20. RLQ and Fourth Corner Analysis

FISH 560: Applied Multivariate Statistics for Ecologists

Topics
- RLQ Analysis
- Fourth-corner Analysis

R Packages: ade4, vegan, dummies
**R Source:**   biostats

---

BACKGROUND

The challenge of associating species traits and environmental variables using species abundance or presence/absence data is known as the fourth-corner problem in ecology (Legendre et al. 1997). The fourth-corner problem can be thought of as a three table problem, which takes matrices of environmental data (R), species abundance or presence/absence data (L) and species trait data (Q), and uses these three tables to infer how species traits relate to the environment (D) (Brown et al. 2014). Two strategies for the fourth-corner problem have been proposed.

RLQ analysis is an exploratory ordination approach to the fourth-corner problem (Doledec et al. 1996). Pairs of ordinations are jointly constructed via a generalized singular value decomposition, which can be understood as jointly relating each of the environmental variables and species traits to species abundance or presence/absence. This can provide a broad qualitative overview of how traits and environmental variables are associated, as opposed to specific details.

A hypothesis testing approach to the fourth-corner problem has also been proposed (Legendre et al. 1997). In this approach, the fourth-corner problem is considered as a matrix algebra problem, and R, L and Q are used to determine a matrix D (the matrix product of R'LQ) which summarizes the association between environmental variables and species traits. Permutation tests are then used to test hypotheses about the environment–trait association. Let's explore both of these approaches in this tutorial.
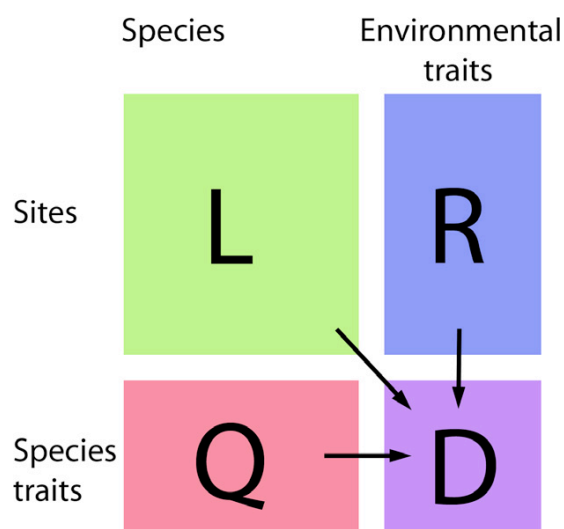
SET-UP

In this exercise you will be working with the MAHA species abundance dataset, environment dataset, and species trait dataset. But first remember to set-up your R work session by defining the current work directory to your folder of choice and loading the appropriate libraries. Also, make sure to source the BIOSTATS file from the *File* pull-down menu. You can also do this using the functions **setwd**, **library** and **source**.

Let's import the data, by typing:

```
speabu <- read.csv('MAHA_speciesabu.csv',header=TRUE, row.names=1)
envdata <- read.csv('MAHA_environment.csv',header=TRUE, row.names=1)
spetra <- read.csv('MAHA_speciestrait.csv',header=TRUE, row.names=1)
```

Let's take a quick look at the species dataset to remind ourselves of the data structure.

```
str(spetra)
```

```
'data.frame':    36 obs. of  8 variables:
 $ MaxLen  : int  78 180 100 410 260 100 860 180 300 360 ...
 $ TempPref: Factor w/ 5 levels "Cold","Cold/cool",..: 5 4 3 5 3 4 1 4 3 3
 $ SubPref : Factor w/ 4 levels "Rubble","Sand",..: 1 1 1 4 1 4 1 1 1 2 ...
 $ VertPos : Factor w/ 2 levels "Benthic","Non-benthic": 1 1 1 2 1 2 2 2 2 2
 $ Trophic : Factor w/ 5 levels "Herbivore-detritivore",..: 2 2 4 3 4 4 3 4 3
 $ Long    : int  4 6 4 5 11 6 15 7 9 7 ...
 $ Maturity: num  1.5 2 2 2.5 1.5 1.2 2.5 2 2 1.5 ...
 $ Fecund  : int  156 692 4758 40950 755 5382 3456 2170 4327 45257 ...
```

You will see that each species (n=36) has information associated with 8 traits. Some traits are quantatitive (MaxLen: maximum length; Long: longevity) and others are categorical (TempPref: temperature preference; Trophic: trophic guild).

Now we need to transform the various datasets before diving into the analysis.

