

Species associations

Pierre Legendre¹ and F. Guillaume Blanchet²

¹Département de sciences biologiques, Université de Montréal

²Department of Renewable Resources, University of Alberta

Introduction

- Classical problem in ecology (Whittaker 1962) –

« Interspecific associations arise when two or more species co-occur either more or less frequently than expected due to chance alone. **Positive associations** between two species can occur when both select the same habitat or have the same environmental requirements. Conversely, **negative associations** can occur if the species have differing ecological requirements (Dale 1977).

Association, in either the positive or negative direction, can also occur as a direct consequence of **biotic interactions** such as mutualism, competition and predation. Although it is not possible to unambiguously infer the action of specific processes from the examination of patterns alone (Schluter 1984, Rejmánek and Leps 1996), association analyses remain a valuable tool for ecologists. Such analyses can be used for generating hypotheses about the factors responsible for the patterns, and hence can be used for identifying particular patterns that may be worthy of further study.

A number of statistical tests have been utilized for detecting species associations. These include correlation analysis (Greig-Smith 1983, O'Connor and Aarssen 1987, Myster and Pickett 1992), analysis by contingency table (Greig-Smith 1983, Dale et al. 1991), variance tests (Schluter 1984, McCulloch 1985), and the use of cross-variograms (Rossi et al. 1992). [...] » (Roxburgh and Chesson 1998).

Roxburgh, S. H., and P. Chesson. 1998. A new method for detecting species associations with spatially autocorrelated data. *Ecology* 79: 2180-2192.

Whittaker, R. R. 1962. Classification of natural communities. *The Botanical Review* 28: 1-239.

Search for species association

An association, in the statistical sense, is a recurrent group of co-occurring (presence-absence data) or correlated (abundance data) species.

Method (in steps): search for concordance among species (Legendre 2005)

1. Look for correlations among species: find clusters of correlated species (several methods available)
2. Compute a global test of concordance among the species in each group. Are there concordant species in these groups?
3. Test the contribution of each species to the concordance of its group.

Legendre, P. 2005. Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics* 10: 226-245.

Small R example using data from Legendre (2005, p. 230)

```
### Obtain mite data from vegan in R
### (70 soil cores x 35 species; Borcard & Legendre 1994)
library(vegan)
data(mite)

### Hellinger transformation of the species data
mite.hel <- decostand(mite, "hel")

### Extract a sub-matrix for test (10 sites, 4 species)
### (not all correlated)
mite.small <- mite.hel[c(4,9,14,22,31,34,45,53,61,69),
                      c(13,14,15,23)]
```

Table 1. Illustrative example. Upper panel: Hellinger-transformed abundances of four mite species at 10 sites selected along the long axis of Figure 5. The Hellinger transformation was computed for the full dataset (70 sites). Lower panel: the same data transformed into ranks (with ties); last column: sum of the ranks for each site.

	<i>Hellinger-transformed abundances</i>				
	<i>Species 13</i>	<i>Species 14</i>	<i>Species 15</i>	<i>Species 23</i>	
Site 4	.25087	.40538	.24380	.08362	
Site 9	.40324	.25503	.39303	.00000	
Site 14	.26577	.47620	.27267	.06097	
Site 22	.32350	.63337	.47003	.00000	
Site 31	.26312	.29089	.39223	.08771	
Site 34	.33675	.44836	.53727	.10153	
Site 45	.07956	.19487	.19487	.11251	
Site 53	.00000	.18570	.26261	.11744	
Site 61	.00000	.15430	.15430	.00000	
Site 69	.12769	.62987	.27584	.34578	
	<i>Ranks (species-wise)</i>				
	<i>Species 13</i>	<i>Species 14</i>	<i>Species 15</i>	<i>Species 23</i>	<i>Sum of ranks R_i</i>
Site 4	5	6	3	5	19.0
Site 9	10	4	8	2	24.0
Site 14	7	8	5	4	24.0
Site 22	8	10	9	2	29.0
Site 31	6	5	7	6	24.0
Site 34	9	7	10	7	33.0
Site 45	3	3	2	8	16.0
Site 53	1.5	2	4	9	16.5
Site 61	1.5	1	1	2	5.5
Site 69	4	9	6	10	29.0

```

### Perform the global test of concordance
global.res <- kendall.global(mite.small, nperm=999)

$Concordance_analysis
      Group.1
W      0.44160305
F      2.37252221
Prob.F  0.04403791
Chi2    15.89770992
Prob.perm 0.04300000

### Perform a posteriori tests of concordance
post.res <- kendall.post(mite.small, mult="holm", nperm=999)

$A_posteriori_tests
      Spec13   Spec14   Spec15   Spec23
Spearman.mean 0.3265678 0.3965503 0.4570402 -0.1681251
W.per.species  0.4949258 0.5474127 0.5927802  0.1239061
Prob           0.0880000 0.0290000 0.0060000  0.7170000
Corrected prob 0.1760000 0.0870000 0.0240000  0.7170000

```

Table 2. Results of (a) the overall and (b) the *a posteriori* tests of concordance among the mite species. P = permutational probability, based upon 9,999 random permutations. (c) Complementary Spearman correlation coefficients (r) with results of one-tailed tests of significance, and partial concordance statistics \bar{r}_j and W_j for each species j described in Section 5. * Reject H_0 at $\alpha = .05$.

