

A (short) introduction to ordination with the vegan package

Eduard Szöcs

Institute for Environmental Sciences - University of Koblenz-Landau



SEFS9, July 5th 2015

Datasets
oooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oooooooooooo

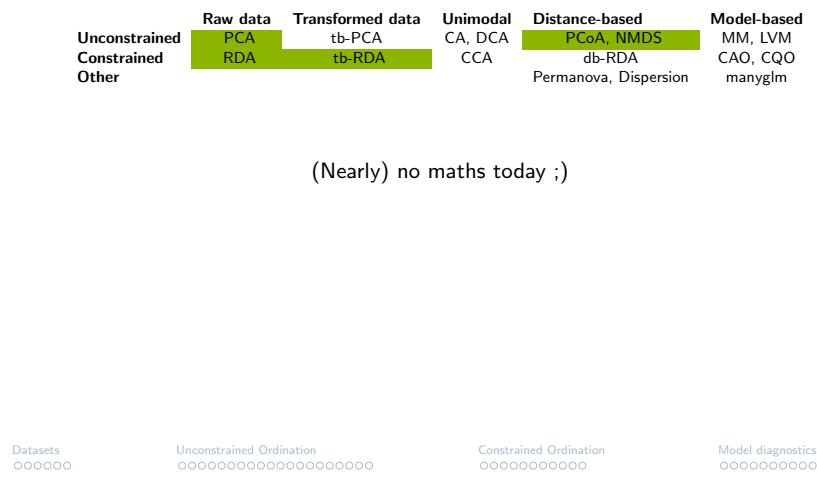
Model diagnostics / testing
oooooooooooo

Datasets
oooooooo

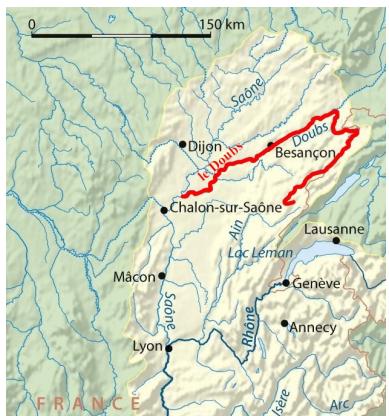
Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oooooooooooo

Model diagnostics / testing
oooooooooooo



Datasets



- ▶ Fish
 - ▶ 30 sites along the Doubs River

Questions

- ▶ How does fish composition change downstream?
 - ▶ Environmental drivers?

Verneaux, 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

Model diagnostics / testing

Demonstration: Doubs river fish communities — Species

```
Dabu <- read.table('doubtsAbu.csv', sep = ',', header = TRUE)
Denv <- read.table('doubtsEnv.csv', sep = ',', header = TRUE)
Dspa <- read.table('doubtsSpa.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	
2	0	5	4	3	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	1	0	0	0	
4	0	4	5	5	0	0	0	0	0	1	0	0	1	2	2	0	0	
5	0	2	3	2	0	0	0	0	5	2	0	0	2	4	4	0	0	
6	0	3	4	5	0	0	0	0	1	2	0	0	1	1	1	0	0	

Datasets

Unconstrained Ordination

Constrained Ordination

Model diagnostics / testing

Demonstration: Doubs river fish communities — Environment

7 / 54

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

[1] 30 11

30 sites, 11 variables

`head(Denv)`

	das	alt	pen	deb	pH	dur	pho	nit	amm	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	18.5	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	0.00	0.20	0.00	10.2	5.3

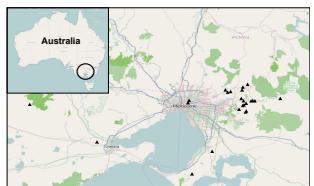
Datasets

Constrained Ordination

Model diagnostics / testing

Exercise: Salinization and Pesticides

8 / 54



- ▶ Macrovertebrates
 - ▶ 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

Unconstrained Ordination

Constrained Ordination

Model diagnostics / testing

Exercise: Salinization and Pesticides

9 / 54

```
# 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

	ID	T	pH	oxygen	Depth	maxwidth	minwidth	rifperc	poolperc	Bedrock
1	1-11	16.8	7.67	80.1	0.9	15	12.0	0	100	0
2	2-11	16.5	7.29	83.0	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	6.97	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

