


Consultation statistique avec le logiciel 
Calcul de l'*Outlier Mean Index*

D. Chessel

14 mars 2006

Suite à une question de Susanne Schmidt (Dresden), la fiche reprend des informations de base sur la méthode de couplage appelée *niche*.

Table des matières

1	introduction	1
2	Statistiques élémentaires dans niche	2
2.1	Une variable	3
2.2	Un test pour une espèce et une variable?	4
2.3	Deux variables	6
3	La fonction souhaitée	8
4	Analyse canonique de la marginalité	9
5	Conclusion	10
	Références	13

1 introduction

Susanne Schmidt pose cette question :

Bonjour,
je suis en train d'apprendre à utiliser "niche" sous ade4 sous R. Je lis l'article de Dolédec 2000 (Ecology) [3] <http://pbil.univ-lyon1.fr/R/articles/arti102.pdf> et le help page de "niche", et spécifiquement l'exemple "trichometeo". Ce que je n'arrive pas à comprendre c'est comment obtenir les nombres Inertia, OMI, Tol et RTol de `nic1`, comme ils sont utilisés dans le dit article (Table 4). Comment le faire? J'ai l'impression que ça serait un bon tableau pour un article.

Et comment faire l'analyse canonique de la marginalité en ade4 sous R, comme sur le page 26f. dans le document "ade4 Niches écologiques et couplages de tableaux"? <ftp://pbil.univ-lyon1.fr/pub/mac/ADE/ADE4/DocThemPDF/Thema47.pdf>

Merci beaucoup pour votre aide,


La question est posée en deux langues, pour plus de persuasion :

Hi there,

I am learning to use "niche" in ade4 in R, reading thoroughly the article by Dolédec 2000 (Ecology) [3] <http://pbil.univ-lyon1.fr/R/articles/arti102.pdf>, and the help page of "niche", especially the example "trichometeo". What I don't seem to be able to understand is how to get the Inertia, OMI, Tol and RTol values out of nic1, to build a table like the Table 4 in the named article. I think this would be a very useful table for publications?

And how to perform the canonical analysis of the marginality, detailed in the French document "ade4 Niches écologiques et couplages de tableaux", available through the ade4 internet site? <ftp://pbil.univ-lyon1.fr/pub/mac/ADE/ADE4/DocThemPDF/Thema47.pdf>

Thanks for your help!

De fait, dans un logiciel fermé comme ADE-4, le moindre calcul afférent à une méthode était programmé et reproductible. Dans  la situation est très différente. La puissance de calcul au clavier est considérable et ne font l'objet de fonction séparée que des calculs ayant une certaine consistance.

2 Statistiques élémentaires dans niche

Reprenons l'exemple de `trichometeo`. C'est un `data` de la librairie `ade4`.

```
library(ade4)
data(trichometeo)
```

Ces données sont décrites dans <http://pbil.univ-lyon1.fr/R/pps/pps034.pdf>. Elles ont pour origine les travaux [9][10] et [11]. On a un tableau de prélèvements faunistiques (piège lumineux pendant une nuit) :

```
fau <- log(trichometeo$fau + 1)
dim(fau)
```

```
[1] 49 17
```

```
names(fau)
```

```
[1] "Che" "Hyc" "Hym" "Hys" "Psy" "Aga" "Glo" "Ath" "Cea" "Ced" "Set" "All" "Han"
[14] "Hfo" "Hsp" "Hve" "Sta"
```

et un tableaux de descripteurs de milieu (conditions météorologique de la nuit de piégeage) :

```
mil <- trichometeo$meteo
dim(mil)
```

```
[1] 49 11
```

```
names(mil)
```

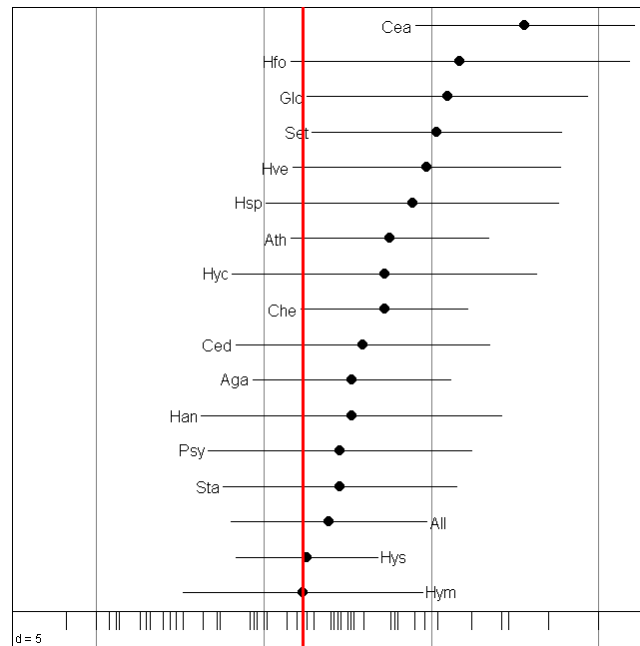
```
[1] "T.max"      "T.soir"      "T.min"      "Vent"      "Pression"
[6] "Var.Pression" "Humidité"   "Nébu.Nuit"  "Précip.Nuit" "Nébu.Moy"
[11] "Précip.Tot"
```