<i>(a) Overall test of the W statistic. H_0: The four species are not concordant with one another</i>							
Kendall's $W =$.44160					
Friedman's chi-square =		15.89771		$P = .0448^*$		Reject H_0	
<i>(b) A posteriori tests</i>		<i>H_0: This species is not concordant with the other three</i>					
Species 13		$P = .0766$		Do not reject H_0			
Species 14		$P = .0240^*$		Reject H_0			
Species 15		$P = .0051^*$		Reject H_0			
Species 23		$P = .7070$		Do not reject H_0			
<i>(c) Spearman correlation table</i>		<i>$H_0: r = 0; H_1: r > 0$ (one-tailed test)</i>					
		Species 13	Species 14	Species 15	Species 23	\bar{r}_j	W_j
Species 13	r	1.0000	.5593	.8389	-.4185	.32657	.49493
	P	—	.0464	.0012	.8856		
Species 14	r	.5593	1.0000	.6242	.0061	.39655	.54741
	P	.0464	—	.0269	.4933		
Species 15	r	.8389	.6242	1.0000	-.0920	.45704	.59278
	P	.0012	.0269	—	.5998		
Species 23	r	-.4185	.0061	-.0920	1.0000	-.16813	.12391
	P	.8856	.4933	.5998	—		

Details of the method

1. Look for correlations among species: find clusters of correlated species (several method available)

- Principal component analysis (PCA): can distinguish two groups.
- Agglomerative clustering on correlation matrix (Pearson, Spearman) among species: the most common method (e.g. Ward agglomerative clustering).
- *K*-mean partitioning of the standardized species.

2. Compute a global test of concordance among the species in each group. Are there concordant species in the groups?

Compute the Kendall (W) coefficient of concordance:

	Species 1	Species 2	Species 3	Sum of ranks
	(ranks of abundances)			in each row
Site 1	1	1	6	8
Site 2	6	5	3	14
Site 3	3	6	2	11
Site 4	2	4	5	11
Site 5	5	2	4	11
Site 6	4	3	1	8

W is the variance of the sums of ranks divided by the maximum value that sum can take. That maximum value is obtained when there is total agreement among the judges (species) with regard to the order of the objects (sites).

$$W = \frac{12 \sum_{i=1}^n (R_i - \bar{R})^2}{p^2(n^3 - n) - pT}$$

Tied ranks should be dealt with in a special manner.

$$T = \sum_{k=1}^m (t_k^3 - t_k)$$

t_k is the number of tied ranks in each of m groups of ties. Ex. from Table 1:

Site	Rank of species 13	t_k
4	5	1
9	10	1
14	7	1
22	8	1
31	6	1
34	9	1
45	3	1
53	1.5	} 2
61	1.5	
69	4	1

W can also be computed from the mean \bar{r} of the Spearman correlations between all species in a group:

$$W = \frac{(p-1)\bar{r} + 1}{p}$$

W can be transformed to a Friedman chi-square statistic before testing:

$$\chi^2 = p(n-1)W$$

W can also be transformed to an F -statistic before testing:

$$F = \frac{(p-1)W}{1-W}$$

The results can be tested parametrically or by permutation. The groups that are not significantly concordant are dropped.

3. Test the contribution of each species to the concordance of its group.

- Compute \bar{r}_j , the mean of the Spearman correlations between species j and all the other species in the same group.
- A partial W_j statistic measuring the contribution of species j to the concordance of its group can be computed as follows:

$$W_j = \frac{(p_{gr} - 1)\bar{r}_j + 1}{p_{gr}}$$

- Permutation test: only the ranks of species j are permuted in the table before recomputing the statistic under permutations. Statistics \bar{r}_j and W_j are equivalent for this permutation test. Either one can thus be used.

⇒ If some species have negative values of \bar{r}_j , that means these species are negatively correlated to several species of the group. The group is thus too inclusive. The analysis needs to be redone with the group subdivided.

Corrections for multiple testing

When more than one test is carried out on the same data, a correction needs to be applied to the P -values resulting from these multiple tests.

Holm correction: $P_{Holm} = (m - k + 1) \times P$

Bonferonni correction: $P_{Bonferonni} = m \times P$

Sidak correction: $P_{Sidak} = 1 - (1 - P)^m$

=====

Numerical simulations (Legendre 2005)

- Correct type I error: when the species are not correlated, the test rejects H_0 in a proportion of the cases equal to the significance level (e.g. 5%). This is true for both the F and χ^2 statistics.