Unconstrained Ordination

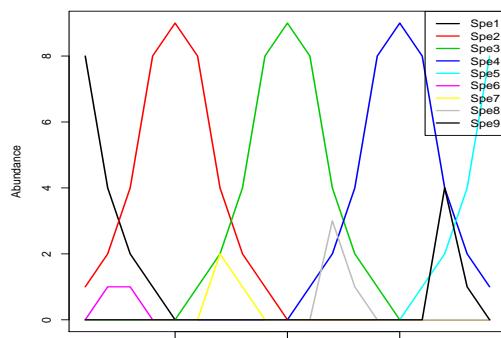
Constrained Ordination

Model diagnostics / testing

Exercise: Dummy abundances

10 / 54

```
# 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, col = 1:9)
legend('topright', legend = colnames(dummy)[-1],
       col = 1:9, lty = 'solid', lwd = 2)
```



Datasets

Unconstrained Ordination

Constrained Ordination

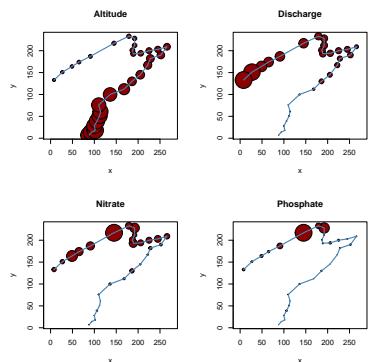
Model diagnostics / testing

Unconstrained Ordination

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

Principal Components Analysis (PCA) — Why?

12 / 54



- ▶ 11 variables

Questions:

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

Solutions?

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

Principal Components Analysis (PCA) — What?

13 / 54

- ▶ *"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

Datasets

Unconstrained Ordination

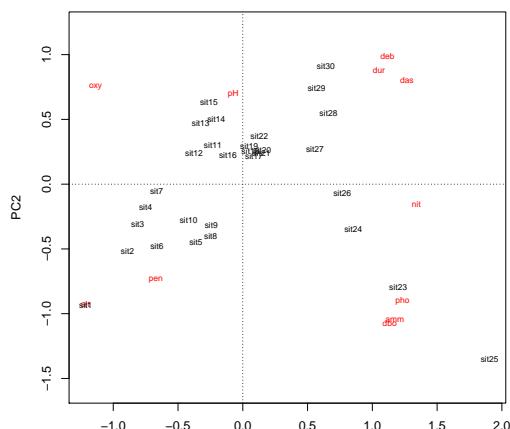
Constrained Ordination

Model diagnostics / testing

Principal Components Analysis (PCA) — How?

14 / 54

```
require(vegan)  
PCA <- rda(Denv, scale = TRUE)  
plot(PCA, scaling = 3)
```



Datasets

Unconstrained Ordination

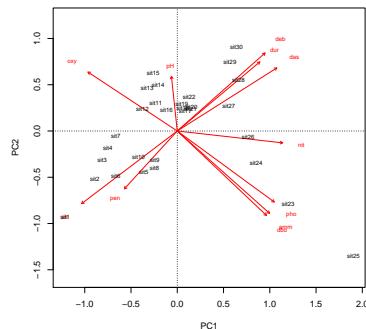
Constrained Ordination

Model diagnostics / testing

Principal Components Analysis (PCA) — Interpretation? (I)

15 / 54

```
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

Datasets

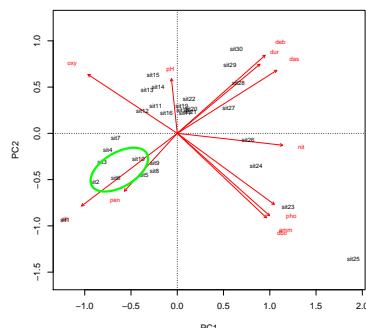
Unconstrained Ordination

Constrained Ordination

Model diagnostics / testing

Principal Components Analysis (PCA) — Interpretation? (II)

16 / 54



- ▶ high altitude + slope, low discharge
 - ▶ high oxygen, low nutrient
 - ▶ intermediate
 - ▶ nutrient rich
 - ▶ high discharge, low altitude, medium nutrient

1

Principal Components Analysis (PCA) — Interpretation? (III)

17 / 54