Ces 11 descripteurs placent 47 points dans \mathbb{R}^{11} . Ceci définit l'espace potentiel occupé par une espèce. C'est le point de vue de N. Perrin[?] et J. Hausser[?] qui ont défini la base de l'*ENFA*[5]. Noter que la partie essentiellement monospécifique et la contrainte forte des absences inconnues propre à l'*Ecological-Niche Factor Analysis* est en œuvre dans la fonction `enfa` de la librairie `adehabitat` de Clément Calenge. Les mêmes idées de niche écologique développée par N. Perrin et J. Hausser sont utilisée ici dans une optique très différente pour des objectifs essentiellement multispécifiques.

2.1 Une variable

Prenons la température du soir. Plaçons les valeurs observée et représentons la *niche* schématique de chaque espèce, c'est-à-dire d'abord un point moyen et un écart-type. La fonction `sco.distri` fait cela.

```
sco.distri(mil$T.soir, fau, incl = F)
abline(v = mean(mil$T.soir), lwd = 3, col = "red")
```



On a déjà l'essentiel du contenu des données. Il y a 36 possibilités sur 2^{17} pour que tous les points soit d'un même côté ou tous les points sauf un seul, soit une probabilité de 0.00027. Le hasard est déjà loin et nous n'en parlerons plus. Il est clair que l'émergence des insectes est favorisée par une température vespérale élevée. Chaque espèce a une abondance qui varie entre les points et donne une distribution de fréquences. Pour calculer les positions moyennes, ou mieux l'écart entre la position moyenne et la moyenne de la distribution uniforme :

```
sapply(fau, function(x) sum(x * mil$T.soir)/sum(x)) - mean(mil$T.soir)
```

Che	Hyc	Hym	Hys	Psy	Aga	Glo
2.41428571	2.41428571	-0.01951850	0.09761905	1.09312450	1.44528282	4.29693718
Ath	Cea	Ced	Set	All	Han	Hfo
2.58202916	6.61194326	1.76853021	3.98422019	0.75969578	1.43415781	4.66490470
Hsp	Hve	Sta				
3.25708958	3.67776809	1.08567276				

Soit n le nombre de stations, (p_1, \dots, p_n) la distribution de fréquence des abondances d'une espèce donnée $\sum(p_i) = 1$ et soit (x_1, \dots, x_n) les valeurs observées de la variable. Chacun sait que la variance vaut la moyenne des carrés moins le carré de la moyenne, ce qui vient du grand Pythagore. Donc :

$$\bar{x} = \sum_{k=1}^{k=n} p_k x_k \quad \sum_{k=1}^{k=n} p_k x_k^2 = \sum_{k=1}^{k=n} p_k (x_k - \bar{x})^2 + \bar{x}^2$$

Écologiquement, on dit que l'inertie autour de l'origine égale la marginalité (le carré de la distance entre l'origine et la moyenne de l'espèce) plus la tolérance (la variance, la variabilité autour de la position moyenne). Pour donner un sens à ces quantités l'origine est placée à la moyenne générale, car dans ce cas une espèce dont la moyenne est la moyenne générale a une marginalité nulle et une inertie entièrement égale à la tolérance. Prenons l'espèce 11 *Set* (*Setodes punctatus*).

```
grad <- mil$T.soir
grad <- grad - mean(grad)
freq <- fau$Set/sum(fau$Set)
posi <- sum(freq * grad)
inertia <- sum(freq * grad * grad)
margi <- posi^2
tol <- sum(freq * (grad - posi)^2)
c(inertia, margi, tol)
```

[1] 29.73316 15.87401 13.85915

Les valeurs ainsi définies dépendent des unités et on préférera exprimer la tolérance et la marginalité en terme de pourcentage de l'inertie.

```
grad <- mil$T.soir
grad <- grad - mean(grad)
fun1 <- function(x) {
  freq <- x/sum(x)
  posi <- sum(freq * grad)
  inertia <- sum(freq * grad * grad)
  margi <- posi^2
  tol <- sum(freq * (grad - posi)^2)
  return(c(margi = margi/inertia, tol = tol/inertia))
}
w <- apply(fau, 2, fun1)
w <- w[, order(w[1, ])]
round(w, dig = 2)
```

