


Fiche TD avec le logiciel  : ter7

Introduction à l'analyse RLQ

S. Dray

La fiche présente les principes d'une méthode d'analyse à trois tableaux utilisant un tableau d'AFC comme lien entre deux tableaux quelconques. La mise en oeuvre de la fonction `rlq` de la librairie `ade4` est détaillée.

Table des matières

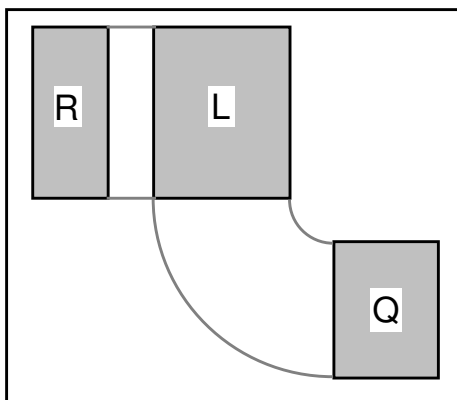
1	Introduction	2
2	Les données	2
3	Le problème	4
4	Les trois analyses simples	5
4.1	L'analyse des correspondances centrale	5
4.2	L'analyse du milieu	8
4.3	L'analyse des traits biologiques	10
5	Les analyses de co-inertie	11
6	L'analyse RLQ	12
	Références	18

1 Introduction

La fiche présente les principes de la méthode RLQ (R pour tableau en mode R, Q pour tableau en mode Q, et L tableau du lien).

Les données analysées forment un triplet de tableaux. Le premier sera appelé **L** parce qu'il sert de lien aux deux autres. Il a I lignes et J colonnes. **R** désigne le second tableau associé au précédent par les lignes. Il a I lignes et p colonnes. **Q** est le troisième, il a q colonnes. Une colonne de **L** est appariée à une seule ligne de **Q**. Comme **L** a J colonnes, **Q** a J lignes.

Dans l'exemple traité les tableaux **R** et **Q** sont des tableaux de variables qualitatives et constituent donc un cas particulier [Dolédec et al., 1996]. I vaut 51 et désigne 51 relevés d'avifaune. p vaut 11 et désigne 11 variables qualitatives décrivant l'environnement autour du relevé. J vaut 40 et désigne 40 espèces d'oiseaux nicheurs de l'agglomération lyonnaise. q vaut 4 et désigne 4 variables qualitatives décrivant quelques traits biologiques des espèces rencontrées.



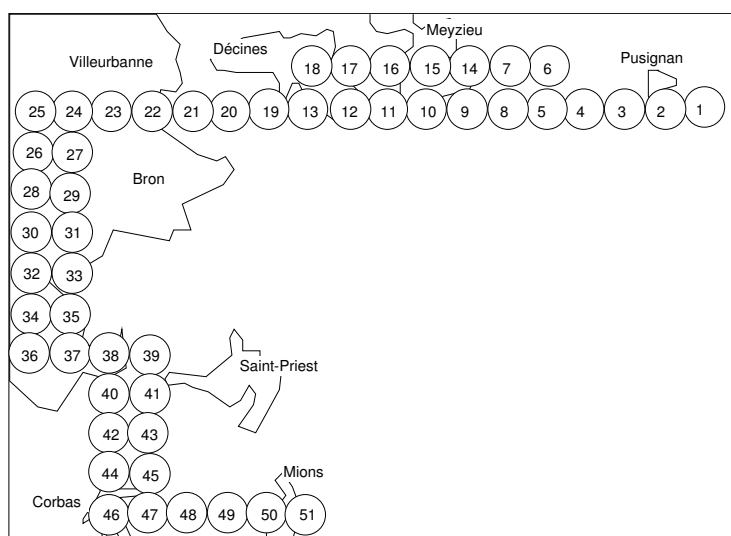
L'analyse met donc en jeu quatre types d'objets, respectivement, les relevés (lignes de **L** et de **R**), les espèces (lignes de **Q** et colonnes de **R**), les variables environnementales (colonnes de **R**) et les variables biologiques (colonnes de **Q**).

2 Les données

Les noms des 40 espèces sont dans l'élément `species.names.fr`

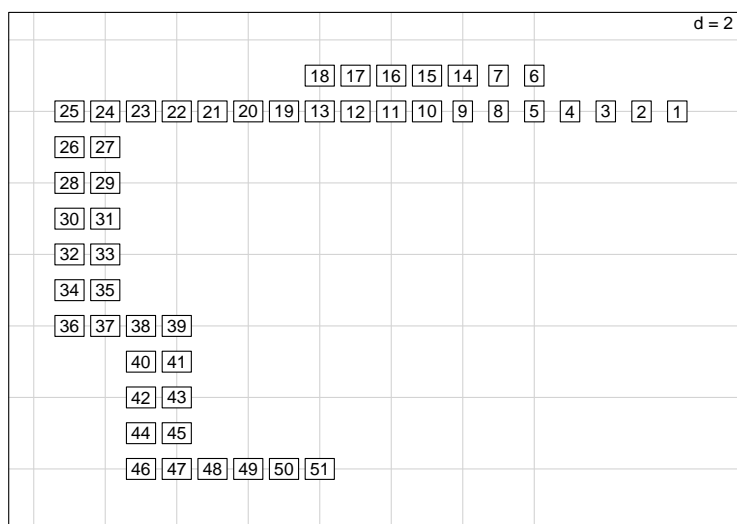
```
library(ade4)
data(aviurba)
head(aviurba$species.names.fr)
[1] "Busard cendré"      "Faucon crecerelle"  "Mouette rieuse"
[4] "Pigeon domestique" "Tourterelle des bois" "Tourterelle turque"
```

Les 51 relevés occupent dans l'espace une position décrite dans la figure suivante. Ils forment un double transect qui part de la zone périurbaine rurale, traverse les villages de l'est lyonnais, parvient au centre ville et ressort de l'agglomération vers le sud. Ils sont simplement numérotés de 1 à 51.