- When species are correlated, the power of the test increases with the number of intercorrelated species included in the analysis.

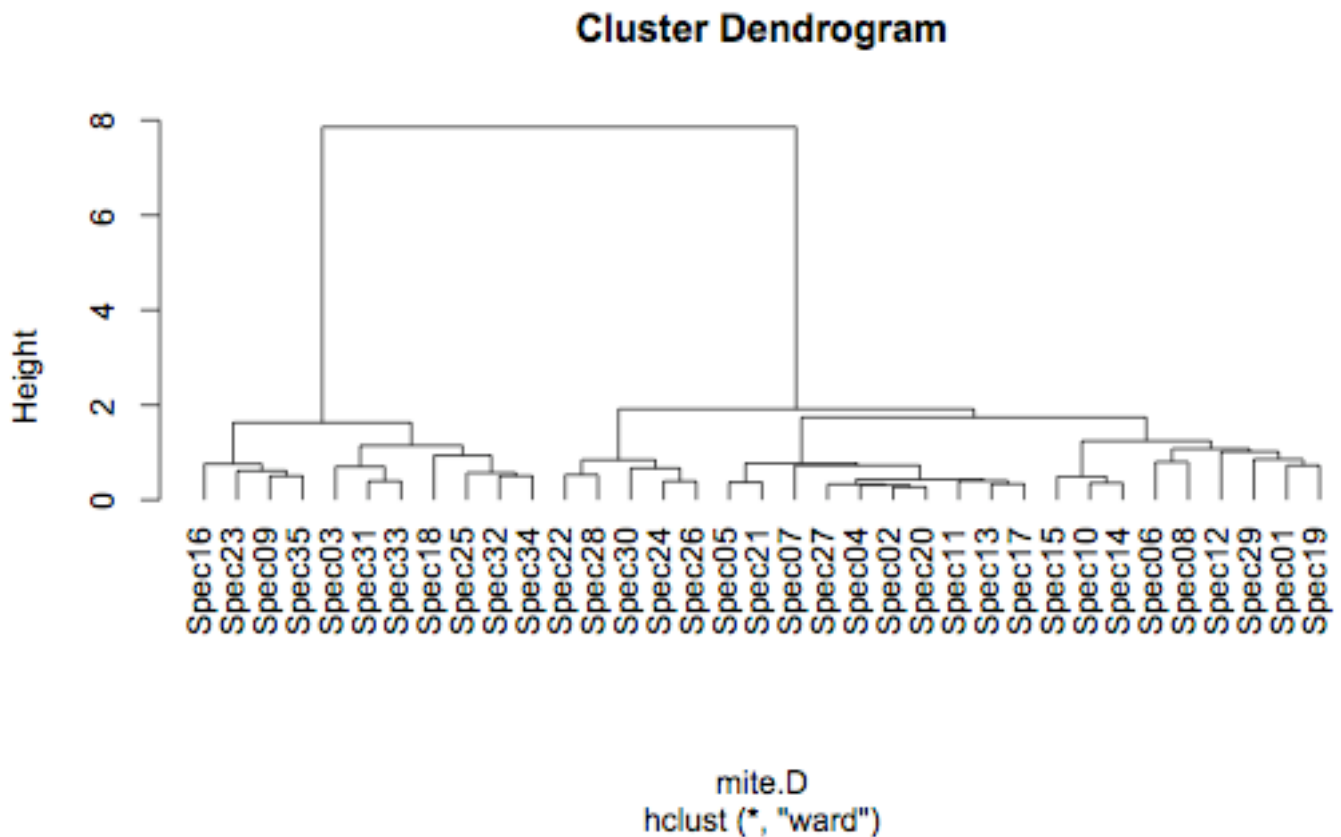
Full example in R: soil oribatid mites

```
### Extract mite data from vegan in R
### (70 soil cores x 35 species ; Borcard & Legendre 1994)
library(vegan)
data(mite)

### Hellinger transformation on the data
mite.hel <- decostand(mite, "hel")

### Spearman correlation matrix between species
mite.cor <- cor(mite.hel, method="spearman")

### Ward agglomerative clustering (ascending classification)
mite.D <- as.dist(1 - mite.cor)
mite.clust <- hclust(mite.D, "ward")
plot(mite.clust, hang=-1)
```




```
# Cut the dendrogram; divide the species in 2 groups
mite.2gr <- cutree(mite.clust, k=2)
```