First, let's recode the trait matrix so that the categorical traits (currently text) are presented as dummy variables (i.e., 0 and 1 to indicate the absence or presence of each trait state).

```
spetra.tran <- dummy.data.frame(spetra)
```

Type `str(spetra.tran)` to take a quick look. Next, let's standardize the traits by re-scaling by minimum-to-maximum following the recommendation of McCune (2015).

```
spetra.tran <- data.stand(spetra.tran,method='range',margin='column')
```

Second, let's conduct a 'row normalization' (i.e., to rescale each row so that the sums equal 1, or in other words to calculate relative species abundances), type the following:

```
speabu.tran<-data.stand(speabu,method='total',margin='row',plot=F)
```

Finally, let's log (based 10) transform the environmental dataset.

```
envdata.tran<-data.trans(envdata,method='log',plot=F)
```

CONDUCTING RLQ ANALYSIS

A preliminary step of RLQ analysis is to perform the separate analyses of each data matrix. In this tutorial we will use the rlq() function in the ade4 library. Its usage is:

```
rlq (dudiR, dudiL, dudiQ, …)
```

Where a selection of the arguments include:

dudiR	a duality diagram providing from one of the functions dudi.hillsmith, dudi.pca, …
dudiL	a duality diagram of the function dudi.coa
dudiQ	a duality diagram providing from one of the functions dudi.hillsmith, dudi.pca, …
scannf	a logical value indicating whether the eigenvalues bar plot should be displayed
…	

For additional documentation type **?rlq**.

Let's perform the RLQ analysis.

First, correspondence analysis is applied to the species relative abundance matrix by typing:

L.species <- dudi.coa(speabu.tran, scannf = FALSE)

Second, principal component analysis is applied on the quantitative trait matrix by typing:

Q.trait <- dudi.pca(spetra.tran, row.w = L.species$cw, scannf = FALSE)

Third, principal component analysis is applied on the environmental matrix by typing:

R.env <- dudi.pca(envdata.tran, row.w = L.species$lw,scannf = FALSE)

Lastly, to complete the RLQ analysis, separate analyses of traits and environmental variables should be weighted by the sites and species weights derived from the previous correspondence analysis.

rlq.MAHA <- rlq(R.env, L.species, Q.trait, scannf = FALSE)

RLQ analysis finds coefficents (in $c1) to obtain a linear combination of traits (species scores in $lQ) and coefficients (in $l1) to obtain a linear combination of environmental variables (site scores in $lR). The covariance between these two sets scores is maximized and equal to the square root of the corresponding eigenvalue.

As RLQ analysis maximizes the covariance between the traits and the environmental variables mediated by the species abundances, it is important to see how the individual parts (i.e. cov(traits, env)^2 = var(traits) × var(env) × cor(traits, env)^2) of the compromise are considered. It is possible to compare the RLQ analysis to the separate analyses which maximize independently the structure of the trait (PCA of the traits), the structure of the environment (PCA of the environmental variables) and the correlation (CA of the sites-species matrix). These comparisons are provided by the summary function.

summary(rlq.MAHA)

```
RLQ analysis

Class: rlq dudi
Call: rlq(dudiR = R.env, dudiL = L.species, dudiQ = Q.trait, scannf = FALSE)

Total inertia: 3.06

Eigenvalues:
    Ax1     Ax2     Ax3     Ax4     Ax5
1.76265 0.70484 0.31385 0.13984 0.06701

Projected inertia (%):
    Ax1     Ax2     Ax3     Ax4     Ax5
  57.60   23.03   10.26    4.57    2.19

Cumulative projected inertia (%):
    Ax1   Ax1:2   Ax1:3   Ax1:4   Ax1:5
  57.60   80.64   90.89   95.46   97.65

(Only 5 dimensions (out of 10) are shown)
```

```
Eigenvalues decomposition:
        eig    covar      sdR      sdQ       corr
1 1.7626529 1.327649 1.679339 1.573950 0.5022898
2 0.7048441 0.839550 1.309862 1.506642 0.4254130

Inertia & coinertia R (R.env):
    inertia       max     ratio
1  2.820178 3.149831 0.8953426
12 4.535917 5.030714 0.9016449

Inertia & coinertia Q (Q.trait):
    inertia       max     ratio
1  2.477317 4.570410 0.5420339
12 4.747288 7.617092 0.6232415

Correlation L (L.species):
       corr       max     ratio
1 0.5022898 0.8707074 0.5768755
2 0.4254130 0.8024854 0.5301193
```

From this output you can see that the variation of the environmental scores is well preserved on the first two axes (90.1%), and the variation of the trait scores are moderately preserved (62.3%). Similarly, the correlations for both CA axes (species scores) are fairly high.

Finally, a multivariate test can be applied to evaluate the global significance of the traits-environment relationships. This test is based on the total inertia of the RLQ analysis. Type:

rlqtest.MAHA <- randtest(rlq.MAHA, modeltype = 6, nrepet = 4999)

rlqtest.MAHA

```
Monte-Carlo tests
Call: randtest.rlq(xtest = rlq.MAHA, nrepet = 4999, modeltype = 6)

Number of tests:   2

Adjustment method for multiple comparisons:    none
Permutation number:   4999
     Test      Obs  Std.Obs   Alter Pvalue
1 Model 2 3.060001 5.854345 greater 0.0002
2 Model 4 3.060001 1.354139 greater 0.0980
```