On reproduit l'échantillonnage

```
x <- c(18:15, rep(14, 2), 13, 13:8, 12:8, 7:3, 2:1, rep(1:2, 6),
      rep(3:4, 5), 5:8)
y <- c(rep(12, 5), rep(13, 2), rep(12, 6), rep(13, 5), rep(12, 7),
      rep(11:7, rep(2, 5)), rep(6, 4), rep(5:3, rep(2, 3)), rep(2,
      6))
xy <- cbind(x, y)
s.label(xy, addaxes = FALSE, include.origin = FALSE)
```



Les 11 variables environnementales sont qualitatives. Elles décrivent le paysage autour du relevé. Les 10 premières codent la présence d'éléments caractéristiques (**yes** avec, **no** sans) :

1. Fermes-Villages
2. Petits immeubles
3. Immeubles de + de 4 étages

4. Chantier/usines/gravières
5. Champs
6. Pelouses
7. Haies-arbustes
8. Feuillus
9. Conifères
10. Bruit permanent

La dernière donne le recouvrement de la végétation naturelle avec le code :

1. R100 : Zone de végétation naturelle
2. R98 (97 à 100) : Périurbain agricole
3. R93 (90 à 97) : Périurbain résidentiel
4. R87 (75 à 90) : Banlieue ouverte
5. R62 (50 à 75) : Banlieue fermée
6. R37 (25 à 50) : Quartiers
7. R22 (10 à 25) : Centre ville
8. R5 (0 à 10) : Zone industrielle, carrière

Les quatre variables biologiques donnent respectivement :

1. Position trophique (1-insectivore, 2-granivore, 3-polyphage)
2. Site d'alimentation (1-sol, 2-air, 3-buissons ou frondaisons)
3. Site de nidification (1-sol, 2-bâtiment, 3-buissons, 4-frondaison)
4. Statut migratoire (1-sédentaire, 2-migrants)

3 Le problème

L'analyse des ces trois tableaux vise à répondre à la question suivante [Legendre et al., 1997, Dray and Legendre, 2008] :

Y-a-t'il dans la variation des traits biologiques des espèces des correspondances avec la variation des caractéristiques de l'environnement ?

Tout tableau floro-faunistique appelle explicitement ou non de l'information externe sur ses deux marges. Sur la marge relevés, on a la problématique classique faune/milieu. Sur la marge espèces, il existe une grande quantité d'information disponible sur la biologie et l'écologie des groupes étudiés. On reconnaît alors dans les données deux problèmes de co-structures :

1. Le premier relie la typologie de relevées induite par les variations de l'abondance des espèces avec la typologie des relevés induite par les variations des variables mésologiques. Des centaines d'articles d'écologie statistique sont consacrées depuis cinquante ans à cette question.
2. Le second relie la typologie des espèces induites par la diversité de leurs modes de distribution entre les relevés et la typologie des mêmes espèces faite par la diversité des choix stratégiques mis en oeuvre pour assurer leur pérennité.

Entre les deux problèmes on reconnaît la partie commune, c'est-à-dire la dualité entre typologie d'espèces et typologie de relevés inhérente à tout tableau écologique. En prenant en compte les deux tableaux associés aux marges, pour des raisons théoriques et pratiques, il est souhaitable d'utiliser dans le couple **R-L** comme dans le couple **Q-L^T** une méthode statistique cohérente.

De ce fait, il y a cinq analyses de base sous-jacentes à la question posée : les trois typologies simples (**L** : relevés-espèces, **R** : relevés-variables mésologiques, **Q** : espèces-variables biologiques) et les deux typologies croisées (**R-L** espèces-variables mésologiques par l'intermédiaire des relevés, qui fonde la théorie des profils écologiques ; **L-Q** relevés-variables biologiques par le biais des espèces, qui fonde les théories modernes de l'habitat).

Nous allons montrer qu'on peut faire ces cinq analyses simultanément et rendre compte de la cohérence du total.

4 Les trois analyses simples

4.1 L'analyse des correspondances centrale

Du fait des deux tableaux associés aux deux marges ligne et colonne, l'exigence de symétrie sur l'analyse du tableau central est forte. L'analyse des correspondances (AFC) s'impose.