```

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 0.1727
Proportion Explained 0.5426 0.1967 0.09683 0.06716 0.03639 0.03051 0.0157
Cumulative Proportion 0.5426 0.7393 0.83616 0.90331 0.93970 0.97022 0.9859
PC8    PC9    PC10   PC11
Eigenvalue 0.10821 0.02368 0.01707 0.005993
Proportion Explained 0.00984 0.00215 0.00155 0.000540
Cumulative Proportion 0.99575 0.99790 0.99946 1.000000

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

```

Datasets

Unconstrained Ordination

Constrained Ordination

Model diagnostics / testing

Your turn!

Load the Melbourne dataset (only environmental variables).

Exclude the variables ID, logCond and logmaxTU.
Perform a PCA.

- Which variables are correlated?
- How much variance is explained by the first 2 axes?
- How could the two PCA axes be interpreted?

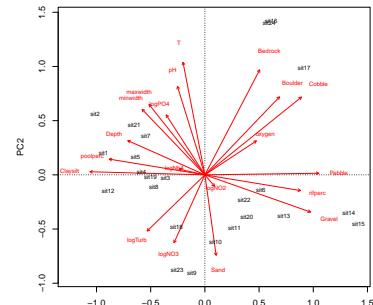
Exercise

19 / 54

```
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*

Datasets

○○○○○○○

Unconstrained Ordination

○○○○○○○●○○○○○○○○○○○

Constrained Ordination

○○○○○○○○○○○

Model diagnostics / testing

○○○○○○○○○○○

Excursus — Principal component regression (PCR)

20 / 54

Question:

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

Problem:

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

A Solution:

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

Datasets

○○○○○○○

Unconstrained Ordination

○○○○○○○●○○○○○○○○○○○

Constrained Ordination

○○○○○○○○○○○

Model diagnostics / testing

○○○○○○○○○○○

Excursus — principal component regression (PCR)

21 / 54

```
# 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)
summary(model)

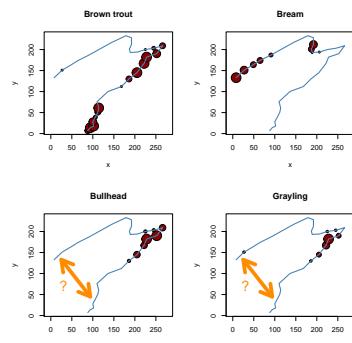
Call:
lm(formula = div ~ PC1 + PC2 + logCond + logmaxTU, data = model_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.64415 -0.15688  0.02063  0.18219  0.57929 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.83079   0.43429  4.216 0.000468 ***
PC1         0.01971   0.16691  0.118 0.907262  
PC2         0.02192   0.19570  0.112 0.911996  
logCond     -0.20942   0.13050 -1.605 0.125049  
logmaxTU   -0.12572   0.07316 -1.718 0.101994  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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
```

22 / 54



- ▶ 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

Method	Number of Datasets
Constrained Ordination	10
Unconstrained Ordination	6

Model diagnostics / testing

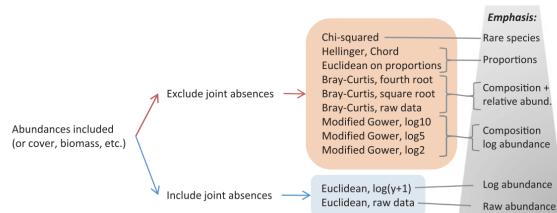
Dissimilarity measures — Species abundance paradox

23 / 54

	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 1.732051
```

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



from: 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.

Principal coordinates analysis (PCoA)

24 / 54

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

Datasets



Unconstrained Ordination

Constrained Ordination

Model diagnostics / testing

Principal coordinates analysis (PCoA)

25 / 54

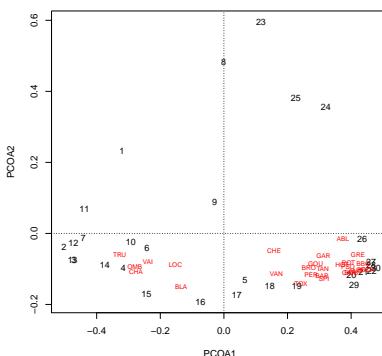
```
# Distance matrix
Dabu_dist <- vegdist(Dabu, method = 'bray')