	Hym	Hys	All	Psy	Sta	Han	Ced	Aga	Hyc	Hsp	Ath	Hve	Hfo	Che	Glo	Set
margi	0	0	0.06	0.07	0.09	0.09	0.18	0.19	0.22	0.36	0.43	0.46	0.46	0.48	0.51	0.53
tol	1	1	0.94	0.93	0.91	0.91	0.82	0.81	0.78	0.64	0.57	0.54	0.54	0.52	0.49	0.47

```
margi Cea
margi 0.8
tol 0.2
```

On obtient l'équivalent numérique de la figure précédente.

2.2 Un test pour une espèce et une variable ?

Je pose cette question pour indiquer combien le débat statistique-écologie est délicat. Tout écologue dira que le lien entre une espèce et une variable de milieu se fait par l'intermédiaire d'une *courbe en cloche*, donc qu'il est stupide de mesurer le lien entre une espèce et une variable par une corrélation, laquelle ne prend en compte que les relations linéaires. On trouve cette assertion répétée à satiété dans la littérature.

Aucun écologue ne contestera qu'un test élémentaire peut légitimement porter sur la marginalité, la première expression du choix d'une espèce portant sur la position moyenne qu'elle occupe, sans doute proche de son *optimum*. Peut-on dire si la moyenne de la position de l'espèce est *significativement* différente de la moyenne générale? Si non, on se posera la question de son *amplitude*, l'espèce pouvant être spécialisée au centre. La statistique utilisé pour la marginalité est $\bar{x}^2 = (\sum_{k=1}^{k=n} p_k x_k)^2$. Si la variable \mathbf{x} est centrée pour la pondération uniforme, la covariance entre la distribution de l'espèce vue comme une variable quantitative \mathbf{p} et la variable \mathbf{x} vaut $\frac{1}{n} \sum_{k=1}^{k=n} p_k x_k$. Tester la nullité de la covariance, ou de la corrélation c'est **exactement** tester la nullité de la marginalité. Alors pourquoi l'un serait stupide et l'autre valide, puisque mathématiquement, c'est la même chose?

```
grad <- mil$T.soir
grad <- (grad - mean(grad))/sd(grad)
freq <- fau$Aga/sum(fau$Aga)
simposi <- fonction(k) (sum(sample(freq) * grad))^2
simobs <- sum(freq * grad)^2
test1 <- as.randtest(sapply(1:9999, simposi), simobs)
test1

Monte-Carlo test
Observation: 0.1356685
Call: as.randtest(sim = sapply(1:9999, simposi), obs = simobs)
Based on 9999 replicates
Simulated p-value: 0.0315

cor.test(freq, grad)

Pearson's product-moment correlation

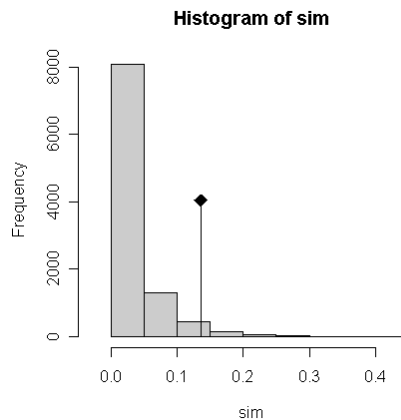
data: freq and grad
t = 2.232, df = 47, p-value = 0.03042
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.03108373 0.54346197
sample estimates:
      cor
0.3095744

anova(lm(freq ~ grad))

Analysis of Variance Table

Response: freq
      Df Sum Sq Mean Sq F value Pr(>F)
grad   1 0.0028264 0.0028264  4.9817 0.03042 *
Residuals 47 0.0266658 0.0005674
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(test1)
```



Cela laisse rêveur. On sait bien que l'analyse de variance est strictement équivalente au test de la corrélation. On a rajouté ici le test non paramétrique sur la marginalité. 9999 fois, on attribue au hasard les abondances de l'espèce aux prélèvements : ceci simule l'indifférence de l'espèce. On calcule sa marginalité et on compare à la valeur observée. On apprend quelque chose : tester la corrélation entre l'abondance d'une espèce et une variable de milieu c'est tester la non nullité de la marginalité de sa niche. C'est dire si les modèles linéaires n'ont aucun sens écologique !