This shows the results of model 2 (permutation of sites, i.e. rows) and 4 (permutation of species, i.e. columns), which should be combined to obtain a test with a correct type I error (Dray et al. 2014). The result indicates marginal statistical significant.
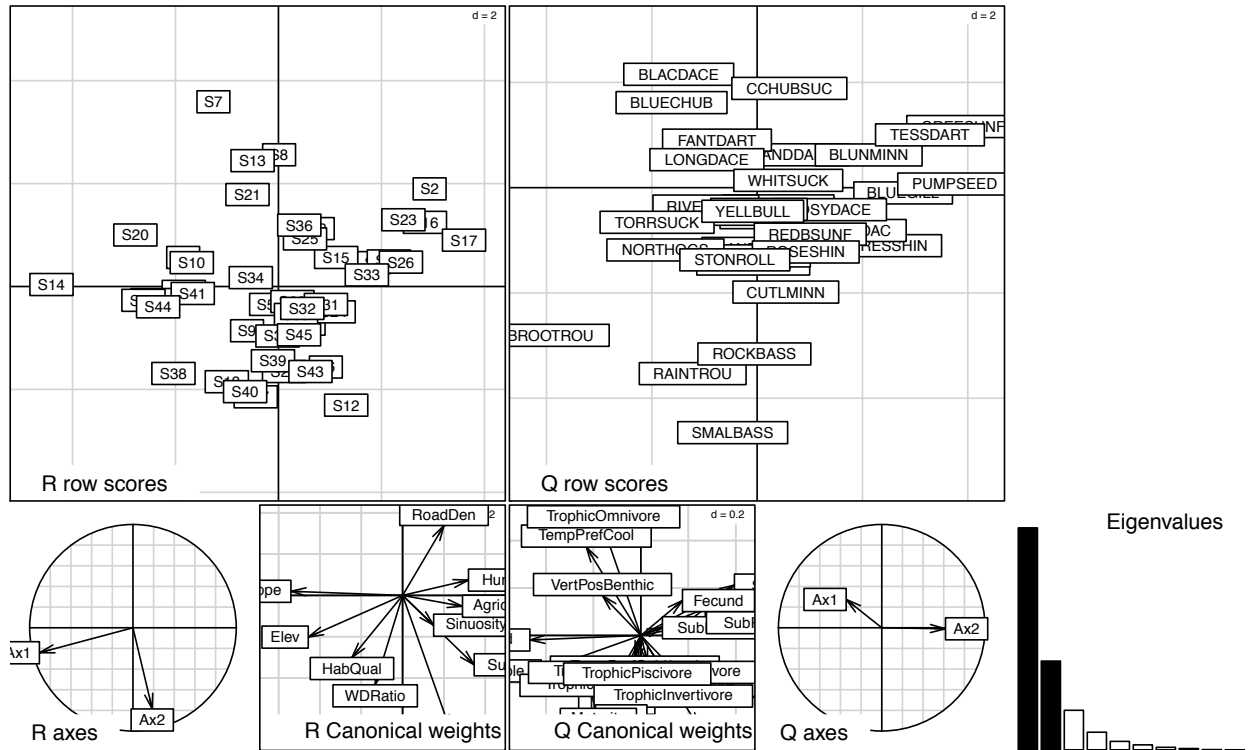
rlqtest.MAHA$comb.pvalue

```
[1] 0.098
```

You can also plot the results from the permutation test by typing **plot(rlqtest.MAHA)**.

PLOTTING THE RLQ RESULTS

The different factorial maps of the analysis are obtained by the plot function:
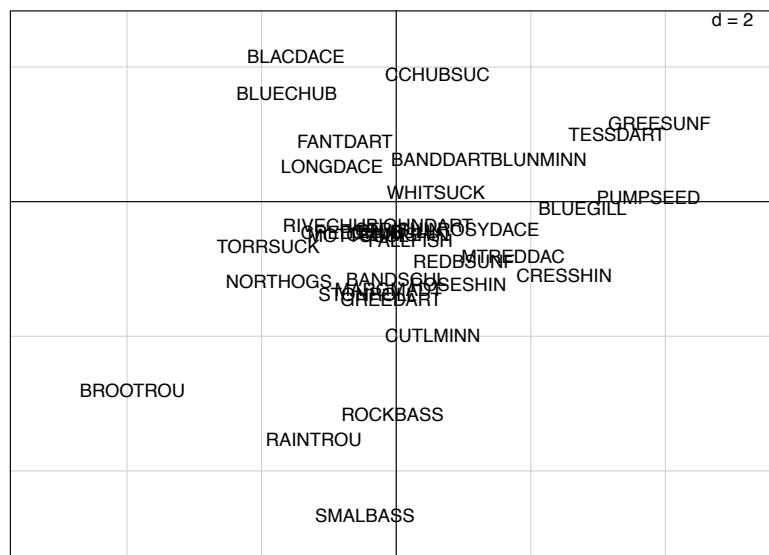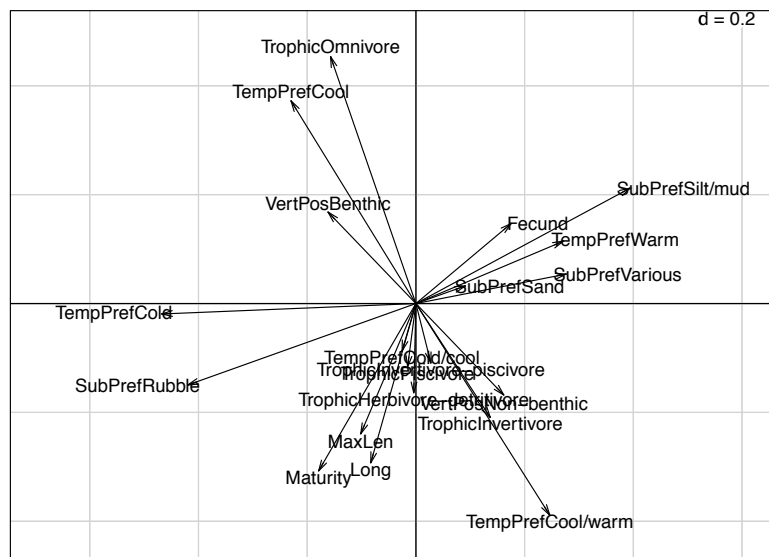
```
plot(rlq.MAHA)
```