La corrélation entre deux variables qualitatives est mesurée par le χ^2 calculé sur la table de contingence :

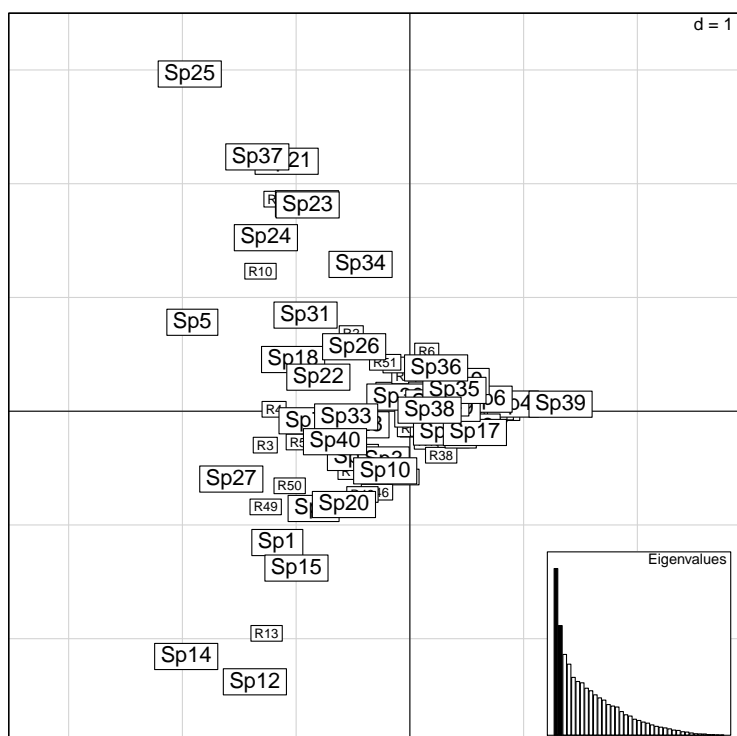
```
test1 <- chisq.test(aviurba$fau)
test1$statistic
X-squared
2005.024
```

L'AFC décompose ce lien sur un ensemble d'axes :

```
afc1 <- dudi.coa(aviurba$fau, scannf = FALSE, nf = 2)
sum(afc1$eig) * sum(aviurba$fau)
[1] 2005.024
```

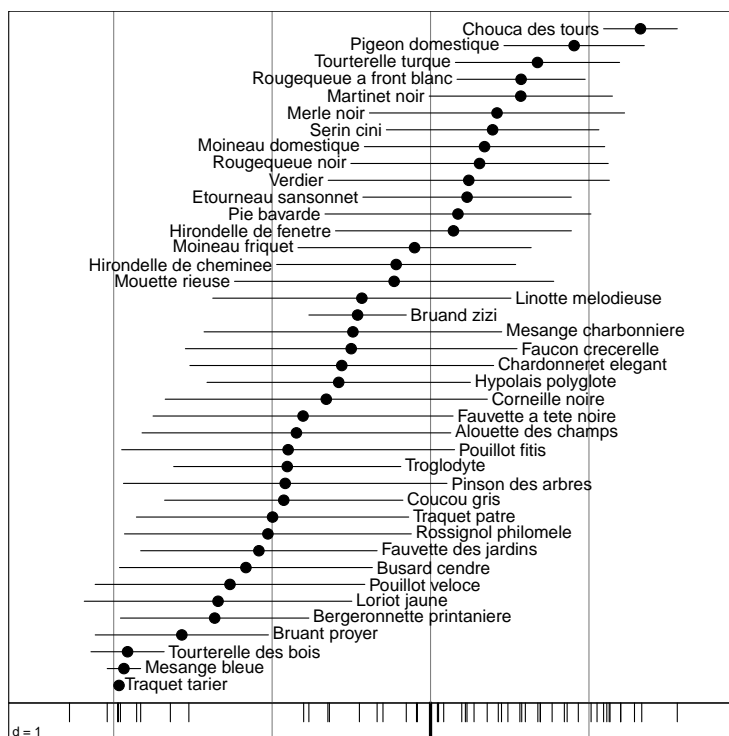
Le plan des deux premiers axes de l'analyse :

```
scatter(afc1, posieig = "bottomright")
```



Parmi les interprétations possibles, l'AFC donne un score de variance unité pour les sites. Les espèces sont alors placées par averaging.

```
sco.distri(score = afc1$l1[, 1], df = aviurba$fau, labels = aviurba$species.names.fr)
```