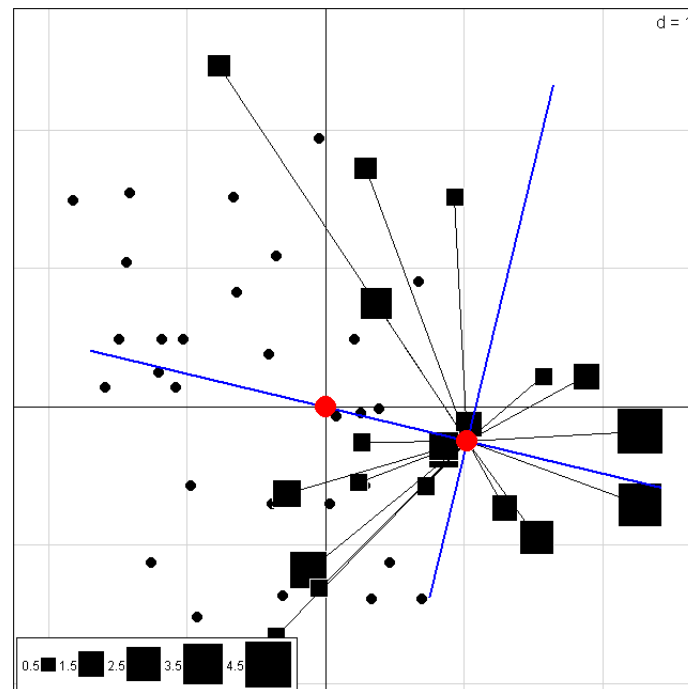
2.3 Deux variables

Dès qu'on a plusieurs variables d'unités différentes comme les pressions, les températures ou les précipitations, la normalisation des variables s'impose. Prenons la température du soir et la pression atmosphérique. Cela se complique un peu mais pas trop. Tout est en deux dimensions.

```

mil0 <- scalewt(mil)
xy <- mil0[, c(2, 5)]
s.label(xy, clab = 0, cpoi = 2)
s.value(xy, fau$Set, add.plot = T)
freq <- fau$Set/sum(fau$Set)
mx <- sum(xy[, 1] * freq)
my <- sum(xy[, 2] * freq)
s.distri(xy, freq, cstar = 1, cell = 0, add.p = T)
abline(c(0, my/mx), lwd = 2, col = "blue")
abline(c((mx^2 + my^2)/my, -mx/my), lwd = 2, col = "blue")
points(0, 0, pch = 19, cex = 3, col = "red")
points(sum(xy[, 1] * freq), sum(xy[, 2] * freq), pch = 19, cex = 3,
       col = "red")

```



L'inertie devient la moyenne (pour la distribution \mathbf{p}) du carré de la distance des points à l'origine, la marginalité devient le carré de la distance à l'origine de la moyenne mais la variabilité résiduelle se décompose en deux dimensions. On peut prendre les variables d'origine pour calculer le total de cette variabilité (l'inertie autour du centre de gravité est souvent appelée variance généralisée) comme somme des variances. Mais écologiquement elle se décompose en deux parties distinctes, la première le long de l'axe de marginalité ou tolérance marginale, le reste devenant la tolérance résiduelle. C'est là qu'intervient l'idée de J. Hausser (*op. cit.*) dont on ne sert pas ici. Dès qu'on a trois dimensions, la variabilité résiduelle s'exprime en deux dimensions et on peut chercher si dans ce sous-espace, on a des directions privilégiées d'étroitesse de la niche.

Pour ce qui nous concerne, les calculs de base sont :

```

inertia <- sum(freq * xy * xy)
margi <- mx^2 + my^2
xy0 <- t(t(xy) - c(mx, my))
tolt <- sum(freq * xy0 * xy0)
u <- c(mx, my)/sqrt(mx^2 + my^2)
z <- xy0 %*% u
tolm <- sum(freq * z * z)
tolr <- tolt - tolm
w <- c(inertia, margi, tolm, tolr)
names(w) <- c("inertia", "margi", "tolm", "tolr")
w

      inertia      margi      tolm      tolr
2.7977358 1.1134586 0.9699923 0.7142849

w <- round(w[2:4]/w[1], dig = 2)
w

      margi      tolm      tolr
0.40      0.35      0.26

```

Remarquer le joli exercice qui consiste à projeter les points sur le vecteur de la marginalité pour avoir la variance des coordonnées des projections. Remarquer aussi le recyclage des poids le long de la matrice des carrés pour sommer les variances en une ligne.