This demonstrates the correspondence analysis of the species-by-site matrix (top panels), the R axes representing the canonical weights of the environmental variables, the Q axes representing the canonical weights of the traits, and the eigenvalues (first two shown in black).

The separate factorial maps can be obtained separately by plotting the different elements contained in the rlq.MAHA object (remember that you can type `rlq.MAHA` to list all the available elements).

```
s.arrow(rlq.MAHA$l1, boxes = FALSE)
s.arrow(rlq.MAHA$c1, boxes = FALSE)
s.label(rlq.MAHA$lQ, boxes = FALSE)
```

d = 0.2

RoadDen
HumanUse
Slope
Agricult
Sinuosity
Elev
HabQual
SubEmbed
WDRatio
BasinAre

d = 0.2

TrophicOmnivore
TempPrefCool
SubPrefSilt/mud
VertPosBenthic
Fecund
TempPrefWarm
SubPrefVarious
TempPrefCold
SubPrefSand
TempPrefCold/cool
TrophicPiscivore
TrophicHerbivore
SubPrefRubble
VertPosNon-benthic
TrophicInvertivore
MaxLen
Long
Maturity
TempPrefCool/warm

d = 2

BLACDACE
CCHUBSUC
BLUECHUB
GREESUNF
TESSDART
FANTDART
LONGDACE
BANDDART BLUNMINN
WHITSUCK
PUMPSEED
BLUEGILL
RIVECHUB RICHDART
TORRSUCK FALLFISH
REDBSUNF MTREDDAC
NORTHOGS BANDSCUL CRESSHIN
STNEREDART
GREEDART
CUTLMINN
BROOTROU
ROCKBASS
RAINTROU
SMALBASS

FOURTH-CORNER ANALYSIS

Fourth-corner analysis will be accomplished using the fourthcorner() function in the ade4 library. Its usage is:

```
fourthcorner(tabR, tabL, tabQ, modeltype = 6, nrepet = 999,
p.adjust.method.G = p.adjust.methods, p.adjust.method.D =
p.adjust.methods, ...)
```

Where a selection of the arguments include:

tabR	a dataframe with the measurements of m environmental variables (columns) for the n sites (rows)
tabL	a dataframe containing the abundances of p species (columns) at n sites (rows)
tabQ	a dataframe describing s species traits (columns) for the p species (rows)
modeltype	an integer (1-6) indicating the permutation model used in testing procedure.
nrepet	the number of permutations
p.adjust.method.G	a string indicating a method for multiple adjustment used for output tabG
p.adjust.method.D	a string indicating a method for multiple adjustment used for output tabD/tabD2

For additional documentation type **?fourthcorner**.

Let's perform the fourth-corner analysis. To obtain a test with a correct type I error, modeltype argument is set to 6 (following Dray et al. 2014). Note that a very high number of repetitions may be required to have enough power, but this is time-consuming and could be modified to speed up the different analyses (e.g., nrepet <- 999). Note that by default, the nrepet argument of the fourthcorner function is set to 999.

fourthcor.MAHA <- fourthcorner(envdata.tran, speabu.tran,spetra, modeltype = 6, p.adjust.method.D = "none", nrepet = 4999)