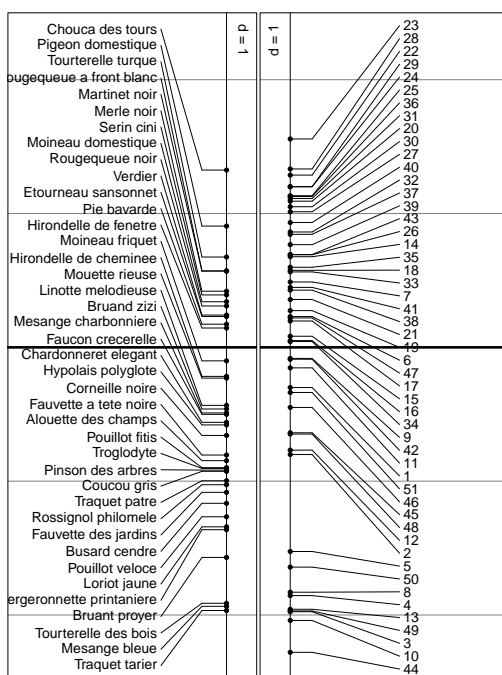


L'ordination des espèces pose la question du lien avec leurs caractéristiques biologiques.

```

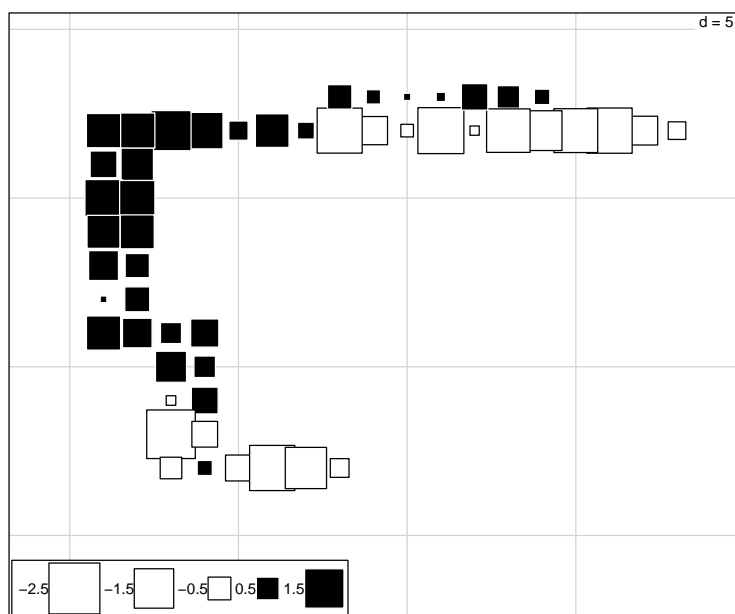
par(mfrow = c(1, 2))
sco.label(afc1$co[, 1], lab = aviurba$species.names.fr, reverse = TRUE,
  boxes = FALSE, horizontal = FALSE, lim = c(-2.5, 2.5), pos.lab = 0.3)
sco.label(afc1$l1[, 1], horizontal = FALSE, lim = c(-2.5, 2.5),
  boxes = FALSE)

```



L'ordination des sites est clairement spatialisée et pose la question du lien avec le milieu :

```
s.value(xy, afc1$l1[, 1], addaxes = FALSE)
```



4.2 L'analyse du milieu

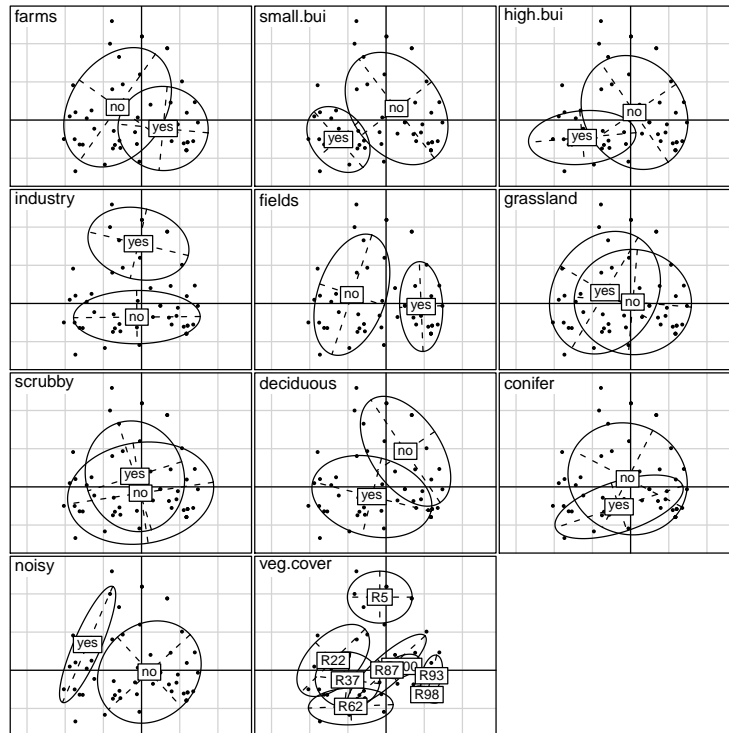
On utilise une analyse des correspondances multiples. En prévision du couplage avec les données espèces, on pondère les lignes avec les poids issus de

l'AFC :

```
acm.mil <- dudi.acm(aviurba$mil, row.w = afc1$lw, scannf = FALSE,
  nf = 4)
```

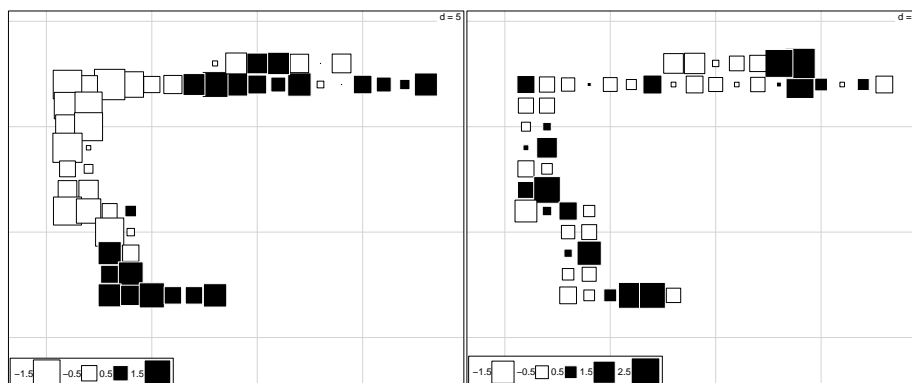
La typologie des variables :

```
scatter(acm.mil)
```