Spec01	Spec02	Spec03	Spec04	Spec05	Spec06	Spec07	Spec08	Spec09	Spec10
1	1	2	1	1	1	1	1	2	1
Spec11	Spec12	Spec13	Spec14	Spec15	Spec16	Spec17	Spec18	Spec19	Spec20
1	1	1	1	1	2	1	2	1	1
Spec21	Spec22	Spec23	Spec24	Spec25	Spec26	Spec27	Spec28	Spec29	Spec30
1	1	2	1	2	1	1	1	1	1
Spec31	Spec32	Spec33	Spec34	Spec35					
2	2	2	2	2					

```
### Global test of each group
```

```
global.res <- kendall.global(mite.hel, group= mite.2gr)
```

```
$Concordance_analysis
```

	Group.1	Group.2
W	3.097870e-01	2.911888e-01
F	1.032305e+01	4.108130e+00
Prob.F	1.177138e-85	4.676566e-22
Corrected prob.F	2.354275e-85	4.676566e-22
Chi2	5.130073e+02	2.210123e+02
Prob.perm	1.000000e-03	1.000000e-03
Corrected prob.perm	2.000000e-03	2.000000e-03

```
### A posteriori test of the contribution of each species
### to the concordance of its group
post.res <- kendall.post(mite.hel, group=mite.2gr,
                        mult="holm", nperm=999)
```