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

# Create plot
plot(PCOA$points, type = 'n',
      xlab = 'PCOA1', ylab = 'PCOA2')
text(PCOA$points,
      labels = rownames(Dabu), cex = 0.9)
abline(h = 0 , lty = 'dotted')
abline(v = 0 , lty = 'dotted')
# Add species as weighted averages
wa <- wascores(PCOA$points, Dabu)
text(wa, 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

○○○○○○○○○○●○○○○○

Constrained Ordination

○○○○○○○○○○

Model diagnostics / testing

○○○○○○○○○○

Nonmetric Multidimensional Scaling (NMDS)

26 / 54

- ▶ 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

Datasets

○○○○○○

Unconstrained Ordination

○○○○○○○○○○●○○○○○

Constrained Ordination

○○○○○○○○○○

Model diagnostics / testing

○○○○○○○○○○

Nonmetric Multidimensional Scaling (NMDS)

27 / 54

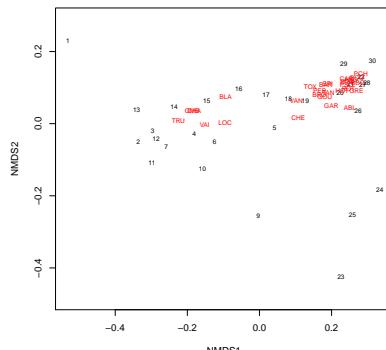
```
# Distance matrix
Dabu_0 <- Dabu[!rowSums(Dabu) == 0, ]
Dabu_dist <- vegdist(Dabu_0, method = 'bray')

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

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

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

# Stress value
NMDS$stress
[1] 0.07429467
```



Datasets
○○○○○○

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

Constrained Ordination
○○○○○○○○○○

Model diagnostics / testing
○○○○○○○○○○

Your turn!

Using the artificial dummy dataset.

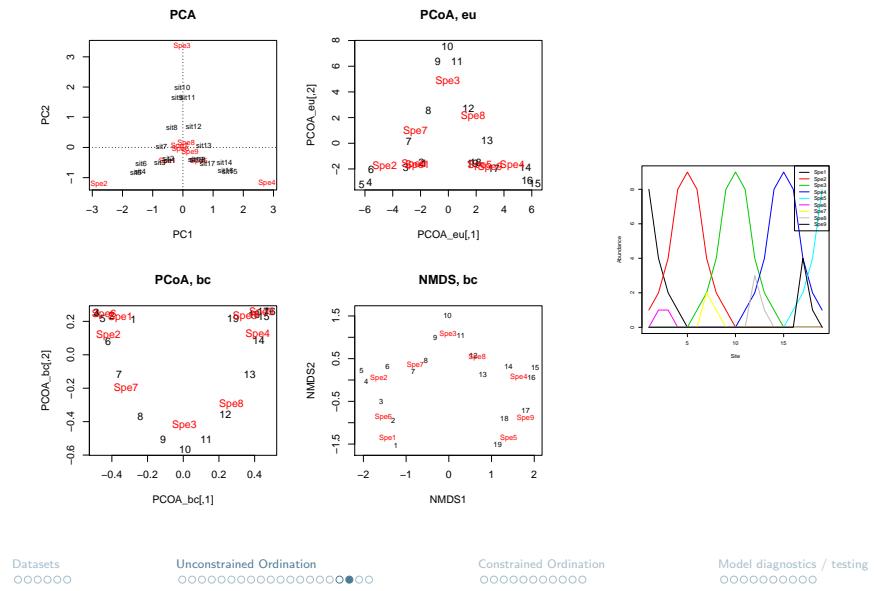
Run:

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 / 54



Fit environmental variables to ordination (I)