On reconnaît sans peine l'opposition ville-campagne sur le facteur 1 et la redondance des modalités Pas de végétation et Usines-gravières caractéristiques des zones industrielles péri-urbaines sur le facteur 2. On décrit clairement le gradient urbain-rural :

```
par(mfrow = c(1, 2))
s.value(xy, acm.mil$li1[, 1], addaxes = FALSE)
s.value(xy, acm.mil$li1[, 2], addaxes = FALSE)
```



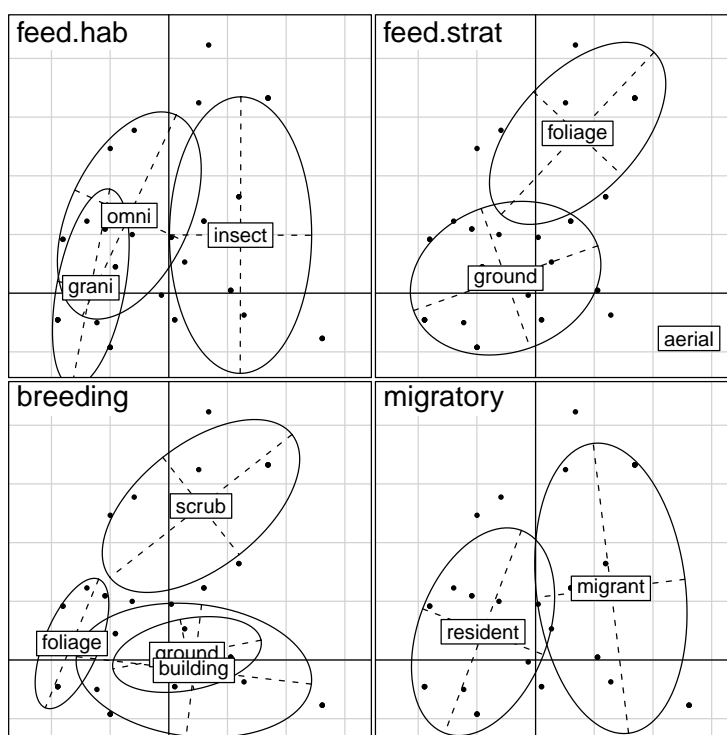
4.3 L'analyse des traits biologiques

L'analyse préliminaire des traits biologiques.

```
acm.traits <- dudi.acm(aviurba$traits, row.w = afc1$cw, scannf = FALSE,
  nf = 4)
```

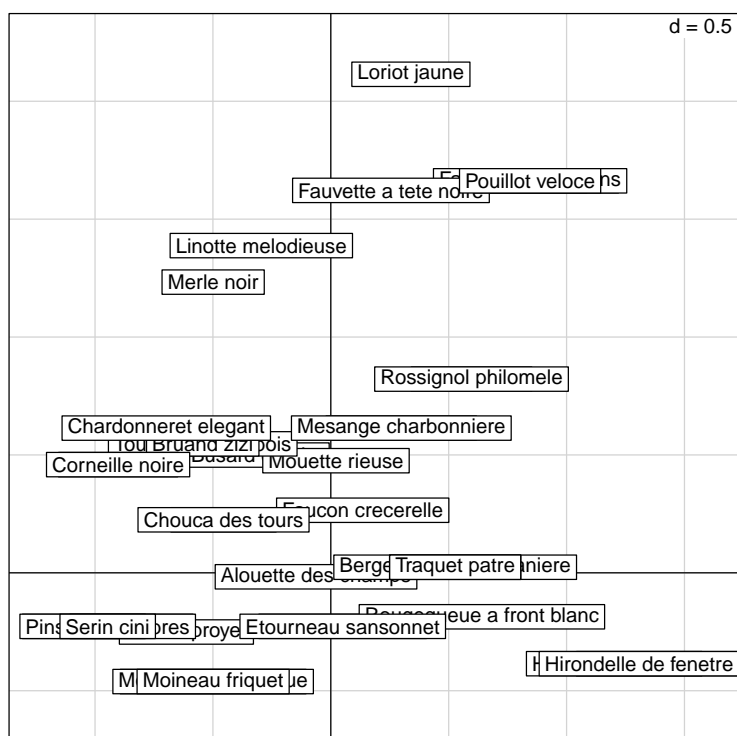
La typologie des variables :

```
scatter(acm.traits)
```



On reconnaît, même sur un cortège faunistique réduit et un ensemble de variables élémentaire, la diversité des stratégies mises en oeuvre. L'association insectivore-migrateur est la plus forte.

```
s.label(acm.traits$li, label = aviurba$species.names.fr)
```