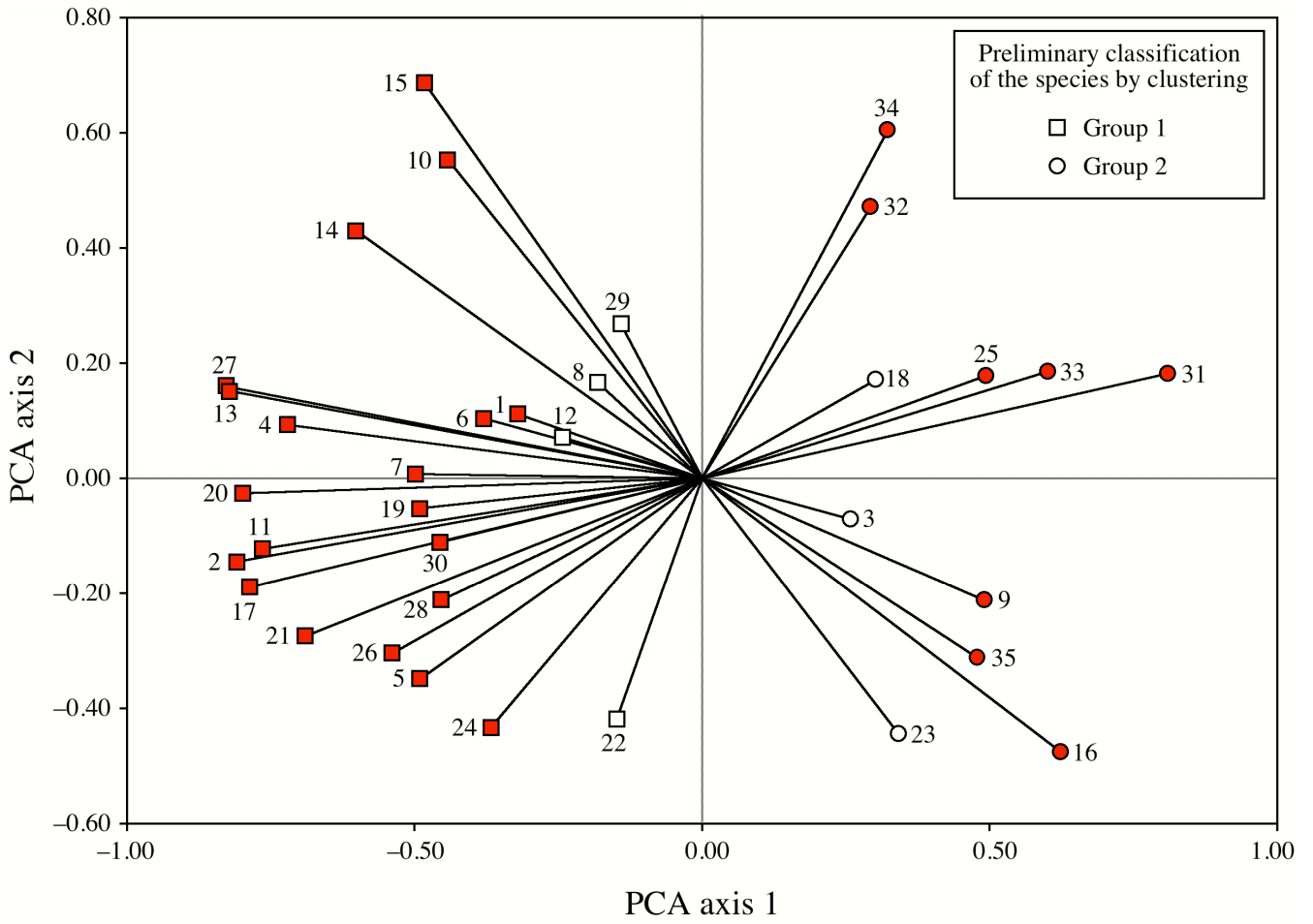
```
$A_posteriori_tests_Group[[1]]
```

	Spec01	Spec02	Spec04	Spec05	Spec06	Spec07
Spearman.mean	0.1851177	0.4258111	0.3590580	0.2505486	0.1802160	0.2833298
W.per.species	0.2190711	0.4497357	0.3857640	0.2817757	0.2143736	0.3131911
Prob	0.0010000	0.0010000	0.0010000	0.0010000	0.0040000	0.0010000
Corrected prob	0.0350000	0.0350000	0.0350000	0.0350000	0.0360000	0.0350000
	Spec08	Spec10	Spec11	Spec12	Spec13	Spec14
Spearman.mean	0.09248024	0.2444656	0.4138494	0.1263751	0.4177343	0.3301159
W.per.species	0.13029357	0.2759462	0.4382723	0.1627761	0.4419954	0.3580278
Prob	0.08900000	0.0010000	0.0010000	0.0330000	0.0010000	0.0010000
Corrected prob	0.17400000	0.0350000	0.0350000	0.1740000	0.0350000	0.0350000
	Spec15	Spec17	Spec19	Spec20	Spec21	Spec22
Spearman.mean	0.2185421	0.4212160	0.2574779	0.4180699	0.3623428	0.1250230
W.per.species	0.2511028	0.4453320	0.2884163	0.4423170	0.3889118	0.1614804
Prob	0.0010000	0.0010000	0.0010000	0.0010000	0.0010000	0.0300000
Corrected prob	0.0350000	0.0350000	0.0350000	0.0350000	0.0350000	0.1740000
	Spec24	Spec26	Spec27	Spec28	Spec29	Spec30
Spearman.mean	0.2188216	0.3016159	0.4217606	0.2577037	0.1108022	0.2301430
W.per.species	0.2513707	0.3307153	0.4458539	0.2886327	0.1478521	0.2622203
Prob	0.0010000	0.0010000	0.0010000	0.0010000	0.0560000	0.0010000
Corrected prob	0.0350000	0.0350000	0.0350000	0.0350000	0.1740000	0.0350000

