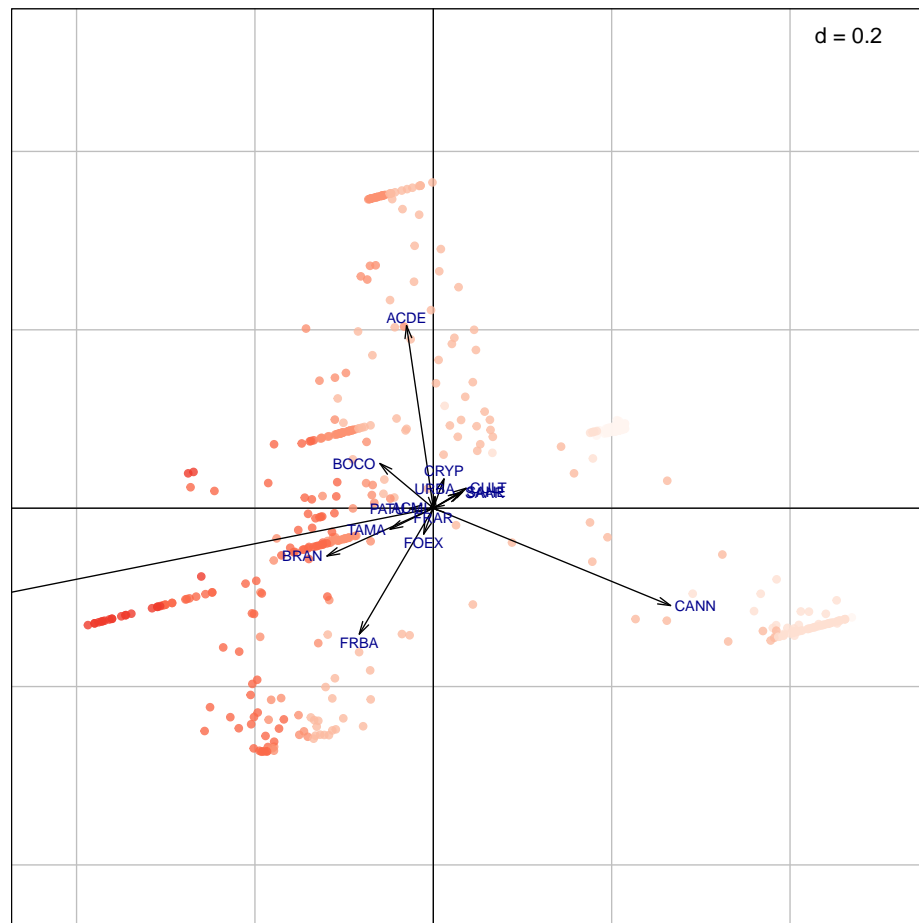
```
sp=supcol(pco.habitat,mil)
s.arrow(sp$cosup,add=T,plabels=list(alpha=1,col="darkblue",cex=0.7,boxes=list(draw=F)))
```



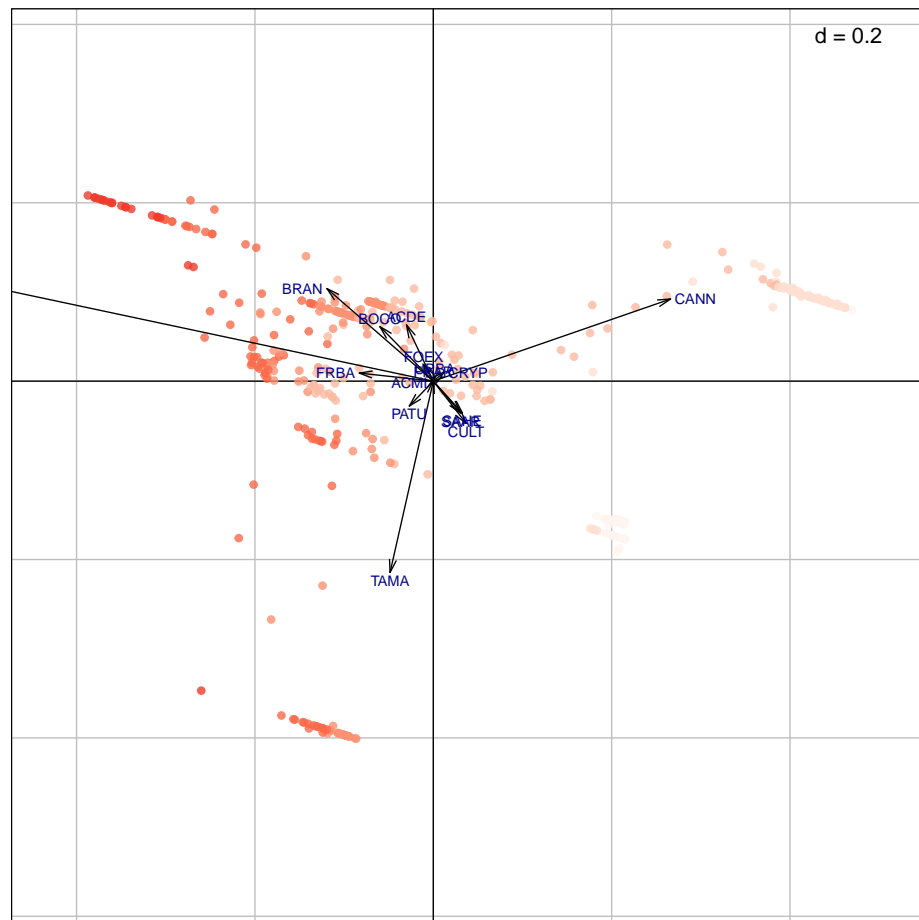
```
# axes 1 vs 3
s.label(pco.habitat$li,xax=1,yax=3,ppoints=list(alpha=0.8,col=c2),plabels=list(alpha=0,boxes=
s.arrow(sp$cosup,xax=1,yax=3,add=T,plabels=list(alpha=1,col="darkblue",cex=0.7,boxes=list(d
```



```
# axes 1 vs 4
s.label(pco.habitat$li,xax=1,yax=4,ppoints=list(alpha=0.8,col=c2),plabels=list(alpha=0,boxes=
s.arrow(sp$cosup,xax=1,yax=4,add=T,plabels=list(alpha=1,col="darkblue",cex=0.7,boxes=list(d
```



```
# axes 1 vs 5
s.label(pco.habitat$li,xax=1,yax=5,ppoints=list(alpha=0.8,col=c2),plabels=list(alpha=0,boxes=
s.arrow(sp$cosup,xax=1,yax=5,add=T,plabels=list(alpha=1,col="darkblue",cex=0.7,boxes=list(d
```



### Interprétation

Difficile sans bien connaître le contexte. Olivier, peux-tu avoir un apport là dessus? Voilà ce que j'arrive à sortir:

- PC1 = altitude - PC2 = habitats exo vs natifs - PC3 = ?? - PC4 = ??
- PC5 = ??