30 / 54

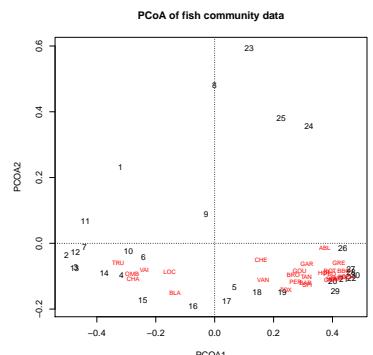
- ▶ This ordination is **only** driven by fish community data

Question:

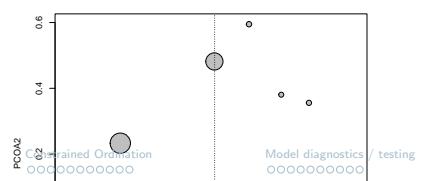
- ▶ How can we interpret the gradients in community composition?

A solution:

- Superimpose environmental variables



PCoA with Altitude superimposed



Fit environmental variables to ordination (II)

31 / 54

```

# PCoA of fish community data
plot(PCOA$points,
      xlab = 'PCOA1', ylab = 'PCOA2',
      cex = 5*Denv$alt / max(Denv$alt),
      main = 'PCoA with Altitude',
      bg = 'grey75', pch = 21)
abline(h = 0 , lty = 'dotted')
abline(v = 0 , lty = 'dotted')

# Fit Altitude to site-scores
ef <- envfit(PCOA, Denv)
plot(ef)
ef    # summary

# Fit GAM
ordisurf(PCOA, Denv$alt, add = TRUE)

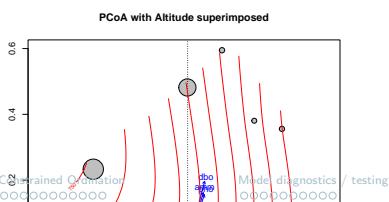
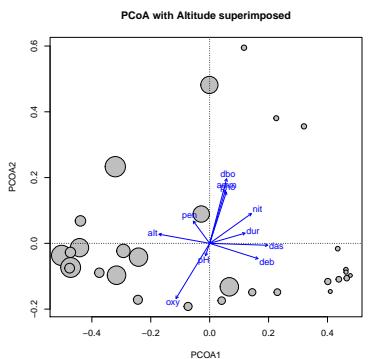
```

- ▶ *Post hoc* method
 - ▶ non-linearity?
 - ▶ be careful with `summary`
 - ▶ Constrained ordination a better alternative

Datasets



Unconstrained Ordination

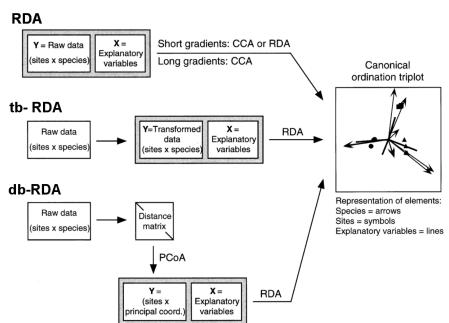


Constrained Ordination

Constrained Ordination

33 / 54

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



Adapted from: Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280.



Redundancy analysis (RDA)

34 / 54

- ▶ Associates both environmental and community data at once
 - ▶ Combination of regression and PCA:
 1. Regress explanatory variables on community data
 2. Run PCA on fitted values
 - ▶ Can test hypotheses about relationships

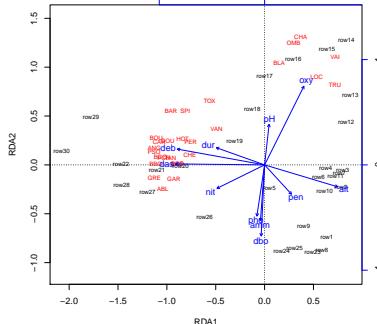


Redundancy analysis (RDA) — How?