5 Les analyses de co-inertie

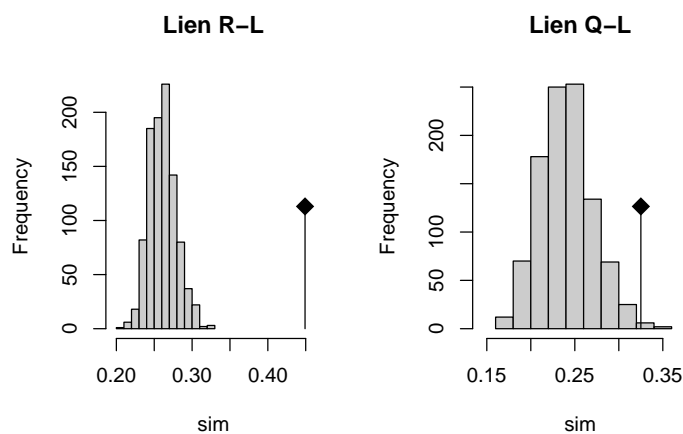
Les analyses précédentes posent les questions du lien entre variables de milieu et cortège faunistique et du lien entre traits biologiques et mode de répartition. On peut y répondre à l'aide des deux analyses de co-inertie :

```

coi1 <- coinertia(afc1, acm.mil, scannf = FALSE, nf = 2)
afc2 <- dudi.coa(t(aviurba$fau), scannf = FALSE, nf = 2)
coi2 <- coinertia(afc2, acm.traits, scannf = FALSE, nf = 2)
test1 <- randtest(coi1, fixed = 1)
Warning: non uniform weight. The results from permutations
are valid only if the row weights come from the fixed table.
The fixed table is table X : aviurba$fau
test2 <- randtest(coi2, fixed = 1)
Warning: non uniform weight. The results from permutations
are valid only if the row weights come from the fixed table.
The fixed table is table X : t(aviurba$fau)

par(mfrow = c(1, 2))
plot(test1, main = "Lien R-L")
plot(test2, main = "Lien Q-L")

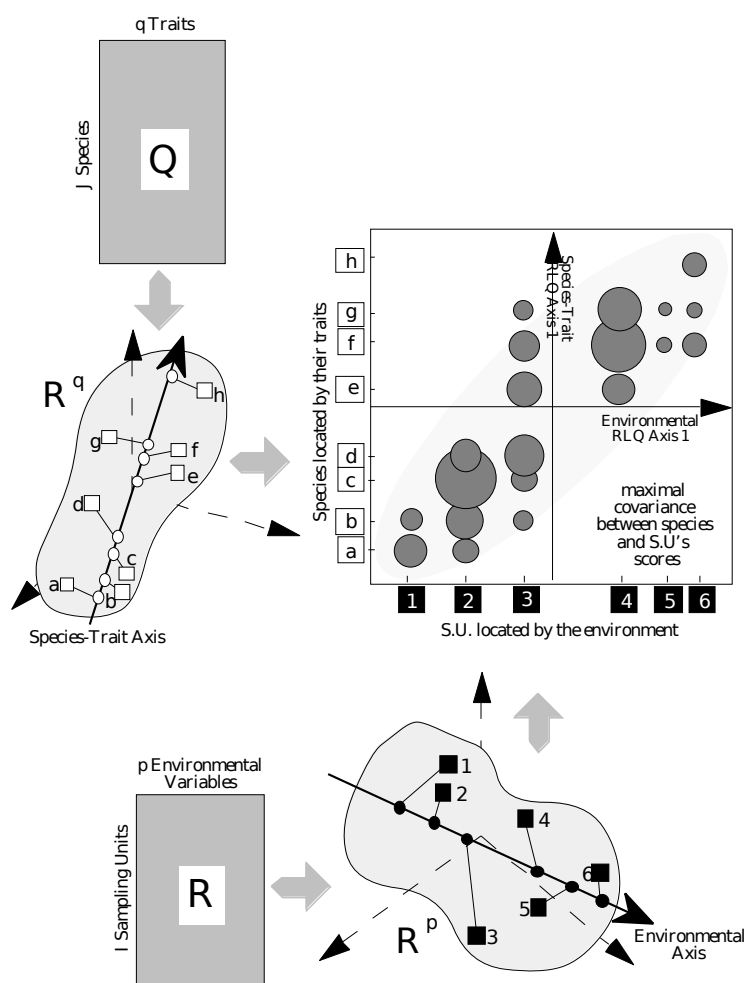
```



Les deux résultats sont significatifs, le lien entre variables de milieu et cortège faunistique est beaucoup plus net que le lien entre traits biologiques et mode de répartition. Tous les liens établis entre les trois tableaux étant significatifs, on cherche maintenant à les synthétiser dans une seule représentation.

6 L'analyse RLQ

L'analyse RLQ étend le principe de co-inertie pour analyser le lien entre les trois tableaux. Le principe est décrit dans la figure suivante :

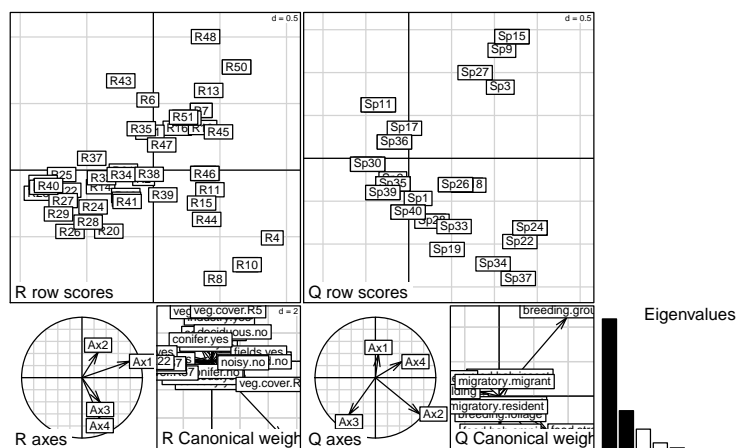


L'analyse recherche des coefficients pour les traits qui donnent un score des espèces et pour les variables de milieu qui donnent un score des sites. Ce couple maximise la covariance exprimée au travers du tableau L .

```
rlq1 <- rlq(acm.mil, afc1, acm.traits, scannf = F, nf = 2)
```