### table Oiseaux

#### Toutes espèces

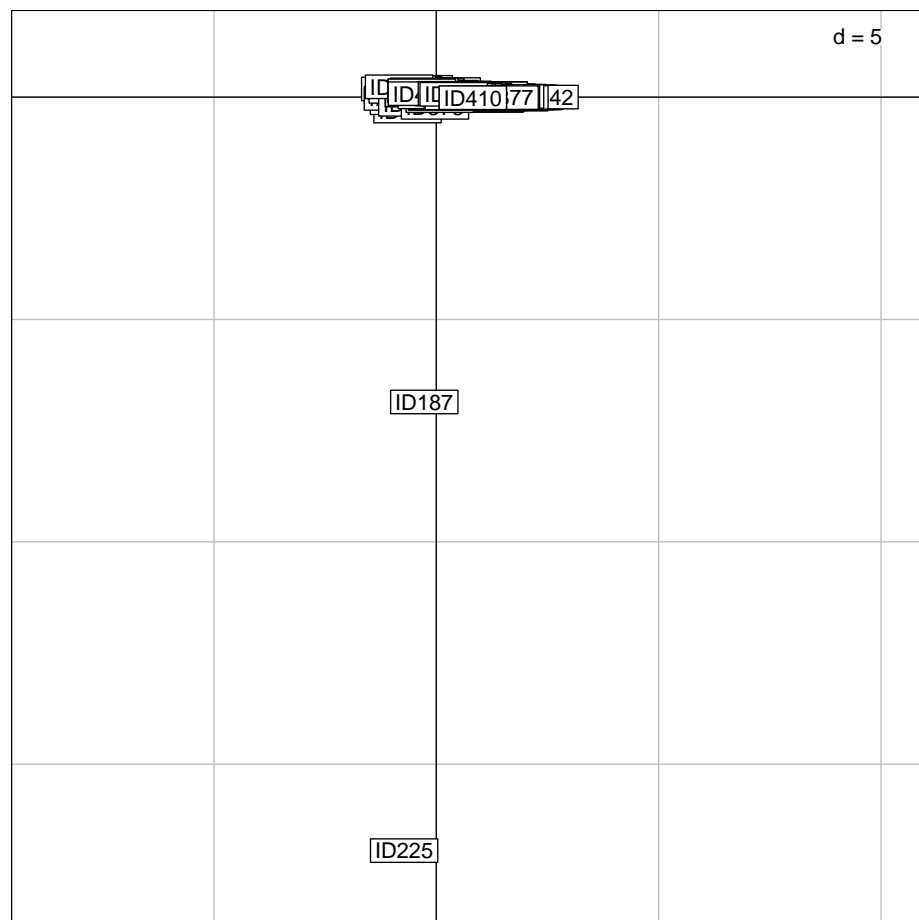
```
# valeurs propres
cumsum(afc.ois$eig)/sum(afc.ois$eig)
```

### AFC sur les abondances

```
## [1] 0.1330918 0.2302468 0.3112237 0.3820102 0.4511414 0.5094170 0.5662222 0.6191457 0.6685437
## [15] 0.8685437 0.8954101 0.9199879 0.9439219 0.9661262 0.9810814 0.9929844 1.0000000
```

Pas évident de retenir des gradients bien clairs, la variation est trop graduelle. Passer en présence/absence ne change rien au problème.

```
# nuage de points
s.label(afc.ois$li)
```



Il y a deux points extrêmes :

```
# 2 points extrêmes
ois[c("ID187", "ID225"),]
```

```
##          ACTR CIMA COCH COCO COFR ESAS FOMA FRPO GEST HYBO LOPU MAMA PADO PEAS PHBO PLCU PY.
## ID187      2    0    0    0    3    0    0    5    4    0    0    0    0    0    0    1
## ID225      2    0    0    0    0    0    0    3    0    0    0    0    0    0    0    0
```

```
apply(ois, 2, "max")
```

```
## ACTR CIMA COCH COCO COFR ESAS FOMA FRPO GEST HYBO LOPU MAMA PADO PEAS PHBO PLCU PYJO SATI
##   34     2     2     8    10    28    57     5    31     6     1     5    17    11     6    42    10    10
```

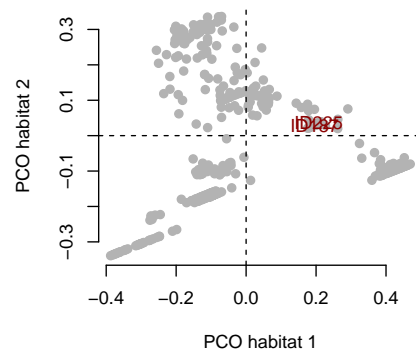
Correspondent en particulier aux seuls points avec Francolin - voir plus bas.

Position dans la PCOA “habitats” (tous les plans axe 1 vs l’un des 4 autres)

:

1 vs 2:

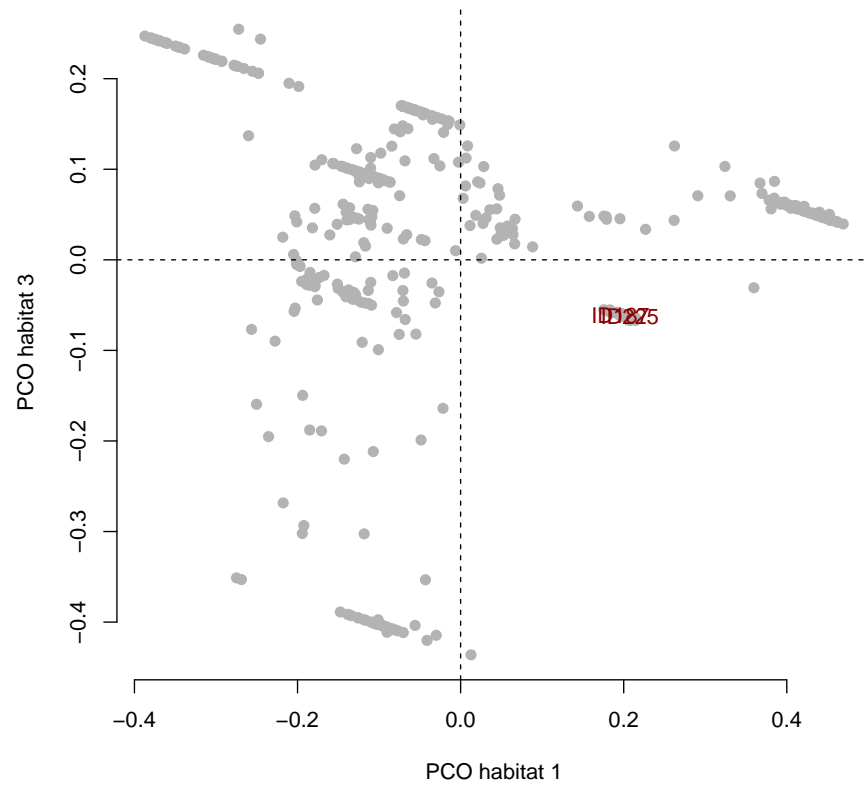
```
par(mfrow=c(2,2))
plot(pco.habitat$li[,1], pco.habitat$li[,2], bty="n", pch=21, bg="gray70", col="gray70", xlab="PCO
abline(h=0, lty="dashed")
abline(v=0, lty="dashed")
text(extr.ois[,1], extr.ois[,2], labels=rownames(extr.ois), col="darkred")
```