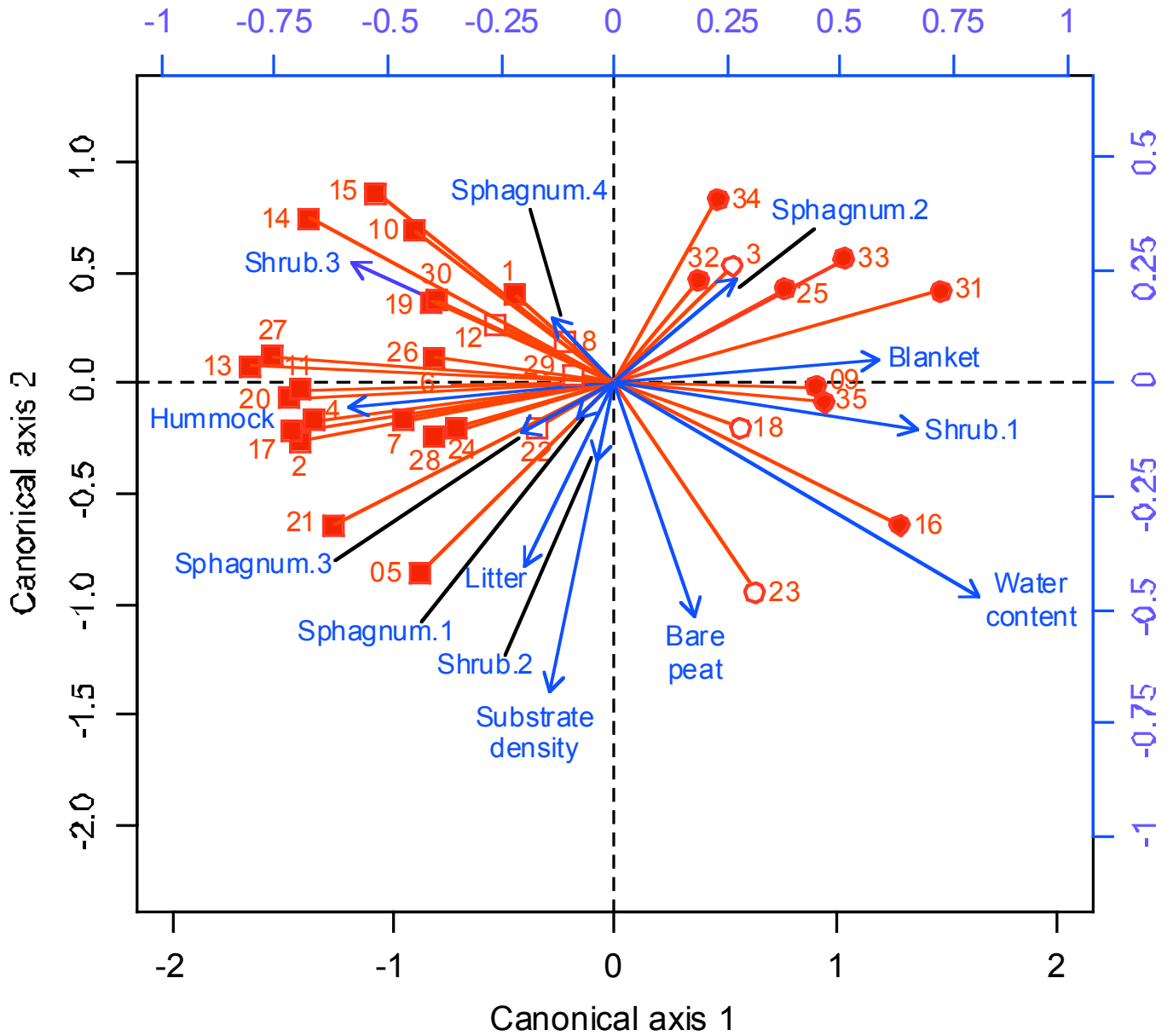
```
$A_posteriori_tests_Group[[2]]
```

	Spec03	Spec09	Spec16	Spec18	Spec23	Spec25
Spearman.mean	0.1222579	0.2712078	0.1906408	0.1375601	0.1342409	0.3342345
W.per.species	0.2020527	0.3374616	0.2642189	0.2159637	0.2129463	0.3947586
Prob	0.0460000	0.0010000	0.0010000	0.0220000	0.0290000	0.0010000
Corrected prob	0.1740000	0.0350000	0.0350000	0.1540000	0.1740000	0.0350000
	Spec31	Spec32	Spec33	Spec34	Spec35	
Spearman.mean	0.3446561	0.1833099	0.3188922	0.1764232	0.2498877	
W.per.species	0.4042328	0.2575544	0.3808111	0.2512938	0.3180797	
Prob	0.0010000	0.0030000	0.0010000	0.0030000	0.0010000	
Corrected prob	0.0350000	0.0350000	0.0350000	0.0390000	0.0350000	

PCA of standardized species



Relate the species associations to environmental variables:
RDA of standardized species with environmental variables