3 La fonction souhaitée

Nous sommes maintenant suffisamment renseigné pour écrire la fonction es-pérée.

```
niche.param <- function(nic) {
  if (!inherits(nic, "niche"))
    stop("Object of class 'niche' expected")
  appel <- as.list(nic$call)
  X <- eval(appel[[2]], sys.frame(0))$tab
  Y <- eval(appel[[3]], sys.frame(0))
  w1 <- apply(Y, 2, sum)
  if (any(w1 <= 0))
    stop(paste("Column sum <=0 in Y"))
  Y <- sweep(Y, 2, w1, "/")
  calcul.param <- function(freq) {
    inertia <- sum(freq * X * X)
    m <- apply(freq * X, 2, sum)
    margi <- sum(m^2)
    X <- t(t(X) - m)
    tolt <- sum(freq * X * X)
    u <- m/sqrt(sum(m^2))
    z <- X %>% u
    tolm <- sum(freq * z * z)
    tolr <- tolt - tolm
    w <- c(inertia, margi, tolm, tolr)
    names(w) <- c("inertia", "OMI", "Tol", "Rtol")
    w1 <- round(w[2:4]/w[1], dig = 3) * 100
    names(w1) <- c("omi", "tol", "rtol")
    return(c(w, w1))
  }
  res <- apply(Y, 2, calcul.param)
  t(res)
}
```

Pour s'en servir :

```
data(trichometeo)
pca1 <- dudi.pca(trichometeo$meteo, scan = FALSE)
nic1 <- niche(pca1, log(trichometeo$fau + 1), scan = FALSE)
niche.param(nic1)
```

	inertia	OMI	Tol	Rtol	omi	tol	rtol
Che	6.433882	2.77316816	1.0214504	2.639263	43.1	15.9	41.0
Hyc	11.914482	4.44884944	2.3877161	5.077916	37.3	20.0	42.6
Hym	10.573796	0.09548554	2.5386420	7.939669	0.9	24.0	75.1
Hys	7.625791	0.63040842	0.7348512	6.260531	8.3	9.6	82.1
Psy	10.470153	0.43447855	3.9237418	6.111932	4.1	37.5	58.4
Aga	7.430579	1.29116377	1.5507447	4.588670	17.4	20.9	61.8
Glo	14.360078	6.17685139	4.7591657	3.424061	43.0	33.1	23.8
Ath	11.244671	1.79679264	2.7654073	6.682471	16.0	24.6	59.4
Cea	18.711518	12.23859181	4.1775853	2.295341	65.4	22.3	12.3
Ced	11.789951	0.87321186	3.2451344	7.671604	7.4	27.5	65.1
Set	12.607986	4.28597109	3.7224679	4.599547	34.0	29.5	36.5
All	6.805252	0.72091250	1.2144331	4.869906	10.6	17.8	71.6
Han	10.368865	1.20620645	3.3672977	5.795361	11.6	32.5	55.9
Hfo	17.543552	6.75786236	7.3444406	3.441250	38.5	41.9	19.6
Hsp	13.976515	2.89982751	5.6222008	5.454487	20.7	40.2	39.0
Hve	12.253601	4.59849113	3.5177233	4.137387	37.5	28.7	33.8
Sta	9.391826	0.58873968	2.5226450	6.280442	6.3	26.9	66.9

C'est bien le tableau de l'article cité p. 2920. On ajoutera cette fonction utilitaire dans la prochaine version d'ade4. En attendant, copier-coller et ça marche.

4 Analyse canonique de la marginalité

La dernière partie du message fait référence à une analyse canonique de la marginalité qui n'a pas été portée dans `ade4`. On pouvait lire dans le texte cité :

Nous revenons au couplage de tableaux écologiques sur la base de l'optimisation de la marginalité. La méthode OMI est une méthode de couplage basée sur le schéma $(\mathbf{F}^t \mathbf{X}, \mathbf{Q}, \mathbf{D}_t)$ où \mathbf{F} est le tableau des profils (colonnes = taxons), \mathbf{X} le tableau de milieu, \mathbf{Q} la norme utilisée dans l'analyse préliminaire du triplet $(\mathbf{X}, \mathbf{Q}, \mathbf{D}_n)$ et \mathbf{D}_t les poids des taxons.