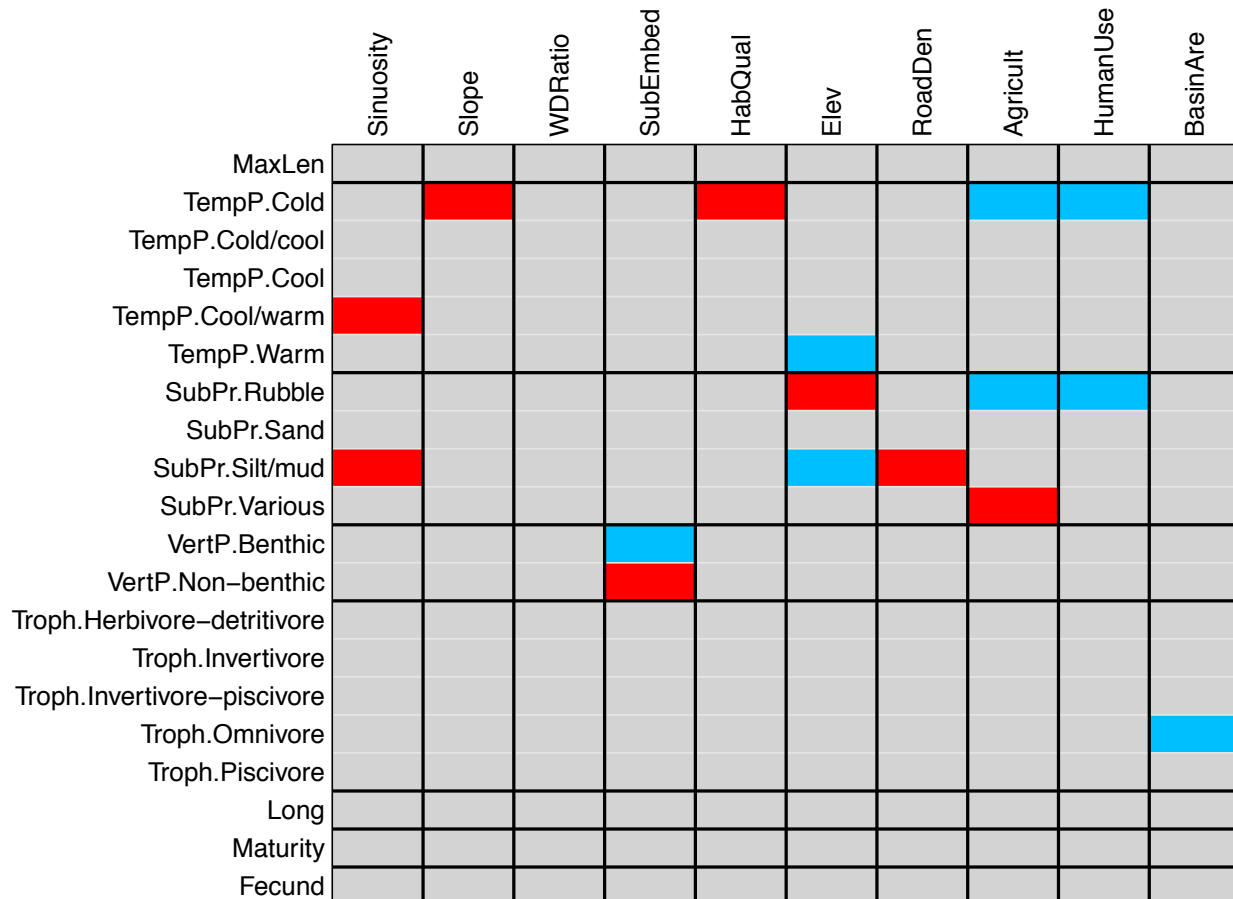**fourthcor.MAHA**

Test	Stat		Obs	Std.Obs	N.perm	Pvalue
1	Sinuosity / MaxLen	r	-0.0321131413	-0.328924753	4999	0.723
2	Slope / MaxLen	r	0.1498299096	1.218982865	4999	0.219
3	WDRatio / MaxLen	r	0.0747526403	0.651665743	4999	0.517
4	SubEmbed / MaxLen	r	0.1284100284	0.922487431	4999	0.3658
5	HabQual / MaxLen	r	0.0856208885	0.937171998	4999	0.337
6	Elev / MaxLen	r	0.0139198589	0.135405142	4999	0.9016
7	RoadDen / MaxLen	r	-0.1343229420	-1.220442196	4999	0.2388
8	Agricult / MaxLen	r	-0.0691995456	-0.764822439	4999	0.4742
9	HumanUse / MaxLen	r	-0.0911340316	-0.998276612	4999	0.332
10	BasinAre / MaxLen	r	0.0858872581	0.468685813	4999	0.6702
11	Sinuosity / TempP.Cold	Homog.	0.00040	-0.473154610	4999	0.0928
12	Slope / TempP.Cold	Homog.	0.029415	2.928102622	4999	0.9714

…

This table displays the results from the pairwise comparions between each environmental variable and trait (trait state when dealing with categorical variables), including the p-value according to the permutation test.

We can also plot the results, by typing:

```
plot(fourthcor.MAHA, alpha = 0.05, stat = "D2")
```

	Sinuosity	Slope	WDRatio	SubEmbed	HabQual	Elev	RoadDen	Agricult	HumanUse	BasinAre
MaxLen										
TempP.Cold		red			red			blue	blue	
TempP.Cold/cool										
TempP.Cool										
TempP.Cool/warm	red									
TempP.Warm						blue				
SubPr.Rubble						red		blue	blue	
SubPr.Sand										
SubPr.Silt/mud	red					blue	red			
SubPr.Various								red		
VertP.Benthic				blue						
VertP.Non–benthic				red						
Troph.Herbivore–detritivore										
Troph.Invertivore										
Troph.Invertivore–piscivore										
Troph.Omnivore										blue
Troph.Piscivore										
Long										
Maturity										
Fecund										

Blue cells correspond to negative significant relationships while red cells correspond to positive significant relationships (this can be modified using the argument col). In this example, there are some associations between categorical traits and quantitative environmental variable which can be measured in three different ways (Legendre et al. 1997). These three methods correspond to three possible values of the stat argument in the plot and print functions:

- stat="D2": the association is measured between the quantitative variable and each category sepa-rately. A correlation coefficient is used to indicate the strength of the association between the given category and the small or large values of the quantitative variable.
- stat="G": the association between the quantitative variable and the whole categorical variable is measured by a global statistic (F).
- stat="D": the association is estimated between the quantitative variable and each category separately by a measure of the within-group homogeneity. The strength of the association is indicated by the dispersion of the values of the quantitative variable for a given category.