1 vs 3

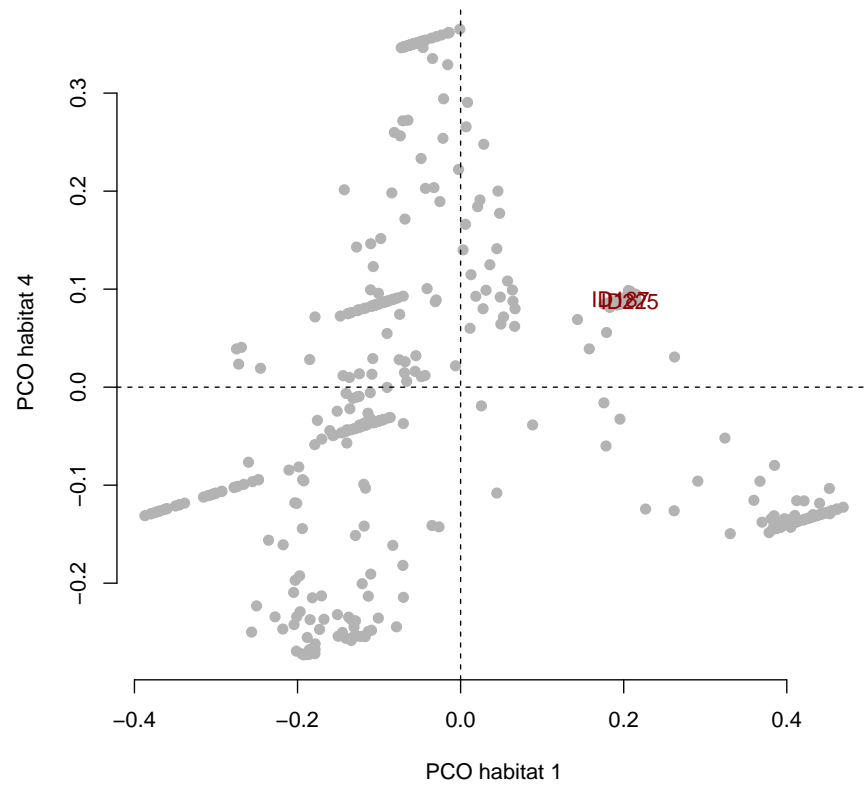
```
plot(pco.habitat$li[,1],pco.habitat$li[,3],bty="n",pch=21,bg="gray70",col="gray70",xlab="PCO habitat 1",ylab="PCO habitat 3")
abline(h=0,lty="dashed")
abline(v=0,lty="dashed")
text(extr.ois[,1],extr.ois[,3],labels=rownames(extr.ois),col="darkred")
```





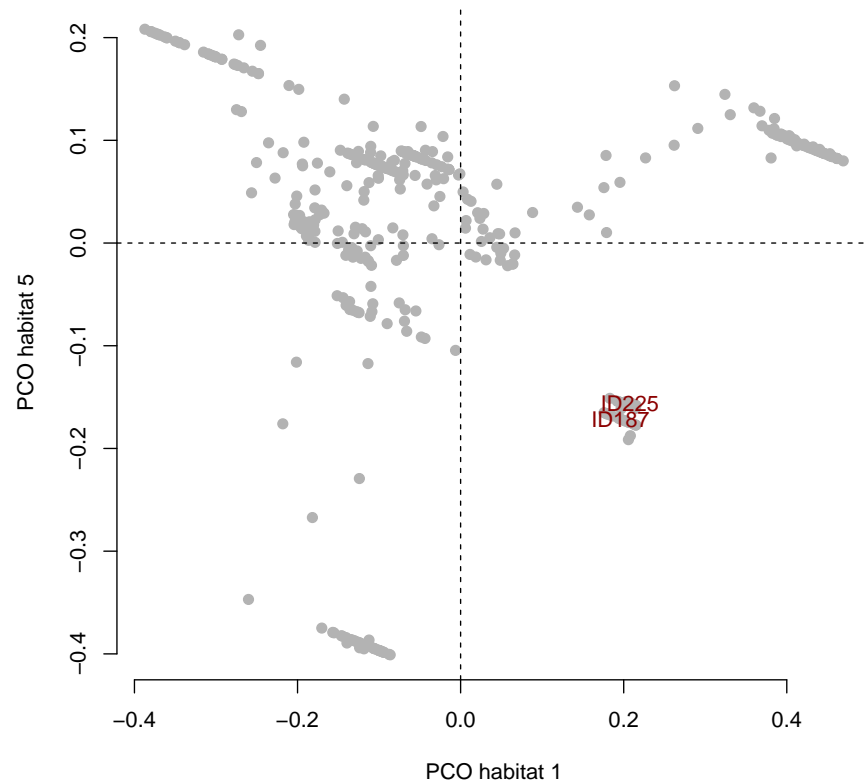
1 vs 4 :

```
plot(pco.habitat$li[,1],pco.habitat$li[,4],bty="n",pch=21,bg="gray70",col="gray70",xlab="PCO habitat 1",ylab="PCO habitat 4")
abline(h=0,lty="dashed")
abline(v=0,lty="dashed")
text(extr.ois[,1],extr.ois[,4],labels=rownames(extr.ois),col="darkred")
```



1 vs 5:

```
plot(pco.habitat$li[,1],pco.habitat$li[,5],bty="n",pch=21,bg="gray70",col="gray70",xlab="PCO habitat 1",ylab="PCO habitat 5")
abline(h=0,lty="dashed")
abline(v=0,lty="dashed")
text(extr.ois[,1],extr.ois[,5],labels=rownames(extr.ois),col="darkred")
```



L'analyse est écrasée par deux points qui n'ont rien de spécifique si ce n'est une espèce qui leur est unique, mais leur position sur les gradients d'habitat n'a rien d'extraordinaire. Je tente une analyse que sur les espèces bien distribuées pour y voir plus clair.