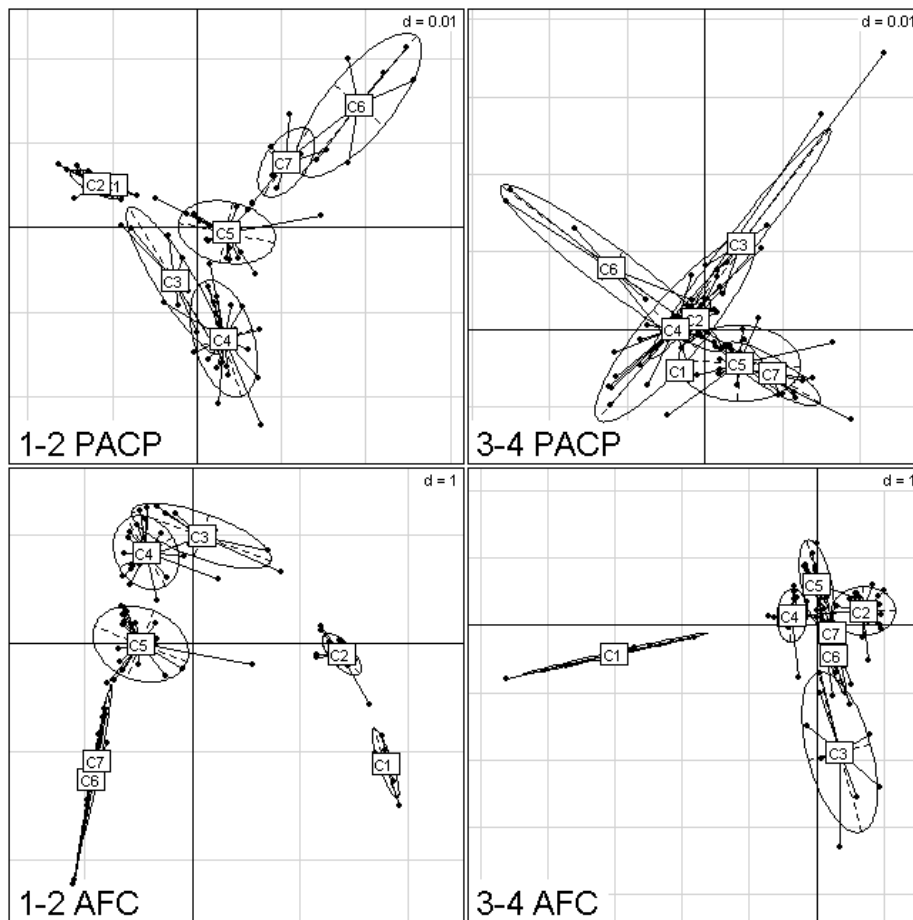
Ceci nous a amené à définir l'analyse du triplet $(\mathbf{F} - \frac{1}{n} \mathbf{U}_{nt}, \mathbf{D}_t, \mathbf{D}_n)$ que nous avons appelé PACP, méthode intermédiaire entre AFC et ACP. Il existe donc une analyse de co-inertie entre la PACP du tableau floro-faunistique et l'analyse de départ puisque les triplets $(\mathbf{X}, \mathbf{Q}, \mathbf{D}_n)$ et $(\mathbf{F} - \frac{1}{n} \mathbf{U}_{nt}, \mathbf{D}_t, \mathbf{D}_n)$ sont correctement appariés par les lignes (relevés). Cette analyse est basée sur le triplet $((\mathbf{F} - \frac{1}{n} \mathbf{U}_{nt})^t \mathbf{D}_n \mathbf{X}, \mathbf{Q}, \mathbf{D}_t)$. Comme \mathbf{D}_n est la pondération uniforme, on retrouve au coefficient n près l'analyse OMI de départ. L'analyse OMI est donc une analyse de co-inertie.

Dans la PACP on cherche des scores des relevés centrés réduits pour optimiser la marginalité moyenne des espèces, dans l'analyse OMI on cherche des combinaisons de variables de milieu pour faire de même. On a donc intérêt à chercher les axes de marginalité ayant des bonnes propriétés d'inertie (forte variance des scores) et des bonnes propriétés de marginalité (éloignement des centres de gravité des espèces de l'origine).

On se retrouve exactement dans la discussion entre ACC et Co-Inertie. Il y a nécessairement une analyse qui cherche des combinaisons de variables de milieu centrées réduites qui optimise la marginalité, ce qui est la version ACC de l'analyse OMI. On peut appeler cette procédure Analyse canonique OMI. Elle est basée sur le triplet $(\mathbf{F}^t \mathbf{X}, (\mathbf{X}^t \mathbf{D}_n \mathbf{X})^{-1}, \mathbf{D}_t)$ où on introduit la norme de Mahalanobis (au sens large, voir fiche thématique 5-3).

