

A (brief) introduction to ordination and the vegan package

Eduard Szöcs

Institute for Environmental Sciences - University of Koblenz-Landau



SEFS9, July 5th 2015

Datasets
oooooooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oo

Permutation Tests
o

End

Datasets
ooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oo

Permutation Tests
o

End

Topics addressed

	Raw data	Transformed data	Unimodal	Distance-based	Model-based
Unconstrained	PCA	tb-PCA	CA, DCA	PCoA, NMDS	MM, LVM
Constrained	RDA	tb-RDA	CCA	db-RDA	CAO, CQO
Other				Permanova, Dispersion	manyglm

No maths today ;)

Datasets
oooooooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oo

Permutation Tests
o

End

Datasets

Demonstration: Doubs river fish communities

5 / 38



- ▶ Fish
 - ▶ 30 sites along the Doubs River
- Questions**
- ▶ How does fish composition change downstream?
 - ▶ Environmental drivers?

Verneau, J. (1973) Cours d'eau de Franche-Comté (Massif du Jura). Recherches écologiques sur le réseau hydrographique du Doubs. Essai de biotypologie. Thèse d'état, Besançon. 1-257.

Datasets Unconstrained Ordination Constrained Ordination Permutation Tests End

Demonstration: Doubs river fish communities — Species

6 / 38

```
Dabu <- read.table('doubsAbu.csv', sep = ',', header = TRUE)
Denv <- read.table('doubsEnv.csv', sep = ',', header = TRUE)
Depa <- read.table('doubsSpa.csv', sep = ',', header = TRUE)
```

```
dim(Dabu)
[1] 30 27
```

30 sites, 27 taxa

```
head(Dabu[, 1:18])
  CHA TRU VAI LOC OMB BLA HOT TOX VAN CHE BAR SPI GOU BRO PER BOU PSO ROT
  1 0 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  2 0 5 4 3 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  3 0 5 5 5 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0
  4 0 4 5 5 0 0 0 0 0 1 0 0 1 2 2 0 0 0 0
  5 0 2 3 2 0 0 0 0 5 2 0 0 2 4 4 0 0 2
  6 0 3 4 5 0 0 0 0 1 2 0 0 1 1 1 0 0 0 0
```

Datasets Unconstrained Ordination Constrained Ordination Permutation Tests End

```
# Dimension and first rows of Environmental data
dim(Denv)
```

```
[1] 30 11
```

30 sites, 11 variables

```
head(Denv)
```

```
das alt pen deb pH dur pho nitr ammon oxy dbo
1 0.3 934 48.0 0.84 7.9 45 0.01 0.20 0.00 12.2 2.7
2 2.2 932 3.0 1.00 8.0 40 0.02 0.20 0.10 10.3 1.9
3 10.2 914 3.7 1.80 8.3 52 0.05 0.22 0.05 10.5 3.5
4 18.6 854 3.2 2.53 8.0 72 0.10 0.21 0.00 11.0 1.3
5 21.5 849 2.3 2.64 8.1 84 0.38 0.52 0.20 8.0 6.2
6 32.4 846 3.2 2.86 7.9 60 0.20 0.19 0.00 10.2 5.3
```

Datasets
000000

Unconstrained Ordination
oooooooooooooooooooooooo

Constrained Ordination
oo

Permutation Tests
0

End

Exercise: Salinization and Pesticides

8 / 38

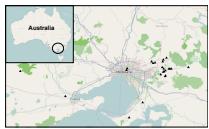
► Macroinvertebrates

► 24 sites

► covering a salinity and toxicity gradient

Questions:

- Interaction between salinization and pesticides?
- Which species are affected?
- Other influences?



The dataset is published in: Szöcs, E., Kefford, B.J., Schäfer, R.B., 2012. Is there an interaction of the effects of salinity and pesticides on the community structure of macroinvertebrates? *Science of the Total Environment* 437, 121–126.