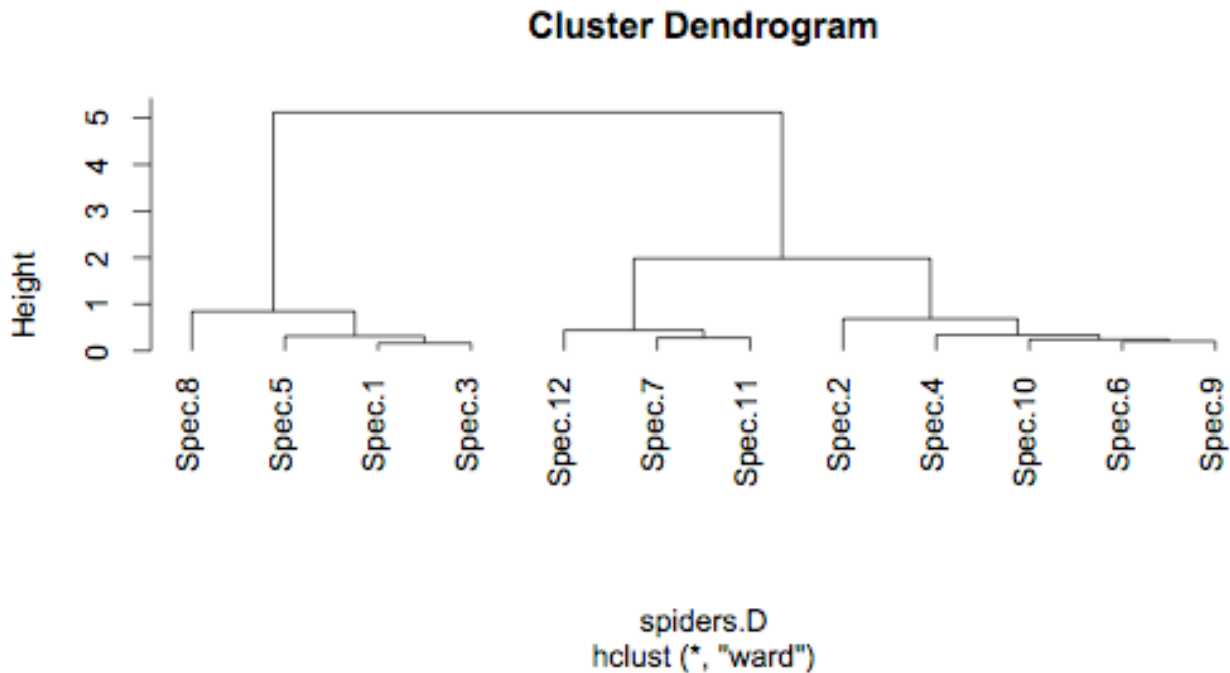


Complete R example: Netherland spiders

```
### Aart & Smeenk-Enserink (1975) spider data from the
### Netherland (28 sites x 12 species)
### The data can be found in the mvpart R library
spiders<-read.table("Spiders_28x12_spe.txt",h=TRUE)
spiders.hel <- decostand(spiders, "hel")

### Spearman correlation matrix between species
spiders.cor = cor(spiders.hel, method="spearman")

### Ward agglomerative clustering (ascending classification)
spiders.D = as.dist(1 - spiders.cor)
spiders.clust = hclust(spiders.D, "ward")
plot(spiders.clust, hang=-1)
```



Is there 2 or 3 groups of spiders?

Test two groups

```
### Form the groups
```

```
group.2 <- c(1,2,1,2,1,2,2,1,2,2,2,2)
global.res.2 <- kendall.global(spiders.hel,group=group.2,
                               mult="sidak")
```

```
$Concordance_analysis
```

	Group.1	Group.2
W	6.634689e-01	4.737846e-01
F	5.914481e+00	6.302537e+00
Prob.F	3.371949e-10	4.126332e-15
Corrected prob.F	3.371949e-10	8.252664e-15
Chi2	7.165464e+01	1.023375e+02
Prob.perm	1.000000e-03	1.000000e-03
Corrected prob.perm	2.000000e-03	2.000000e-03

```
out.post.2 <- kendall.post(spiders.hel, group=group.2,
                            mult="sidak")
```

```
$A_posteriori_tests_Group[[1]]
```

	Spec.1	Spec.3	Spec.5	Spec.8
Spearman.mean	0.71560252	0.67146276	0.49143813	0.3567842
W.per.species	0.78670189	0.75359707	0.61857860	0.5175882
Prob	0.00100000	0.00100000	0.00100000	0.0270000
Corrected prob	0.01193422	0.01193422	0.01193422	0.2799642

```
$A_posteriori_tests_Group[[2]]
```

	Spec.2	Spec.4	Spec.6	Spec.7	Spec.9	Spec.10
Spearman.mean	0.40329685	0.42007080	0.46382969	0.1904927	0.52020655	0.44015787
W.per.species	0.47788474	0.49256195	0.53085098	0.2916811	0.58018073	0.51013814
Prob	0.00200000	0.00100000	0.00100000	0.0930000	0.00100000	0.00100000
Corrected prob	0.02373775	0.01193422	0.01193422	0.6900530	0.01193422	0.01193422

	Spec.11	Spec.12
Spearman.mean	0.3407120	0.45620365
W.per.species	0.4231230	0.52417819
Prob	0.0090000	0.00100000
Corrected prob	0.1028112	0.01193422

Test three groups

```
group.3 <- c(1,2,1,2,1,2,3,1,2,2,3,3)
global.res.3 <-kendall.global(spiders.hel,group.3,
                             mult="sidak")
```

\$Concordance_analysis

	Group.1	Group.2	Group.3
W	6.634689e-01	7.000662e-01	7.606553e-01
F	5.914481e+00	9.336275e+00	6.356150e+00
Prob.F	3.371949e-10	1.427026e-17	6.829889e-09
Corrected prob.F	1.011585e-09	0.000000e+00	2.048967e-08
Chi2	7.165464e+01	9.450893e+01	6.161308e+01
Prob.perm	1.000000e-03	1.000000e-03	1.000000e-03
Corrected prob.perm	2.997001e-03	2.997001e-03	2.997001e-03

```
post.res <- kendall.post(spiders.hel, group=group.3,
                        mult="sidak")
```

\$A_posteriori_tests_Group[[1]]

	Spec.1	Spec.3	Spec.5	Spec.8
Spearman.mean	0.71560252	0.67146276	0.49143813	0.3567842
W.per.species	0.78670189	0.75359707	0.61857860	0.5175882
Prob	0.00100000	0.00100000	0.00200000	0.0270000
Corrected prob	0.01193422	0.01193422	0.02373775	0.2799642