Personne n'ayant fait allusion à cette direction, je n'y suis pas revenu. On peut simplement faire deux remarques. La PACP n'existe pas mais les outils généraux de `ade4` permettent de la reconstruire en un instant. Il suffira de préparer une pondération des lignes, une pondération des colonnes et un tableau transformé pour avoir un schéma de dualité. Voici un exemple pour faire la PACP :

```
data(mafragh)
flo <- mafragh$flo
flo.afc <- dudi.coa(flo, scannf = FALSE, nf = 4)
w1 <- apply(flo, 2, sum)
if (any(w1 <= 0)) stop(paste("Column sum <=0 in Y"))
flo <- sweep(flo, 2, w1, "/")
flo <- as.data.frame(scalewt(flo, scale = FALSE))
flo.wc <- w1/sum(w1)
flo.wl <- rep(1, nrow(flo))/nrow(flo)
flo.pacp <- as.dudi(flo, flo.wc, flo.wl, scannf = FALSE, nf = 4,
  call = NULL, type = "pacp")
par(mfrow = c(2, 2))
s.class(flo.pacp$li, mafragh$partition, sub = "1-2 PACP", csub = 2)
s.class(flo.pacp$li[, 3:4], mafragh$partition, sub = "3-4 PACP",
  csub = 2)
s.class(flo.afc$li, mafragh$partition, sub = "1-2 AFC", csub = 2)
s.class(flo.afc$li[, 3:4], mafragh$partition, sub = "3-4 AFC", csub = 2)
```



Les différences sont évidentes. Les groupe 6 et 7 indissociables en AFC sont bien isolés et le groupe 1 semble difficile à distinguer du groupe 2.

Les groupes 1 et 2 sont des faciès de l'association *Scirpetum maritimi*. Les autres unités appartiennent à l'association *Suaedetum fruticosae* : 3 est dans la sous-association à *Salicornia fruticosa* et les unités 4 6 et 7 sont des groupes floristiques de la sous-association à *Inula crithmoïdes*. Il est curieux que les discriminations 1 contre 2 et 6 contre 7 soient exprimées de manière si différentes. Ce sont aux extrémités des gradients d'ordination que les contraintes numériques sont les plus fortes.

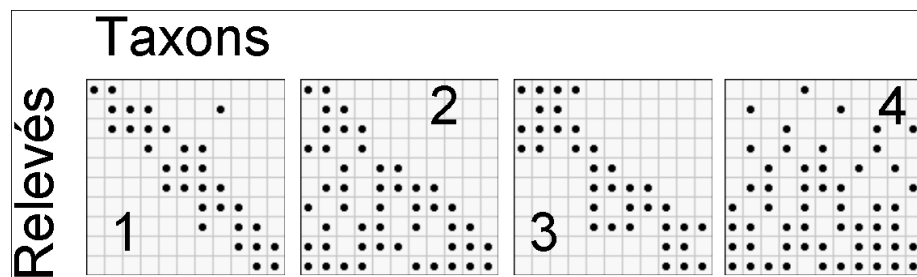
Pour faire une analyse canonique de la marginalité, la *recette* consiste à remplacer les données par les coordonnées normées (composante c_1) de l'ACP initiale. Il faudrait avoir une bonne raison pour essayer cela. Il vaut mieux s'en tenir aux bases solides pour le moment. La conclusion de la fiche citée mérite un petit rappel.

5 Conclusion

L'introduction d'une variante des méthodes de base n'est pas sans soulever un problème sérieux. Faut-il encore perturber l'utilisateur occasionnel par de

nouveaux boutons? La question renvoie à la fonction de la statistique, ici la statistique euclidienne destinée à voir des structures multidimensionnelles. Cet objectif n'est pas simple et l'énorme littérature biométrique sur l'ordination en témoigne. Faire des citations relève du tirage aléatoire.

C'est d'abord une question de modèles conceptuels. Il n'y a aucune méthode qui repère avec une même précision des organisations des types 1 (ordination des niches sur un gradient), 2 (ordination sur un facteur limitant), 3 (partition dans le modèle phyto-sociologique) ou 4 (facteur d'abondance pur) :



Il y a une très forte contradiction entre la connaissance déjà ancienne de la généralité du modèle ACP et de la plasticité potentielle des paramètres d'une part, de la domination de deux programmes (ACP sur matrice de covariances et AFC) d'autre part. Les décompositions de l'inertie dans les ACP centrées par espèces (mode R) et par relevés (mode Q) avec ou sans normalisation sont explicites dans [7] et l'effet du centrage discuté dans [8]. Neuf transformations préalables sont étudiées dans [1].

1. aucune (*Unmodified data*) ;
2. standardisation par espèces : l'abondance est exprimée en pourcentage de la valeur maximum observée pour une espèce (*Species maximum = 100*) ;
3. fréquence par relevés : l'abondance est exprimée en pourcentage de la valeur totale observée pour un site (*Stand total = 100*) ;
4. fréquence par relevés après standardisation par espèces (*Stand total = 100 after species maximum = 100*) ;
5. normalisation sans centrage par espèces (division par la longueur d'après [8]) ;
6. centrage et normalisation par sites (*Standard deviation*) ;
7. passage en présence/absence ;
8. double centrage additif (*Deviation from expected value*) ;

Pour chacune de ces transformations, la matrice des produits scalaires est diagonalisée et le rapport des deux premières valeurs propres édité sur les cartes factorielles. Pour M.P. Austin, la notion de la généralité de l'ACP est clairement acquise et l'ordination inventée par les écologues (Méthode de Bray et Curtis[?]) est abandonnée au profit de l'ACP dans [2].