Datasets
000000

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oo

Permutation Tests
0

End

Exercise: Salinization and Pesticides

9 / 38

```
# setwd('3-Ordination/data/')
abu <- read.table('melbourneAbu.csv', sep = ';', header = TRUE)
env <- read.table('melbourneEnv.csv', sep = ';', header = TRUE)
```

```
# dimensions of data.frame
dim(env)
```

```
[1] 24 23
```

```
dim(abu)
```

```
[1] 24 76
```

24 sites, 22 environmental variables, 75 taxa

```
head(env[, 1:10])
```

```
ID T pH oxygen Depth maxwidth minwidth rifperc poolperc Bedrock
1 1-11 16.8 7.67 80.0 0.9 15 12.0 0 100 0
2 2-11 16.4 7.29 85.0 0.9 30 15.0 0 100 0
3 3-11 17.3 7.20 77.9 0.4 4 2.5 0 100 0
4 4-11 15.6 7.84 72.0 0.7 8 2.5 0 100 0
5 5-11 17.2 6.87 69.9 0.9 7 4.0 0 100 0
6 6-11 15.5 7.26 80.0 0.2 3 2.0 5 95 0
```

Datasets
000000

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oo

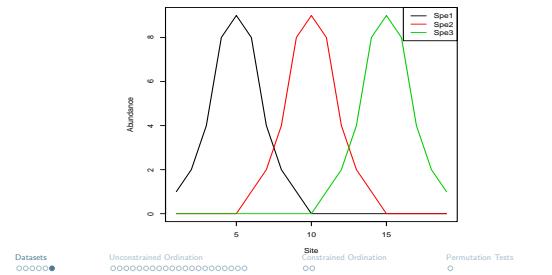
Permutation Tests
0

End

Exercise: Dummy abundances

10 / 38

```
# Load dummy data
dummy <- read.table('dummydata.csv', header = TRUE, sep = ';')
# plot dummy data
matplot(dummy[, -1], type = 'l', xlab = 'Site', ylab = 'Abundance',
        lty = "solid", lwd = 2)
legend('topright', legend = colnames(dummy)[-1],
       col = 1:3, lty = 'solid', lwd = 2)
```

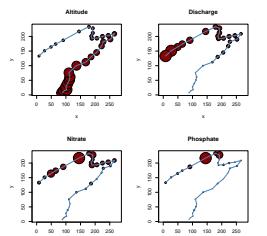


Unconstrained Ordination

- Principal Components Analysis (PCA)
- Principal coordinates analysis (PCoA)
- Nonmetric Multidimensional Scaling (NMDS)

Principal Components Analysis (PCA) — Why?

12 / 38



- 11 variables

- Questions:**

 - ▶ Which variables are correlated?
 - ▶ Which sites have similar conditions?
 - ▶ How do conditions change

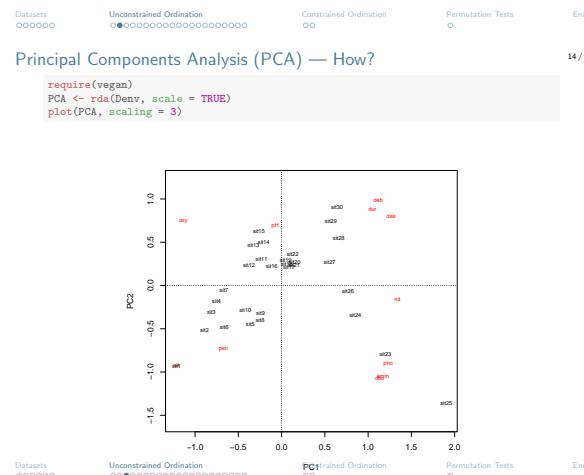
- ▶ pairwise comparisons
 - ▶ 3D possible
 - ▶ more than 3 dimensions?

- ▶ "Look from another angle on the data"
 - ▶ PCA is just a rotation of the coordinate system
 - ▶ The rotation is done so that the first axis contains as much variation as possible
 - ▶ Second axis than most of remaining variation

Short Demo.