In the rest of the tutorial, we focus on the D2 statistic.

Next, we can adjust p-values for multiple comparisons (here we used the fdr method using the p.adjust.4thcorner function).

```
fourthcor2.MAHA <- fourthcorner(envdata.tran, speabu.tran,spetra, modeltype =
6, p.adjust.method.D = "fdr", nrepet = 4999)
```

Plot these results and take a look what happens!

COMBINING RLQ AND FOURTH-CORNER ANALYSIS

Both approaches can be combined if RLQ scores are used to represent traits and environmental variables on a biplot. Then, significant associations revealed by the fourth-corner approach can be represented using segments (blue lines for negative associations, red lines for positive associations, see the argument col). Only traits and environmental variables that have at least one significant association are represented on the factoral map below.

```
plot(fourthcor.MAHA,x.rlq=rlq.MAHA,alpha=0.05,stat="D2",type="biplot")
```
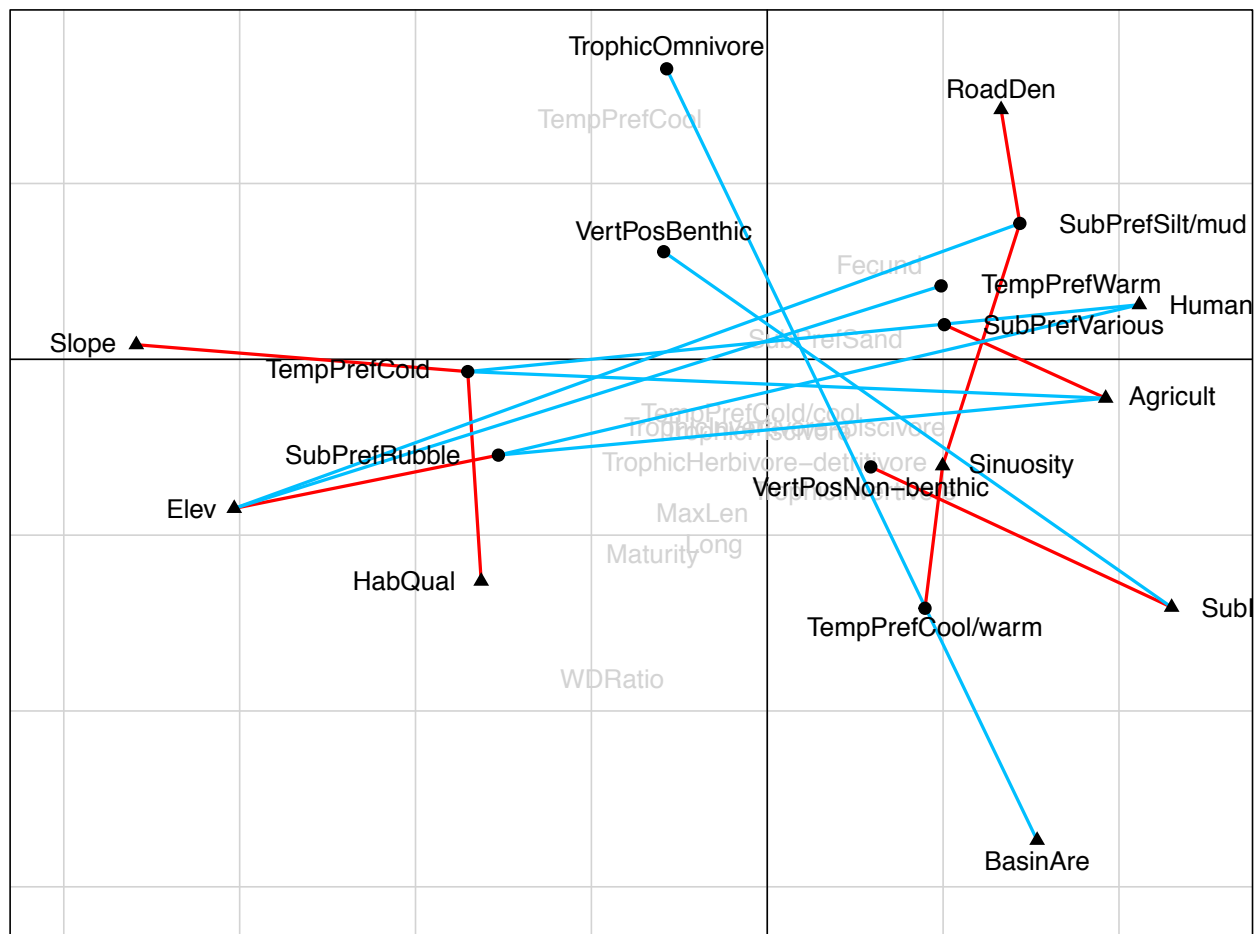