Douze transformations initiales sont introduites dans la théorie générale de la diagonalisation des matrices $\mathbf{X}^t\mathbf{X}$ et $\mathbf{X}\mathbf{X}^t$ dans [6].

1. aucune (*original data*) ;
2. centrage par espèces (*centering by species mean*) ;

3. centrage par relevé (*centering by site mean*);
4. double centrage additif (*double centering by both species and site*);
5. profils par espèces (*standardization by species total*);
6. profils par sites (*standardization by site total*);
7. normalisation sans centrage par espèces (*standardization by species norm*);
8. normalisation sans centrage par sites (*standardization by site norm*);
9. double normalisation sans centrage ($x_{ik} / \sqrt{\sum_i x_{ik}^2} \sqrt{\sum_k x_{ik}^2}$);
10. centrage et normalisation par sites (*standardization by standard deviation*);
11. centrage et normalisation par espèces (*standardization by standard deviation*);
12. double centrage normalisé alternatif jusqu'à convergence.

En référence à Gower (1966) [4] on a donc un usage intensif du schéma de dualité à deux produits scalaires canoniques dans l'article de Noy-Meir (1973) [6]. D'autres transformations préalables sont encore citées dont

$$x_{ik} / \sqrt{\sum_i x_{ik}} \sqrt{\sum_k x_{ik}}$$

(*double standardization by site total and species total*) en référence à [?] car il s'agit bien d'AFC. Le terme *standardization* est utilisé en général pour division par une quantité. On a donc passage en fréquences (profils) par *standardization* par la marge, passage en longueur unité par *standardization* par la norme et passage en variance unité par *standardization* par l'écart-type.

Le terme de dualité (*The R/Q problem : duality*) est introduit en écologie dans l'article fondamental de Noy-Meir et Whittaker 1977[?].

After twenty-five years of development of continuous multivariate techniques in ecology, some of the early optimistic promises, as well as some of the septical criticisms, seem to have been overstated. No single method has emerged as a solution to all problems of describing and explaining patterns of compositional variation in natural communities. However, several ordination techniques have proved to be helpful tools in the analysis community data for various scientific and practical purposes.

Avec l'ACP centrée par espèce sur les profils espèces (PACP), l'analyse de co-inertie qui lui est associée (Analyse OMI) et peut-être l'analyse canonique (au sens Canonical Correspondence Analysis) qui en dérive basée sur la marginalité des niches peuvent faire partie de ces outils utiles.

Références

- [1] M.P. Austin and P. Greig-Smith. The application of quantitative methods to vegetation survey. ii some methodological problems of data from rain forest. *Journal of Ecology*, 56 :827–844, 1968.
- [2] M.P. Austin and L. Orłóci. Geometric models in ecology ii an evaluation of some ordination techniques. *Journal of Ecology*, 54 :217–227, 1966.
- [3] S. Dolédec, D. Chessel, and C. Gimaret. Niche separation in community analysis : a new method. *Ecology*, 81 :2914–1927, 2000.
- [4] J.C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53 :325–338, 1966.
- [5] A.H. Hirzel, J. Hausser, D. Chessel, and N. Perrin. Ecological niche factor analysis : How to compute habitat-suitability maps without absence data ? *Ecology*, 83 :2027–2036, 2001.
- [6] I. Noy-Meir. Data transformations in ecological ordination. i. some advantages of non-centering. *Journal of Ecology*, 61 :329–341, 1973.
- [7] L. Orłóci. Geometric models in ecology. i. the theory and applications of some ordination methods. *Journal of Ecology*, 54 :193–215, 1966.
- [8] L. Orłóci. Data centering : a review and evaluation with reference to component analysis. *Systematic Zoology*, 16 :208–212, 1967.
- [9] F. Rojas-Camousseight. *Etudes préliminaires sur l'utilisation des Trichoptères adultes comme descripteurs écologiques*. PhD thesis, 1985.
- [10] P. Usseglio-Polatera. *Evolution des peuplements de Trichoptères et d'Ephéméroptères du Rhône à Lyon (1959-1982) : résultats de piégeages lumineux*. PhD thesis, 1985.
- [11] P. Usseglio-Polatera and Y. Auda. Influence des facteurs météorologiques sur les résultats de piégeage lumineux. *Annales de Limnologie*, 23 :65–79, 1987.