On retrouve les éléments d'interprétation dans la figure globale :

```
plot(rlq1)
```



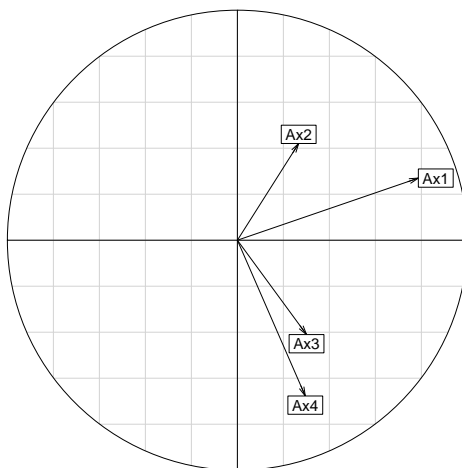
C'est le site de nidification et le recouvrement de la végétation qui font l'essentiel du lien.

L'analyse RLQ a de nombreuses aides à l'interprétation. En maximisant une covariance, elle cherche à maximiser variances et corrélations, donc à faire partiellement une analyse d'inertie de **R**, de **Q** et de **L**. On retrouve ce point de vue dans :

```
summary(rlq1)
Eigenvalues decomposition:
      eig      covar      sdR      sdQ      corr
1 0.010870058 0.10425957 0.4698488 0.5791786 0.3831293
2 0.003223881 0.05677923 0.3943920 0.6096836 0.2361331
Inertia & coinertia R:
      inertia      max      ratio
1 0.2207579 0.2724464 0.8102800
12 0.3763030 0.4859191 0.7744148
Inertia & coinertia Q:
      inertia      max      ratio
1 0.3354478 0.6507848 0.5154512
12 0.7071619 1.0507454 0.6730098
Correlation L:
      corr      max      ratio
1 0.3831293 0.6435487 0.5953384
2 0.2361331 0.5220054 0.4523576
```

Pour R, on compare les deux points de vue en projetant les axes d'inertie de l'analyse simple sur les axes de l'analyse RLQ. Il n'y a pas identité entre les deux, mais un point de vue original bien que contenant en partie l'axe 1 de l'analyse simple.

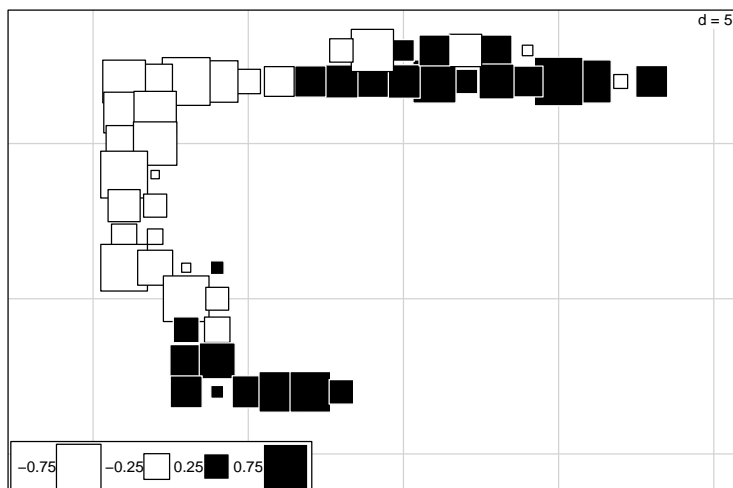
```
s.corcircle(rlq1$aR)
```



Comme en analyse de co-inertie on peut discuter de la valeur des axes RLQ (*Inertia & Coinertia R*). L'inertie projetée sur les axes RLQ (0.2208 pour l'axe 1 et 0.3763 pour le plan 1-2) est assez loin de l'optimum atteint sur les axes d'inertie (0.2724 et 0.4859).

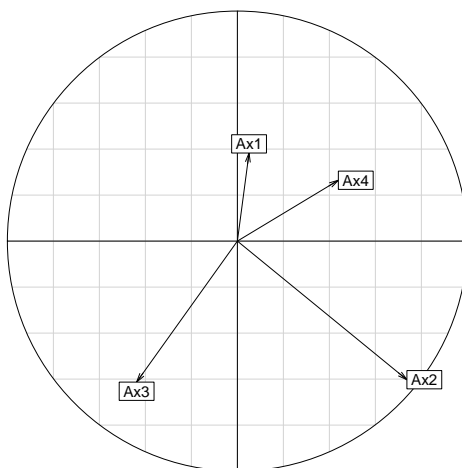
On peut cartographier le score des sites :

```
s.value(xy, rlq1$1R[, 1], addaxes = FALSE)
```