Another approach is provided by the fourthcorner.rlq function and consists in testing directly the links between RLQ axes and traits (typetest="Q.axes") or environmental variables (typetest="R.axes").

```
testQaxes<-fourthcorner.rlq(rlq.MAHA,modeltype = 6, typetest="Q.axes",
p.adjust.method.D = "fdr", nrepet = 4999)
```

```
testRaxes<-fourthcorner.rlq(rlq.MAHA,modeltype = 6, typetest="R.axes",
p.adjust.method.D = "none", nrepet = 4999)
```

Rather than exploring the detailed tables with the results, let's just plot the p-values based on no multiple comparison correct (but you could use `p.adjust.method.D = "fdr"` if you wanted).

```
par(mfrow = c(1, 2))
plot(testQaxes, alpha = 0.05, type = "table", stat = "D2")
plot(testRaxes, alpha = 0.05, type = "table", stat = "D2")
```
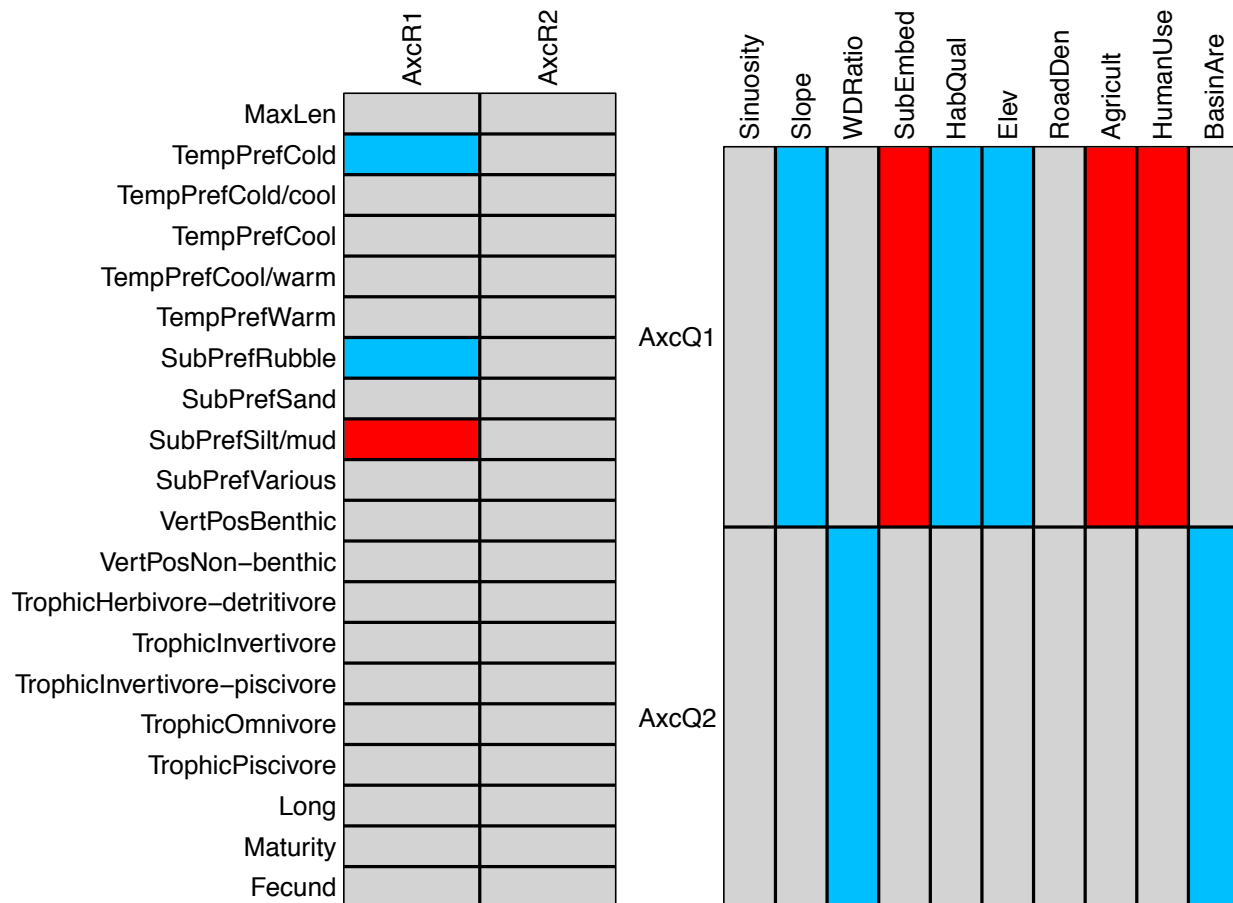

You can examine at the RLQ plots presented previously to see that this test is identifying the traits and environmental variables with the largest canonical loadings in the factoral map. In fact, significance with axes can also be reported on the factorial map of RLQ analysis. Here, significant associations with the first axis are represented in blue, with the second axis in orange, with both axes in green (variables with no significant association are in black). Type:

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

plot(testQaxes, alpha = 0.05, type = "biplot", stat = "D2", col = c("black",
"blue", "orange", "green"))

plot(testRaxes, alpha = 0.05, type = "biplot", stat = "D2", col = c("black",
"blue", "orange", "green"))
```
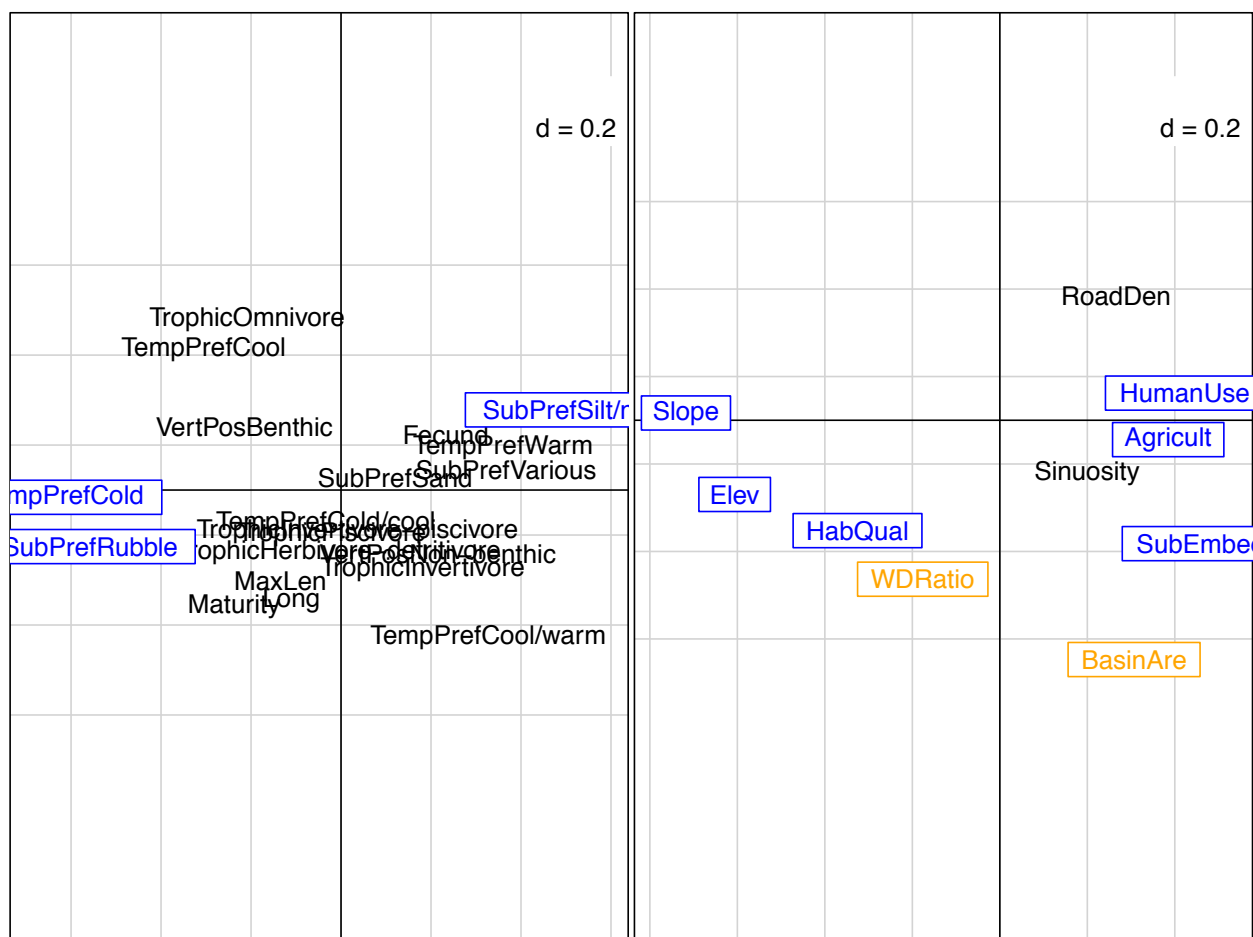
OPTIONAL READINGS

Doledec, S., Chessel, D., ter Braak, C.J.F. and Champely, S. 1996. Matching species traits to environmental variables: a new three-table ordination method. *Environmental and Ecological Statistics* **3**: 143–166.

Dray, S., Pettorelli, N., Chessel, D. 2002. Matching data sets from two different spatial samplings. *Journal of Vegetation Science* **13**: 867–874.

Dray, S. and Legendre, P. 2008. Testing the species traits-environment relationships: the fourth-corner problem revisited. *Ecology* **89**: 3400–3412.

Dray, S., P. Choler, S. Dolédec, P. R. Peres-Neto, W. Thuiller, S. Pavoine and C. J. F. ter Braak. 2014. Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation. *Ecology* **95**: 14-21.

Legendre, P., Galzin, R. and M. L. Harmelin-Vivien. 1997. Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* **78**: 547–562.

McCune, B. 2015. The front door to the fourth corner: variations on the sample unit × trait matrix in community ecology. *Community Ecology* **16**: 267-271.

ter Braak, C., Cormont, A., Dray, S. 2012. Improved testing of species traits-environment relationships in the fourth corner problem. *Ecology* **93**: 1525-1526.