**table oiseaux : restreint aux espèces présentes sur >5% des points**

**AFC oiseaux sans espèces sous-représentées**

```
# abundances
sum(ois)
```

```
## [1] 6529
```

```
0.05*sum(ois)
```

```
## [1] 326.45
```

```
0.01*sum(ois)
```

```
## [1] 65.29
```

```
# fréquence  
frq.cum=colSums(ois)/nrow(ois)  
frq.cum[frq.cum<0.05]
```

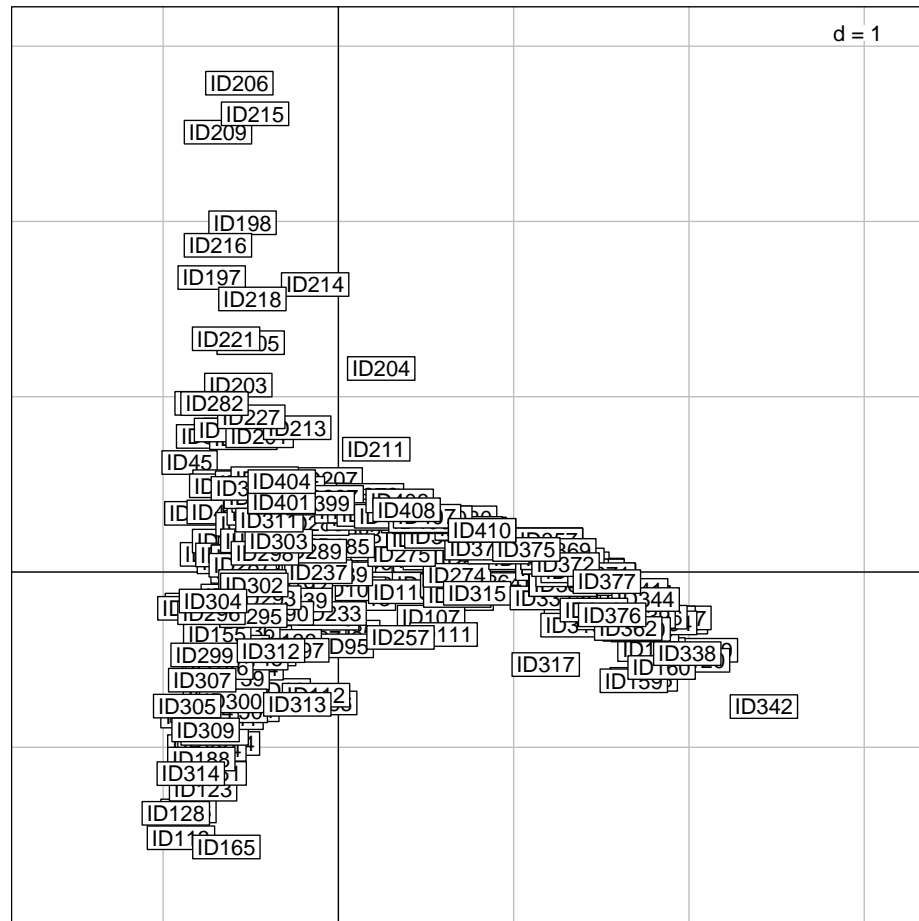
```
##          CIMA          COCH          FRPO          LOPU          MAMA          PHBO  
## 0.026829268 0.004878049 0.019512195 0.002439024 0.026829268 0.021951220
```

## AFC sans les espèces vues sur moins de 5% des points

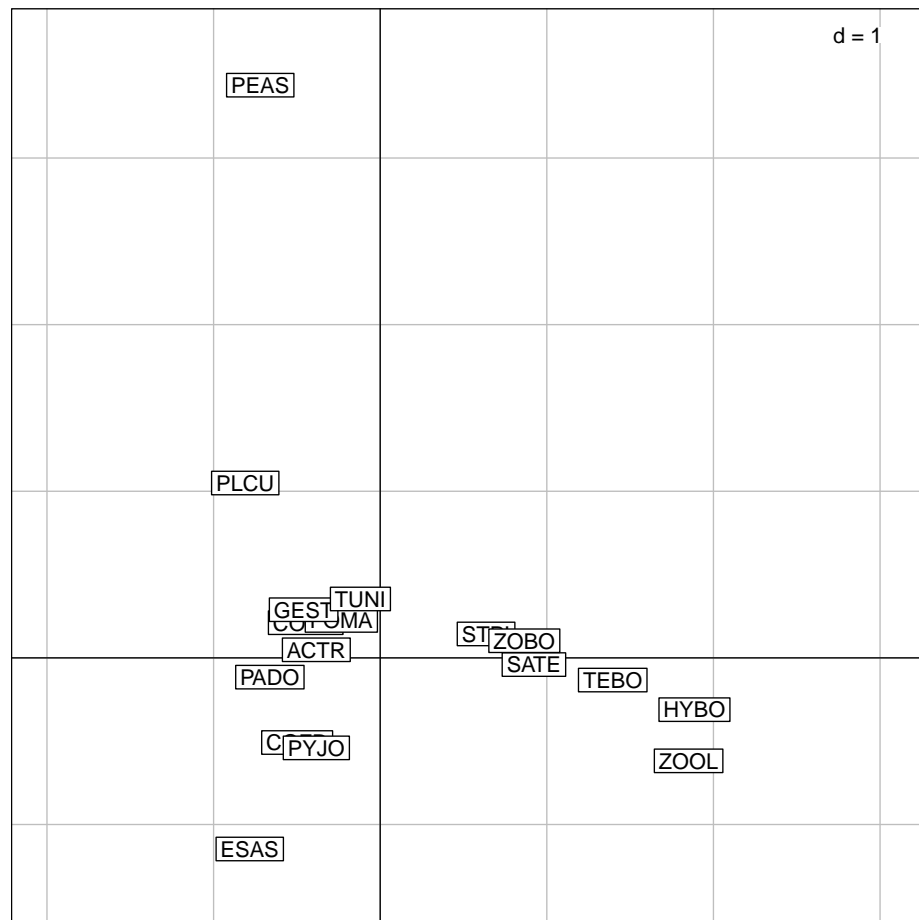
```
ois.com=ois[,names(frq.cum[frq.cum>0.05])]  
afc.ois.com=dudi.coa(ois.com,scannf=F,nf=2)  
cumsum(afc.ois.com$eig)/sum(afc.ois.com$eig)
```

```
## [1] 0.1685298 0.2707443 0.3599813 0.4456636 0.5192498 0.5897839 0.6562875 0.7063508 0.75  
## [15] 0.9718166 1.0000000
```

```
s.label(afc.ois.com$li)
```



```
s.label(afc.ois.com$co)
```