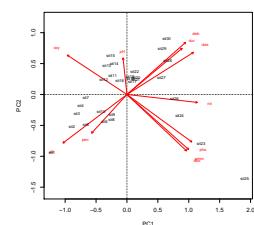
Maths:

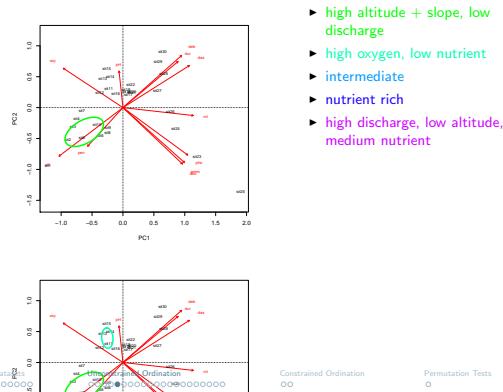
- ▶ The covariance (or correlation) matrix is decomposed into its Eigenvectors and Eigenvalues.
 - ▶ The Eigenvectors give the rotation needed
 - ▶ The Eigenvalues stretch the axes



```
biplot(PCA, cex = 5, scaling = 3)
```

- ▶ angle between variables **approx.** their correlation
 - ▶ distance between sites **approx.** their euclidean distance
 - ▶ projecting a site on a variable
approx. the relative value
 - ▶ scaling = 1 - to interpret (only) distances between sites
 - ▶ scaling = 2 - to interpret (only) correlations between variables





```

summary(PCA, display = NULL, scaling = 3)

Call:
rda(X = Denv, scale = TRUE)

Partitioning of correlations:
           Inertia Proportion
Total          11             1
Unconstrained 11             1

Eigenvalues, and their contribution to the correlations

Importance of components:

PC1   PC2   PC3   PC4   PC5   PC6   PC7
Eigenvalue 5.9687 2.1638 1.06516 0.73873 0.40027 0.33565 0.17275
Proportion Explained 0.5420 0.1962 0.09614 0.06711 0.03639 0.03051 0.01575
Cumulative Proportion 0.5420 0.7389 0.83616 0.90331 0.93970 0.97022 0.98599

PC8   PC9   PC10  PC11
Eigenvalue 0.10821 0.02368 0.01707 0.00593
Proportion Explained 0.00984 0.00215 0.00155 0.000540
Cumulative Proportion 0.99575 0.99790 0.99946 1.00000

Scaling 3 for species and site scores
* Both sites and species are scaled proportional to eigenvalues
on all dimensions
* General scaling constant of scores:
```

Exercise

19 / 38

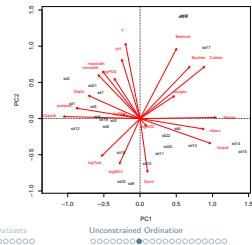
```

take <- env[, !names(env) %in% c('ID', 'logCond', 'logmaxTU')]
PCA <- rda(take, scale = TRUE)
cumsum(PCA$CA$eig / PCA$tot.chi)[1:2]

    PC1      PC2
0.2839873 0.4537452

```

```
biplot(PCA, scaling = 3)
```



- ▶ multiple variables interrelated
 - ▶ 1st axis can be interpreted as *hydrological gradient*
 - ▶ 2nd axis can be interpreted as *chemistry gradient*

Excusus — Principal component regression (PCR)

20 / 38

Question:

- ▶ How is diversity related to salinity, pesticides and other variables?

Problem:

- ▶ Only 24 sites
 - ▶ but 22 (partly correlated) explanatory variables
 - ▶ strong hypotheses about salinity and pesticides

A Solution:

- ▶ Reduce number of variables to *Principal Components*
 - ▶ regress these

```

cursus — principal component regression (PCR)

# calculate shannon diversity index
div <- diversity(abu[, -1], index = 'shannon')
pc <- scores(PCA, choices = c(1, 2), scaling = 1, display = 'sites')
model_data <- data.frame(div, pc, logCond = env$logCond, logmaxTU = env$logmaxTU)
model <- lm(div ~ PC1 + PC2 + logCond + logmaxTU)
data <- model_data

```

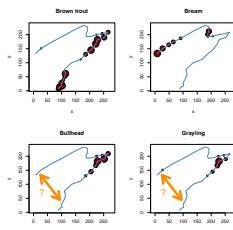
```
Call:  
lm(formula = dix ~ PC1 + PC2 + logCond + logmargTH, data = model.data)
```

Residuals:

Coefficients:						
	Estimate	Std. Error	t value	Pr(> t)		
(Intercept)	1.83079	0.43429	4.216	0.000468	***	
PC1	0.01971	0.16691	0.118	0.90726		
PC2	0.02192	0.15970	0.120	0.911996		
logCond	-0.20942	0.13050	-1.605	0.125049		
logmaxTU	-0.12572	0.07316	-1.718	0.101994		

Residual standard error: 0.3645 on 19 degrees of freedom
Multiple R-squared: 0.2682, Adjusted R-squared: 0.1141
F-statistic: 1.741 on 4 and 19 DF, p-value: 0.1827

Datasets Unconstrained Ordination



- Species may be absent due to different factors (too high flow, too saline, etc.)
- Absence contains less information than Presence
- PCA preserves the Euclidean distance between sites
- Need another measure of similarity for (raw) abundances

Datasets
000000

Unconstrained Ordination
0000000000●0000000000

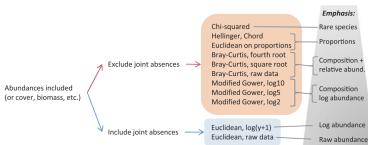
Constrained Ordination
00

Permutation Tests
0

End

Dissimilarity measures

23 / 38



	Spe1	Spe2	Spe3
sit1	0	4	8
sit2	0	1	1
sit3	1	0	0

```
vegdist(mat, method = 'euclidean')
      sit1     sit2
sit2 7.615773
sit3 9.000000
```

```
vegdist(mat, method = 'bray')
      sit1     sit2
sit2 0.7142857
sit3 1.0000000 1.0000000
```

Anderson, M.J., Crist, T.O., et al., 2011. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. Ecology Letters 14, 19–28.

Datasets
000000

Unconstrained Ordination
0000000000●0000000000

Constrained Ordination
00

Permutation Tests
0

End

Principal coordinates analysis (PCoA)

24 / 38

- Works on distance matrices
- Species can be added as *weighted averages*
- Eigenvalue based
- PCoA with Euclidean distance == PCA

Datasets
000000

Unconstrained Ordination
0000000000●0000000000

Constrained Ordination
00

Permutation Tests
0

End

Principal coordinates analysis (PCoA)

25 / 38

```

# Distance matrix
Dabu_dist <- vegdist(Dabu, method = 'bray')

# PCoA
PCoA <- cmdscale(Dabu_dist, eig = TRUE)

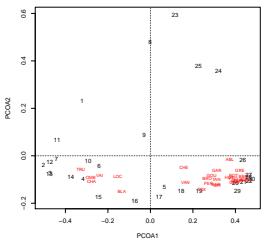
# Create plot
plot(PCoA$points, type = 'n',
      xlab = 'PCoA1', ylab = 'PCoA2')
text(PCoA$points,
      labels = colnames(Dabu), cex = 0.9)
abline(v = 0, lty = 'dotted')
abline(h = 0, lty = 'dotted')

# Add species as weighted averages
va <- wascores(PCoA$points, Dabu)
text(va, labels = colnames(Dabu),
      col = 'red', cex = 0.7)

# explained variance
(PCoA$eig / sum(PCoA$eig))[1:2] * 100

[1] 49.24914 15.95758

```



Datasets Unconstrained Ordination

26 / 39

- ▶ Similar to PCoA
 - ▶ Does not preserve exact distances between objects
 - ▶ Possibly better representation in low dimensions
 - ▶ **Not eigenvalue based, iterative algorithm**
 - ▶ Axes have no meaning, just the relative distances

The figure is a Non-Metric Multidimensional Scaling (NMDS) ordination plot. The x-axis is labeled "Environmental Gradient 1" and the y-axis is labeled "Environmental Gradient 2". The plot area contains numerous small dots representing different species or samples. A horizontal dashed line at approximately y=0.5 represents the mean species distribution. A vertical dashed line at approximately x=0.5 represents the mean environmental gradient. A diagonal dashed line from the bottom-left to the top-right represents the 1:1 relationship between the two axes.