Curieusement cette coordonnée RLQ est plus nettement caractéristique de la zone proprement urbaine que la première coordonnée de l'ACM de **R**. Le lien avec les variables biologiques décrit donc essentiellement la rupture ville/banlieue. Pour **Q** (*Inertia & Coinertia Q*), l'inertie projetée sur les axes RLQ (0.3354 et 0.7072) peut être comparée à l'optimum (0.6508 et 1.0507). Délibérément, l'axe 1 (insectivore/migrateur) de l'analyse simple est abandonné dans l'analyse RLQ. La direction choisie dans l'espace est originale. Sur cette carte on s'est éloignée ainsi de l'analyse de **Q** :

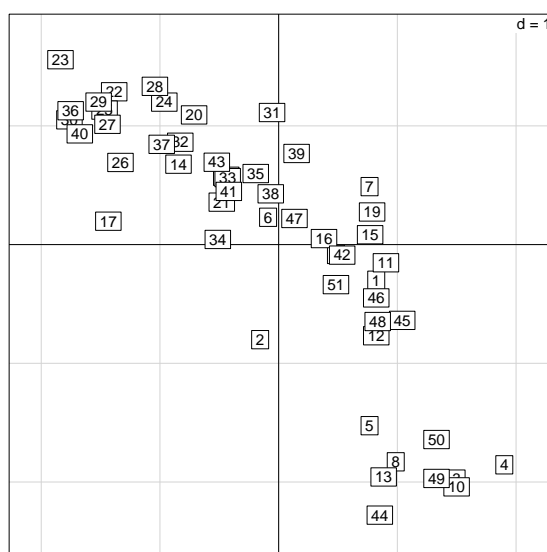
```
s.corcircle(rlq1$aQ)
```



En maximisant la covariance, on maximise en partie la corrélation. Hors l'optimum de la corrélation est obtenu par la racine carrée de la première valeur propre de l'AFC [Williams, 1952]. On peut comparer et voir que les coordonnées factorielles sur les premiers axes RLQ sont loin d'être optimaux en termes de coordonnées factorielles (cf. *Correlation L*).

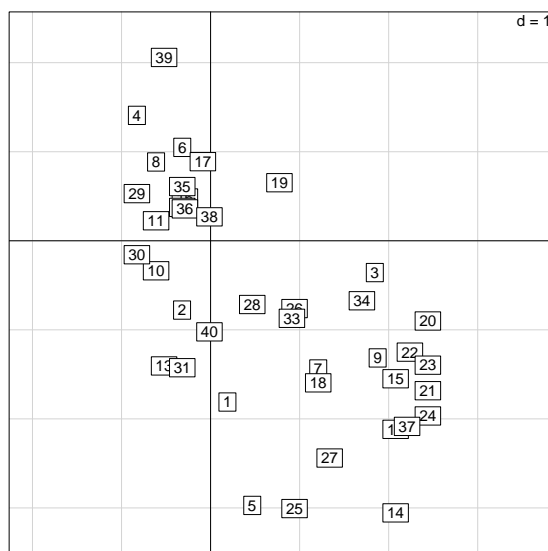
Pourtant, il y a bonne corrélation manifeste (-0.817 avec les poids marginaux) entre la première coordonnée normalisée des lignes de l'AFC de \mathbf{L} et la première coordonnée normalisée des lignes de \mathbf{R} sur l'axe 1 de l'analyse RLQ :

```
t(rlq1$mR[, 1]) %*% diag(afc1$lw) %*% as.matrix(afc1$l1[, 1])
[1,] -0.8171083
s.label(cbind(rlq1$mR[, 1], afc1$l1[, 1]))
```



Cela signifie que l'on trouve dans le tableau \mathbf{R} une combinaison de variables reconstituant à peu près l'ordination des relevés faite par le cortège faunistique et cela n'est pas surprenant. Le gradient urbain-rural s'exprime tant par la faune que par les indications de milieu. Si on a perdu sur la corrélation canonique, c'est que l'enregistrement des variables biologiques est plus sommaire. On n'y trouve pas de quoi reconstituer l'ordination des espèces faites par les relevés (ce qui est parfaitement cohérent avec les remarques précédentes). La corrélation (-0.723 avec les poids marginaux) entre la première coordonnée normalisée des colonnes de l'AFC de \mathbf{L} et la première coordonnée normalisée des lignes de \mathbf{Q} sur l'axe 1-RLQ est cependant non négligeable :

```
t(rlq1$mQ[, 1]) %*% diag(afc1$cw) %*% as.matrix(afc1$c1[, 1])
      [,1]
[1,] -0.7230557
s.label(cbind(rlq1$mQ[, 1], afc1$c1[, 1]))
```



Références

- S. Dolédec, D. Chessel, C.J.F. Ter Braak, and S. Champely. Matching species traits to environmental variables : a new three-table ordination method. *Environmental and Ecological Statistics*, 3 :143–166, 1996.
- S. Dray and P. Legendre. Testing the species traits-environment relationships : the fourth-corner problem revisited. *Ecology*, 89 :3400–3412, 2008.
- P. Legendre, R. Galzin, and M.L. Harmelin-Vivien. Relating behavior to habitat : solutions to the fourth-corner problem. *Ecology*, 78(2) :547–562, 1997.
- E.J. Williams. Use of scores for the analysis of association in contingency tables. *Biometrika*, 39 :274–289, 1952.