Il y a un effet arche assez fort, mais au moins les points sont bien étalés dans l'ordination

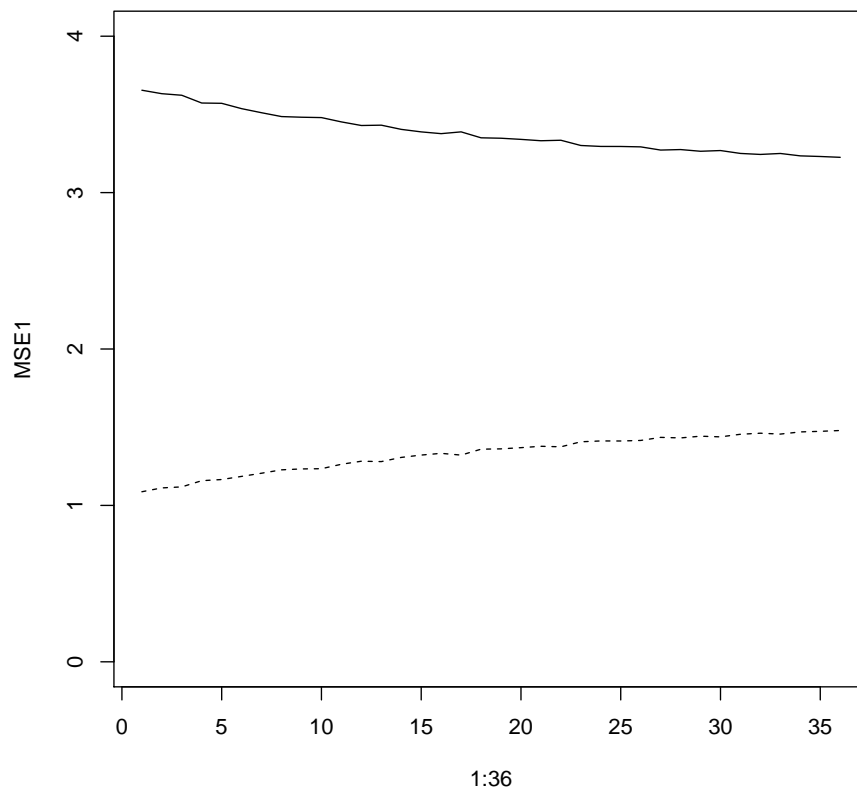
## Analyse fonctionnelle

Cette analyse fonctionnelle ne porte **que sur l'altitude** et sur **les abondances d'espèces log-transformées**. C'est donc très préliminaire, mais l'altitude étant très structurante ça donne déjà une première idée des distributions d'espèces.

## base optimale

Le choix de la base fonctionnelle optimale se fait par une comparaison entre des écarts fonctions-données pour différentes bases. Le but est d'avoir un compromis entre sur- et sous-lissage.

```
# on représente la variation du MSE avec l'ordre : on veut minimiser l'écart à la courbe moy  
plot(1:36,MSE1,type="l",ylim=c(0,4))  
lines(1:36,MSE2,type="l",lty="dashed")
```



Quelle que soit la base, la représentation n'est de toute façon pas très bonne, probablement à cause des patterns de réponse à l'altitude mal marqués (gros pics en plein milieu du gradient qui ne sont pas modélisables facilement et / ou grosse dominance de 0 partout). Faute de mieux on reste sur une base 4 (classique, permet de marquer 1 ou 2 optimums mais sans surlisser).

## Analyse fonctionnelle

On travaille sur abondances log transformées. Les abondances et les altitudes sont normalisées pour limiter l'impact de la rareté des espèces et faciliter l'estimation numérique.

```
# normaliser les abondances
oiscb=log(oisc+1)
maxab=apply(oiscb,2,max)
oisd=oiscb/maxab

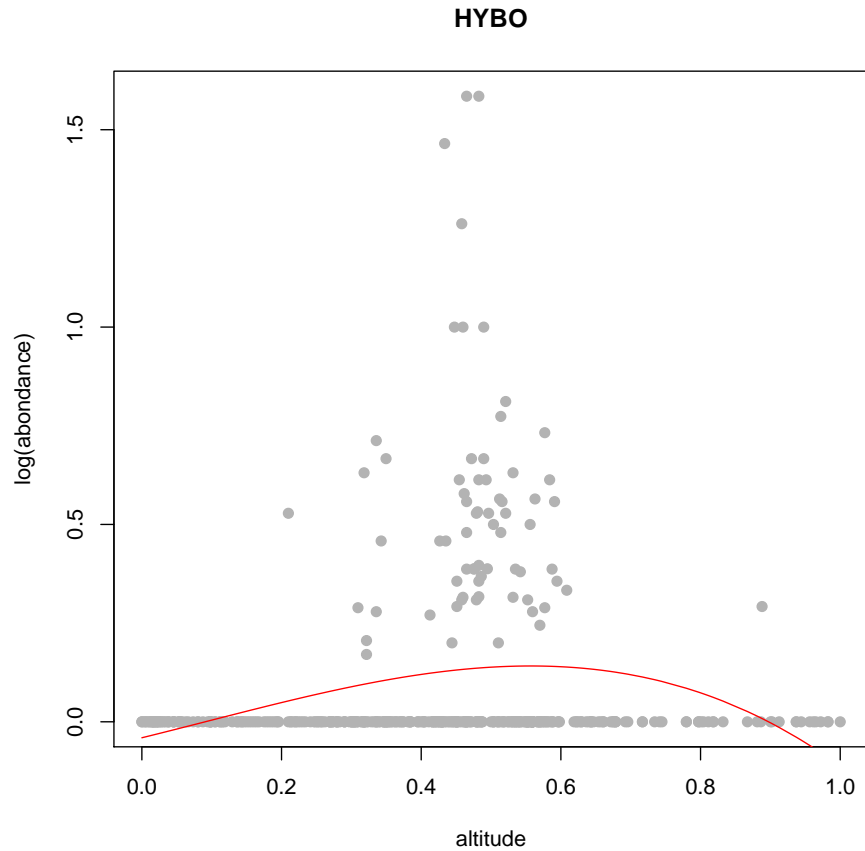
# normaliser les altitudes
alti4=(alti3-min(alti3))/max((alti3)-min(alti3))
```