35 / 54

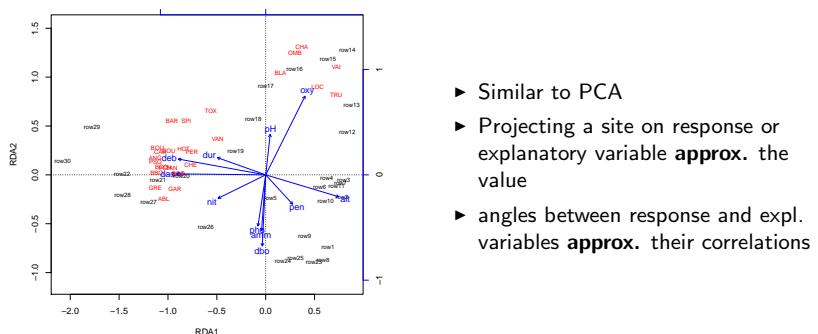
```
RDA <- rda(Dabu ~ ., data = Denv,  
            scale = TRUE)  
plot(RDA, scaling = 3)
```

- ▶ Formula interface
 - ▶ Left side: Response **matrix**
 - ▶ Right side: Response variables from Denv



Redundancy analysis (RDA) — Interpretation? (I)

36 / 54



- ▶ Similar to PCA
 - ▶ Projecting a site on response or explanatory variable **approx.** the value
 - ▶ angles between response and expl. variables **approx.** their correlations



```
...
Partitioning of correlations:
      Inertia Proportion
Total      27.000   1.0000
Constrained 20.177   0.7473
Unconstrained 6.823   0.2527

Eigenvalues, and their contribution to the correlations

Importance of components:
          RDA1   RDA2   RDA3   RDA4   RDA5   RDA6   RDA7
Eigenvalue     14.714  2.6433  1.1341  0.76821  0.33807  0.28135  0.09356
Proportion Explained  0.545  0.0979  0.0420  0.02848  0.01252  0.01042  0.00347
Cumulative Proportion 0.545  0.6429  0.6849  0.71333  0.72585  0.73627  0.73974
          RDA8   RDA9   RDA10  RDA11    PC1    PC2
Eigenvalue     0.08411  0.07592  0.02314  0.02129  2.44703  1.4094
Proportion Explained  0.00312  0.00281  0.00086  0.00079  0.09063  0.0522
Cumulative Proportion 0.74285  0.74566  0.74652  0.74731  0.83794  0.8901
...

```

Datasets
ooooooooUnconstrained Ordination
ooooooooooooooooooooConstrained Ordination
oooooooo●ooooModel diagnostics / testing
oooooooooooo

transformation-based RDA

- ▶ RDA (as PCA) preserves the euclidean distance.
- What about the species abundance paradox?
- ▶ Can transform data to use with euclidean distance
 - ▶ Remove differences in total abundance, while keeping the variations relative abundance
- ▶ Chord and Hellinger transformations useful.

Hellinger:

$$y'_{ij} = \sqrt{\frac{y_{ij}}{\sum_{j=1}^p y_{ij}}}$$

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

decostand(mat, 'hellinger')

      Spe1      Spe2      Spe3
sit1    0  0.5773503  0.8164966
sit2    0  0.7071068  0.7071068
sit3    1  0.0000000  0.0000000
attr(,"decostand")
[1] "hellinger"
```

Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280.

Datasets
ooooooooUnconstrained Ordination
ooooooooooooooooooooConstrained Ordination
oooooooo●ooooModel diagnostics / testing
oooooooooooo

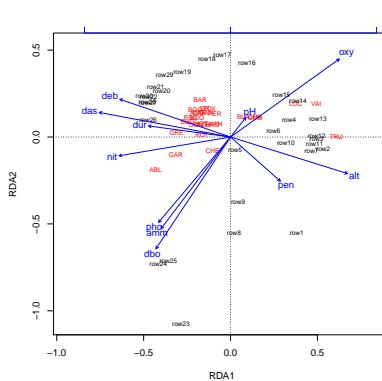
transformation-based RDA

39 / 54

```
# Hellinger transformation
Dabu_h <- decostand(Dabu, 'hellinger')
# RDA on Hellinger transformed abundances
tbRDA <- rda(Dabu_h ~ ., data = Denv)