27 / 29

```
# Distance matrix
Dabu_0 <- Dabu[!rowSums(Dabu) == 0, ]
Dabu_dist <- vegdist(Dabu_0, method = 'bray')

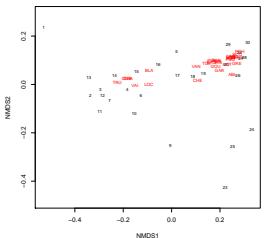
# NMDS
NMDS <- metaMDS(Dabu_dist, k = 2, trace = 0)

# Plot
plot(NMDS, type = 't')

# Add species as weighted averages
va <- wascores(NMDS$points, Dabu_0)
text(va, labels = colnames(Dabu),
      col = 'red', cex = 0.7)

# Stress value
NMDS$stress

[1] 0.07376279
```



Datasets Unconstrained Ordination
○○○○○○○○○○○○○○○●○○○○

Your turn!

Using the artificial dummy dataset.

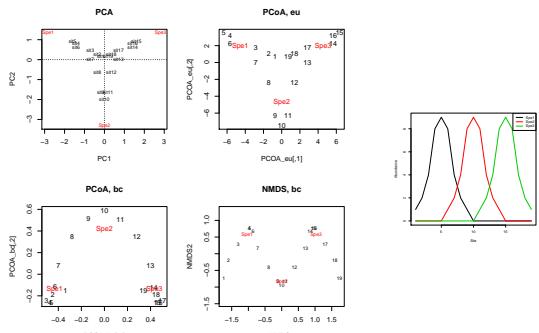
Run:

1. PCA
 2. PCoA with euclidean distance
 3. PCoA with Bray-Curtis dissimilarity
 4. NMDS with Bray-Curtis dissimilarity

What are the differences between ordinations?
Which represent better the underlying gradient?

Exercise

29 / 38



Datasets

Unconstrained Ordination

Constrained Ordination

Permutation Tests

End

Indirect Gradient Analysis

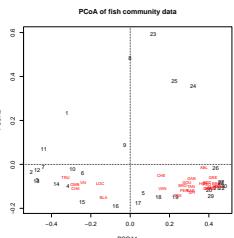
- This ordination is **only** driven by fish community data

Question:

- ▶ How can we interpret the gradients in community composition?

In conclusion

- ▶ Superimpose environmental variables



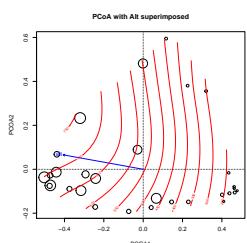
Unconstrained Ordination

Constrained Ordination

Permutation Tests

End

1



Datasets
000000

Unconstrained Ordination

Constrained Ordination
oo

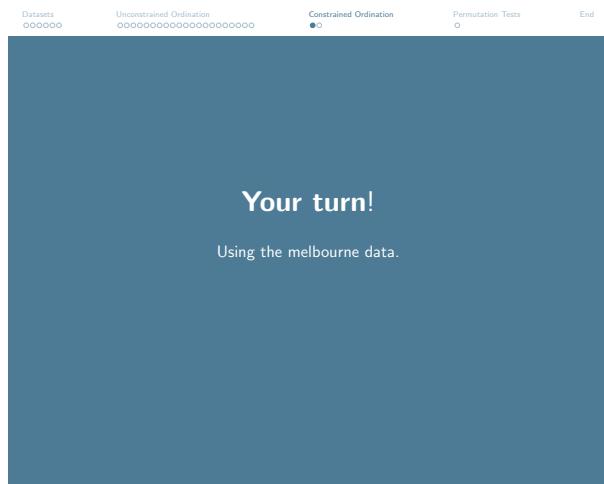
Permutation Tests

Your turn!

Using the melbourne data.

Constrained Ordination

- ▶ Redundancy analysis (RDA)
- ▶ Transformation-based RDA
- ▶ Distance-based RDA



Permutation Tests

Your turn!

Using the melbourne data.

Usefull topics not covered here

38 / 38

- ▶ Distance-based hypothesis testing ((PER-)MANOVA, SIMPER, ANOSIM)
- ▶ Dispersion measures (β -Diversity, Functional diversity)
- ▶ Model-based ordination / hypothesis testing (See work of David Warton et al.)

Datasets
oooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oo

Permutation Tests
o

End