## Estimation des courbes

Ces courbes décrivent, pour chaque espèce, la réponse à l'altitude estimée à partir de splines de base 4.

```
# plot d'une courbe
plot(alti4,oisd[,10],pch=21,bg="gray70",col="gray70",xlab="altitude",ylab="log(abondance)",m
lines(alti4,evalmyb6[,10],type="l",col="red")
```

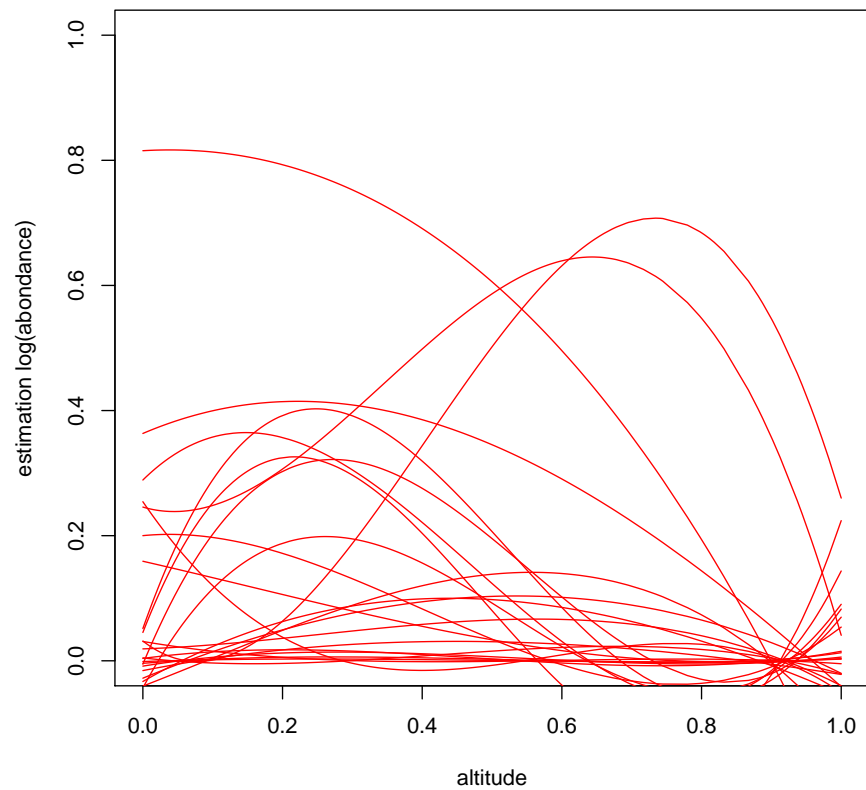




La dispersion des points est toujours forte et il y a souvent des gros pics d'abondance en milieu de gradient qui sont mal modélisables. En général les optimums sont au bon endroit sur l'axe des altitudes mais leur amplitude n'est pas bonne.

Comparaison des courbes de toutes les espèces :

```
# courbes toutes espèces
plot(x=0,y=0,xlim=c(0,1),ylim=c(0,1),type="n",xlab="altitude",ylab="estimation log(abondance)")
for(i in 1:ncol(evalmyb6)){
  lines(alti4,evalmyb6[,i],type="l",col="red")
}
```



## ACP fonctionnelle

```
mypca=pca.fd(datamyb6,nharm=2)
```

Cette ACP permet de créer des typologies de courbes. Les axes opposent des courbes de formes différentes.

Variance expliquée :

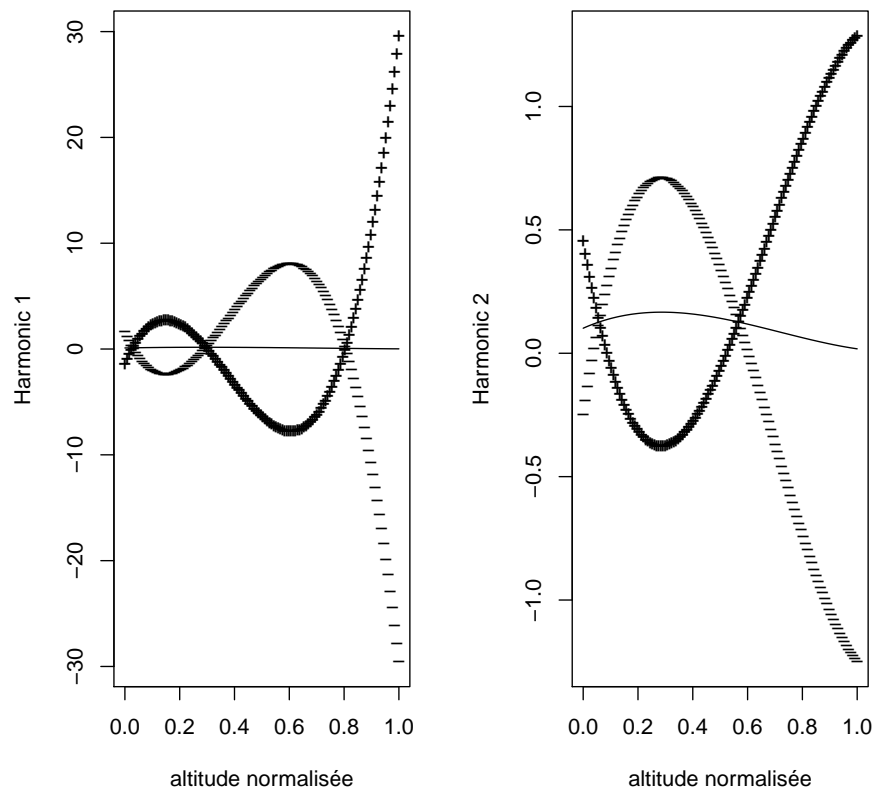
```
100*mypca$varprop
```

```
## [1] 87.45767 11.89144
```

## Interprétation des axes

```
# courbes extrêmes (et moyenne) qu'on a théoriquement aux extrémités des axes  
par(mfrow=c(1,2))  
plot(mypca,xlab="altitude normalisée")
```

➤ A function 1 (Percentage of variability) ➤ A function 2 (Percentage of variability)