# Plot
plot(tbRDA, type = 't')
```

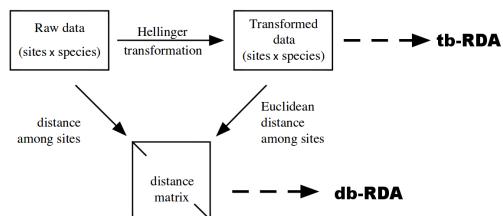
- ▶ alt, oxy and nutrients important
 - ▶ Trout (TRU) and minnow (VAI) found at high sites with high oxygen
 - ▶ Bleak (ABL) is found a low oxygen and high nutrients
 - ▶ Other species in similar environments



distance-based RDA

40 / 54

- ▶ db-RDA is a related method
 - ▶ hellinger transformation can also be expressed as distance matrix
 - ▶ *Constrained PCoA*
 - ▶ Can use every distance metric



modified from: Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280.



distance-based RDA

41 / 54

```
# dbRDA
dbRDA <- capscale(Dabu ~ ., data = Denv,
                     distance = 'bray')
plot(dbRDA, type = 't')
```

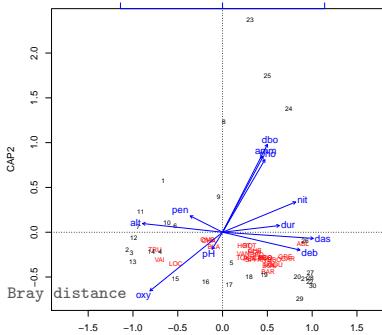
```
summary(dbRDA)

...
Partitioning of squared Bray distance:
  Inertia Proportion
Total      7.802    1.0000
Constrained 5.715    0.7325
Unconstrained 2.087   0.2675

Eigenvalues, and their contribution to the squared Bray distance oxy
Importance of components:
```

	CAP1	CAP2	CAP3	CAP4	CAP5	CAP6
Eigenvalue	3.2627	1.0283	0.53633	0.37863	0.24913	0.10755
Proportion Explained	0.4182	0.1318	0.06874	0.04853	0.03193	0.01379
Cumulative Proportion	0.4182	0.5500	0.61875	0.66729	0.69922	0.71300

```
...
```



Your turn!

Using the artificial dummy dataset and Site as only constraining variable

Run:

1. RDA
2. tbRDA (Hellinger)
3. dbRDA with Bray-Curtis
4. dbRDA with $x^{0.25}$ transformed abundances and Bray-Curtis

What ordination presents best the gradient?

What method explains most of variance?

See Demo.

Datasets

Constrained Ordination

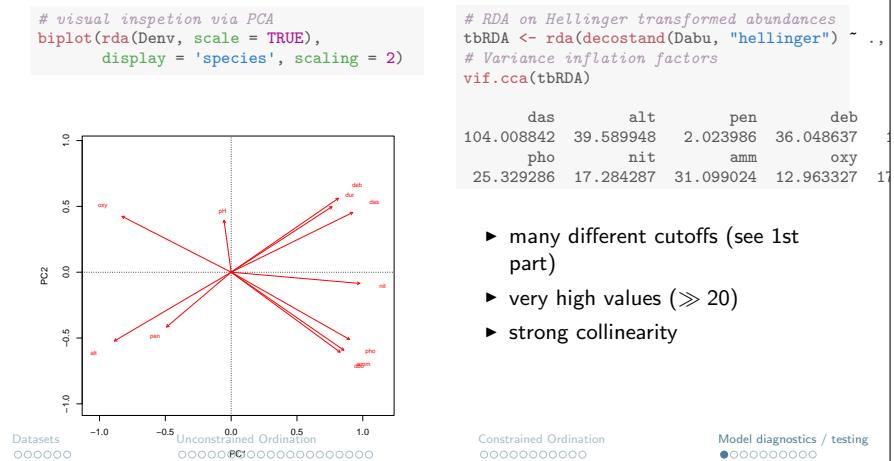
Model diagnostics / testing
○○○○○○○○○○

Model diagnostics and testing

Collinearity

45 / 54

- ▶ RDA et al. are *regression methods* - everything you know applies also here
- ▶ Collinearity of predictors may lead to wrong conclusions.
- ▶ Many methods available (see ref.). Additional: *Use your ecological knowledge!*



Goodness of fit

46 / 54

- ▶ cumulative variance explained by constraints
- ▶ available for *sites* or *species*
- ▶ summarize = TRUE gives the accumulated total variance (= last column)