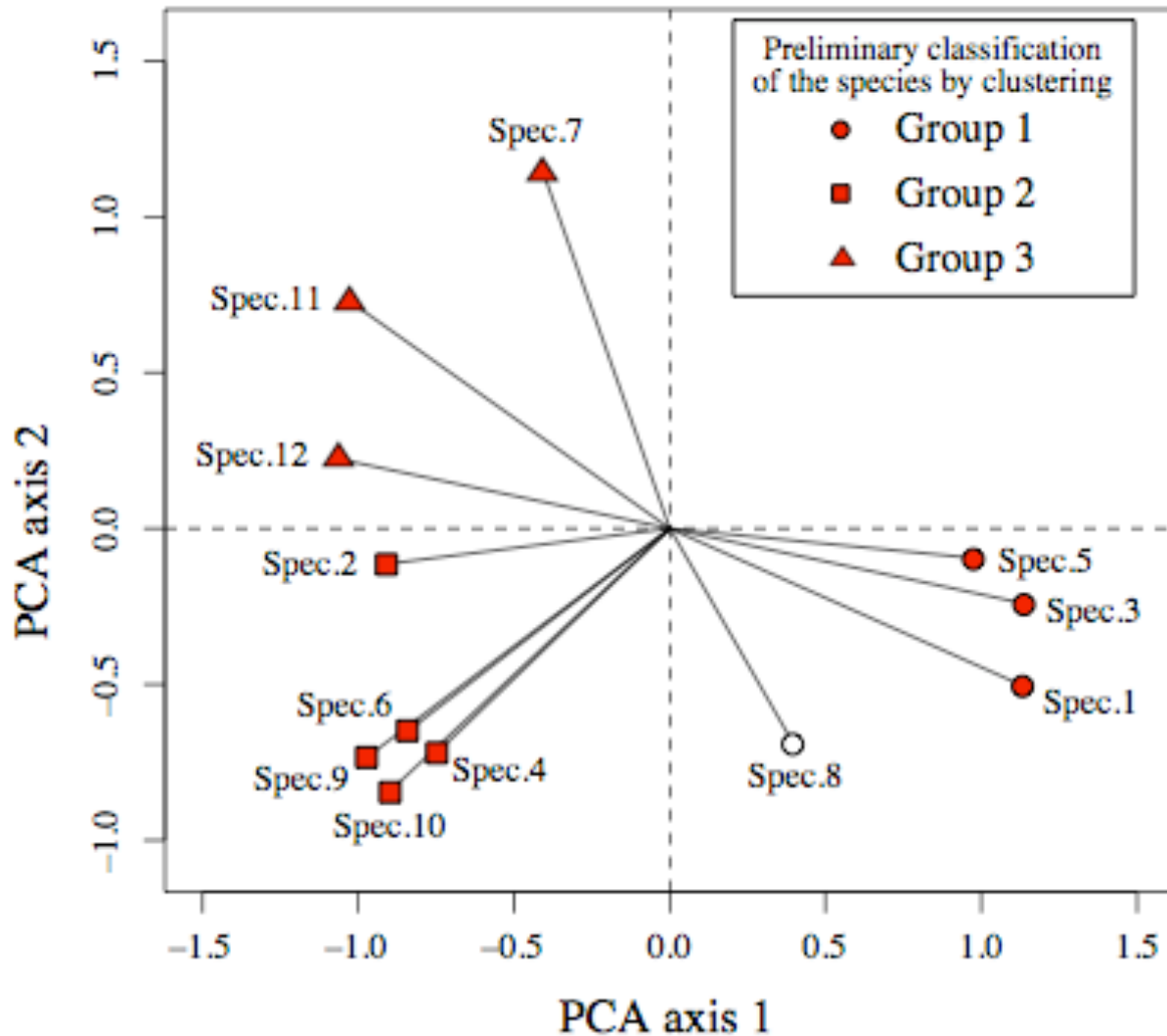
\$A_posteriori_tests_Group[[2]]

	Spec.2	Spec.4	Spec.6	Spec.9	Spec.10
Spearman.mean	0.47356894	0.62159400	0.66388990	0.72146269	0.68455036
W.per.species	0.57885515	0.69727520	0.73111192	0.77717015	0.74764028
Prob	0.00400000	0.00100000	0.00100000	0.00100000	0.00100000
Corrected prob	0.04695795	0.01193422	0.01193422	0.01193422	0.01193422

\$A_posteriori_tests_Group[[3]]

	Spec.7	Spec.11	Spec.12
Spearman.mean	0.61601977	0.70451843	0.59905641
W.per.species	0.74401318	0.80301229	0.73270427
Prob	0.00100000	0.00100000	0.00100000
Corrected prob	0.01193422	0.01193422	0.01193422

PCA of standardized species



References

Blanchet, F. G. and P. Legendre. 2008. Kendall.W. R package version 2.8. <http://www.bio.umontreal.ca/legendre/>. Also available in R library 'vegan'.

Legendre, P. 2005. Species associations: the Kendall coefficient of concordance revisited. *Journal of Agricultural, Biological, and Environmental Statistics* 10: 226-245.

Legendre, P. 2010. Coefficient of concordance. In: *Encyclopedia of Research Design*. SAGE Publications (in press).