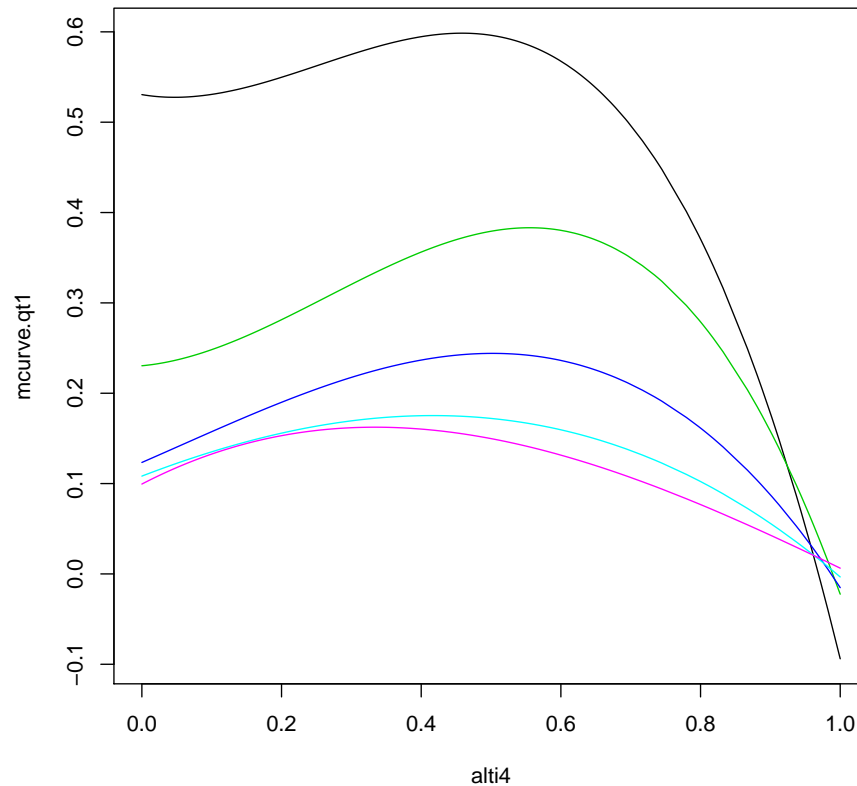
La courbe en trait plein est la courbe moyenne (ici plate : en moyenne les espèces ne répondent pas à l'altitude). La courbe “-” est la déformation de cette courbe moyenne vers le négatif de l'axe d'ACP, et la courbe “++” est la déformation vers les positifs.

- **axe 1** (courbe de gauche): sépare des espèces qui diminuent avec l'altitude (-) ou augmentent avec l'altitude (+)
- **axe 2** (courbe de droite) : sépare des espèces qui ont un pic d'abondance (-) ou un déficit d'abondance (+) dans les plaines.

Cette interprétation manque de netteté, probablement parce que les réponses à l'altitude sont un peu anarchiques et mal définies. Il faudra probablement réitérer l'exercice sur les axes de PCOA si on les comprend bien. On peut aussi interpréter les axes à partir des courbes quantiles, afin d'affiner la compréhension:

```
# axe 1 en quantiles
qt1=quantile(mypca$scores[,1],p=c(0.05,0.25,0.5,0.75,0.95))
scr.qt1=which(mypca$scores[,1]<qt1[1])
eva.qt1=evalmyb6[,scr.qt1]
mcurve.qt1=apply(eva.qt1,1,mean)
plot(alti4,mcurve.qt1,type="l")

for(i in 2:length(qt1)){
  scr.qt1=which(mypca$scores[,1]<qt1[i])
  eva.qt1=evalmyb6[,scr.qt1]
  mcurve.qt1=apply(eva.qt1,1,mean)
  lines(alti4,mcurve.qt1,type="l",col=i+1)
}
```



Les quantiles vont du noir (-) au violet (+) et montrent bien une typologie de patrons de réponse allant d'espèces présentes sur tout le gradient altitudinal jusqu'à un seuil après lequel elles déclinent brusquement à des espèces qui ont un optimum d'altitude mais une variation peu marquée. C'est déjà plus interprétable que les déformations.

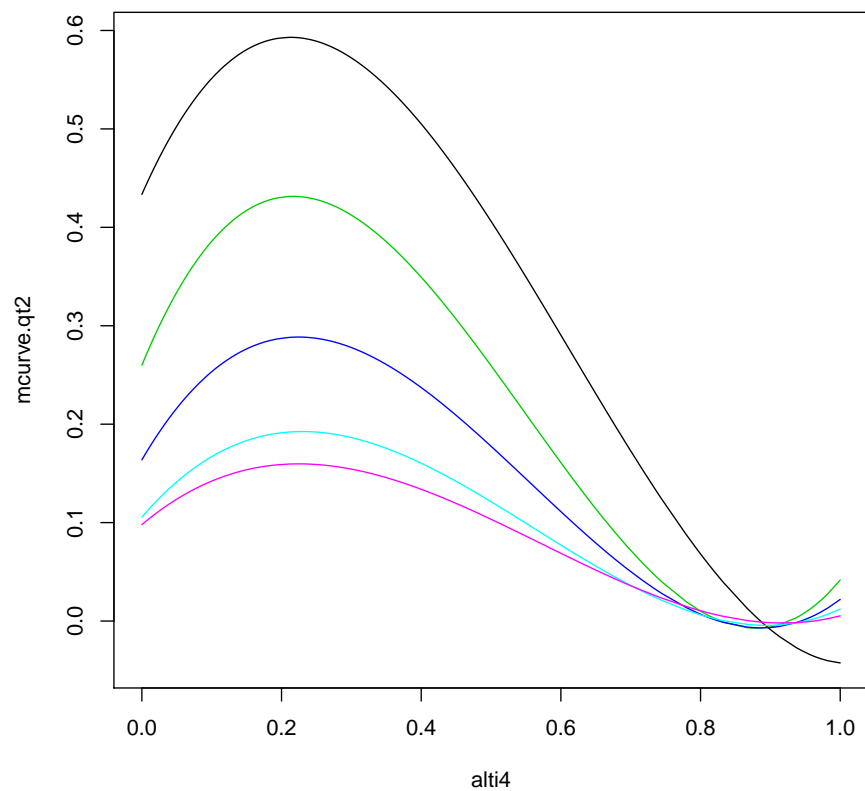
```
# axe 2 en quantiles
qt2=quantile(mypca$scores[,2],p=c(0.05,0.25,0.5,0.75,0.95))
scr.qt2=which(mypca$scores[,2]<qt2[1])
eva.qt2=evalmyb6[,scr.qt2]
mcurve.qt2=apply(eva.qt2,1,mean)
plot(alti4,mcurve.qt2,type="l")

for(i in 2:length(qt2)){
  scr.qt2=which(mypca$scores[,2]<qt2[i])
  eva.qt2=evalmyb6[,scr.qt2]
```

```

mcurve.qt2=apply(eva.qt2,1,mean)
lines(alti4,mcurve.qt2,type="l",col=i+1)
}

```



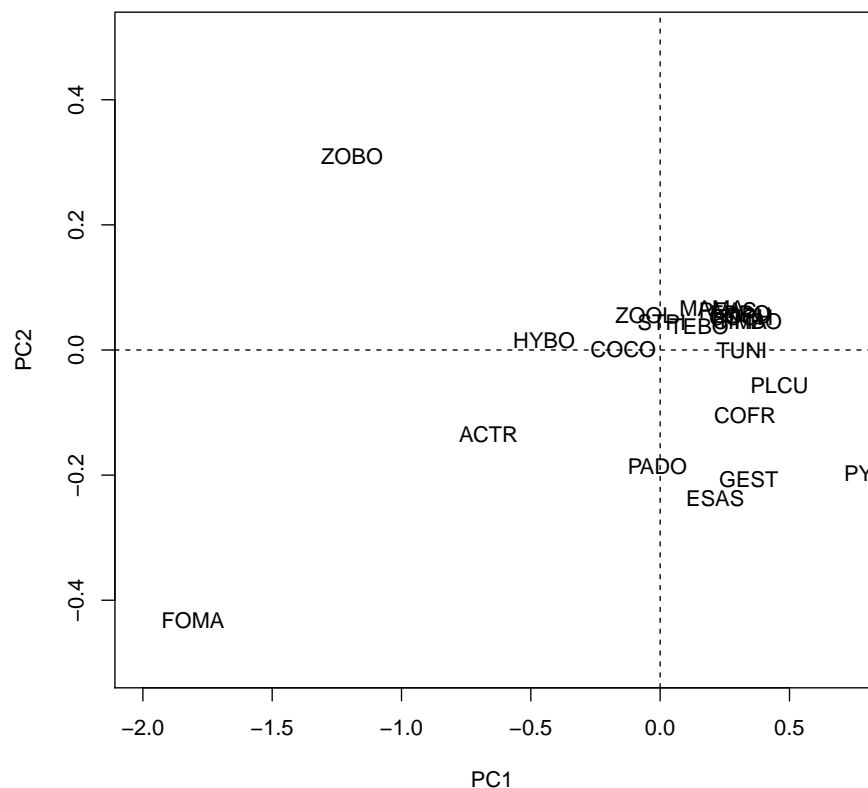
Le deuxième axe semble séparer les espèces en termes de fréquence, toutes devenant plus rares en altitude.

Le nuage de points de cette ACP:

```

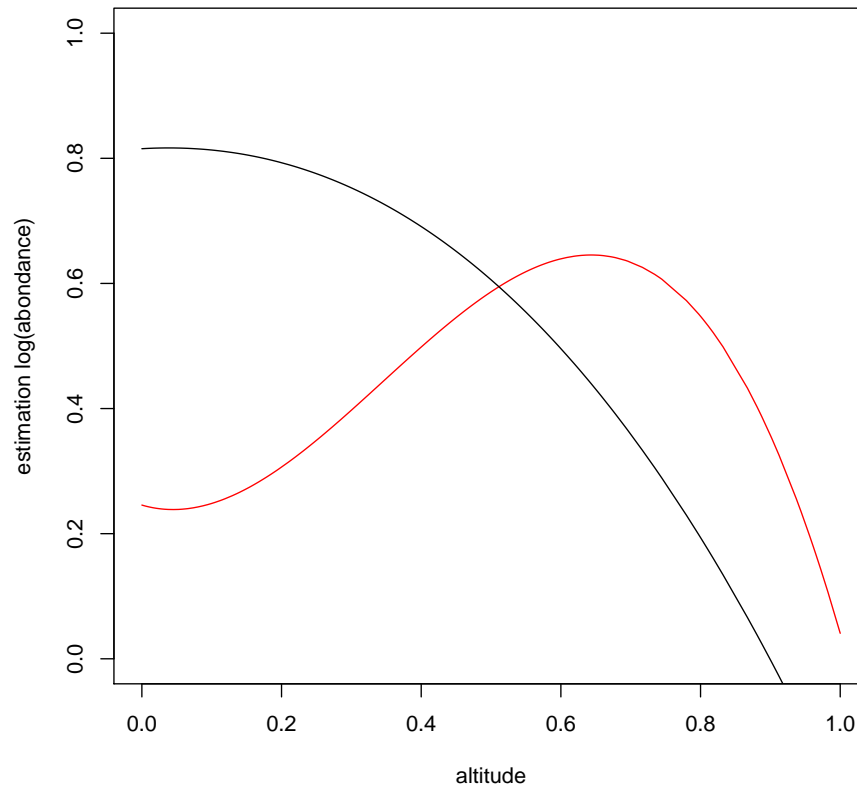
# nuage de points
rownames(mypca$scores)=colnames(oisd)
plot(mypca$scores,type="n",xlab="PC1",ylab="PC2",xlim=c(-2,0.7),ylim=c(-0.5,0.5))
abline(h=0,lty="dashed")
abline(v=0,lty="dashed")
text(mypca$scores,labels=rownames(mypca$scores))

```



La plupart des espèces est décalée dans le positif de l'axe 1, soit une réponse assez graduelle (et faible) à l'altitude avec une baisse d'abondance vers les sommets. Le négatif est dominé par le Foudi et le zosterops des Mascareignes, deux espèces qui sont donc présentes sur une bonne partie du gradient altitudinal avant une chute d'effectifs en haute altitude. On peut le vérifier sur les courbes individuelles (Zosterops : rouge ; Foudi : noir) :

```
plot(x=0,y=0,xlim=c(0,1),ylim=c(0,1),type="n",xlab="altitude",ylab="estimation log(abondance)
lines(alti4,evalmyb6["ZOBO"],type="l",col="red")
lines(alti4,evalmyb6["FOMA"],type="l",col="black")
```



## Conclusions

– L’approche à choisir : une approche statique (type PCOA / coinertie / RLQ) ou une approche par gradient (type FDA). Dépend d’à quel point on a confiance en l’une ou l’autre. Je n’estime pas avoir une connaissance suffisante de l’avifaune réunionnaise pour y répondre clairement : Olivier, ton retour serait vraiment utile.

– la suite des analyses : là on décrit juste l’organisation des espèces le long des gradients, il faut aller plus loin. Il y a l’approche traits qui est facile à ajouter aux deux types d’analyses, une répartition des ordinations sur la phylogénie, etc : à voir ce qui est le plus intéressant à vos yeux

– la robustesse : les éboulis de variance ne sont pas terrible, les 3-4 premiers axes des ordinations tentées ne représentent pas énormément de variance. Ca



suggère des résumés pas hyper pertinents (en tout cas bruités) des gradients qui nous intéressent. Je suis pas très clair sur leur représentation de la réalité du terrain (en particulier pour la matrice habitats).