```
# GOF for each species on
goodness(tbRDA)[ , 1:3]
```

	RDA1	RDA2	RDA3
CHA	0.16597466	0.28222282	0.72531373
TRU	0.68728597	0.68728738	0.75205565
VAI	0.66278142	0.76121936	0.77215444
LOC	0.50167024	0.63319561	0.63433575
OMB	0.16050386	0.25551654	0.64853919
BLA	0.04766735	0.17402885	0.47152162

...

```
# total var explained by constraints
goodness(tbRDA, summarize = TRUE)[1:6]
```

	CHA	TRU	VAI	LOC	OMB	BLA
1	0.8010478	0.8798422	0.8676846	0.8223702	0.7889173	0.5669931



- ▶ can decompose variance into constrained (CCA) and unconstrained (CA) parts

```
# variance explained by constrained (CCA) and unconstrained (CA) axes  
inertcomp(tbRDA, proportional = TRUE)
```

	CCA	CA
CHA	0.8010478	0.1989522
TRU	0.8798422	0.1201578
VAI	0.8676846	0.1323154
LOC	0.8223702	0.1776298
OMB	0.7889173	0.2110827
BLA	0.5669931	0.4330069
...		

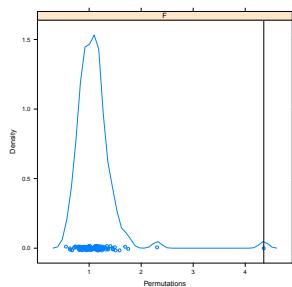


Permutation Tests (I)

- ▶ Cannot use parametric test theory
 - ▶ Use of *permutation tests* :

1. Shuffle the data (H_0 : No effect)
 2. Fit model to shuffled data
 3. Compute (pseudo-)F statistic for each model (Null distribution if H_0 is true)

$$p = \frac{(\text{No. of } F_{\text{perm}} \geq F))}{\text{Total No. of } F_{\text{perm}} + 1}$$



A number of different test can be applied to RDA:

- ▶ Test overall significance of model
- ▶ Test RDA axes
- ▶ Test terms:
 - ▶ sequential
 - ▶ marginal

Datasets
oooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oooooooooooo

Model diagnostics / testing
oooo●oooo

Test overall model

50 / 54

I omit some variables due to collinearity:

```
# RDA on Hellinger transformed abundances
tbRDA <- rda(decostand(Dabu, "hellinger") ~ alt + oxy + pH + nit + pho,
               data = Denv)
vif.cca(tbRDA)

alt      oxy      pH      nit      pho
2.929759 2.198671 1.047617 6.577783 4.164368
```

```
# Tests if the overall model is significant
anova(tbRDA)
```

```
Permutation test for rda under reduced model
Permutation: free
Number of permutations: 999

Model: rda(formula = decostand(Dabu, "hellinger") ~ alt + oxy + pH + das + nit + pho, data
          Df Variance   F Pr(>F)
Model    6  0.31490 6.4397  0.001 ***
Residual 23  0.18745
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Datasets
oooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oooooooooooo

Model diagnostics / testing
oooo●oooo

Test of RDA axes

51 / 54

```

# Tests RDA axes
anova(tbRDA, by = 'axis')

Permutation test for rda under reduced model
Marginal tests for axes
Permutation: free
Number of permutations: 999

Model: rda(formula = decostand(Dabu, "hellinger") ~ alt + oxy + pH + das + nit + pho, data = Dabu)
        Df Variance      F Pr(>F)
RDA1     1  0.217577 26.6970  0.001 ***
RDA2     1  0.047459  5.8233  0.001 ***
RDA3     1  0.030110  3.6946  0.001 ***
RDA4     1  0.012113  1.4863  0.143
RDA5     1  0.005714  0.7012  0.716
RDA6     1  0.001923  0.2359  0.995
Residual 23  0.187447
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Legendre, P., Oksanen, J. and ter Braak, C.J.F. (2011). Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution* 2, 269–277.



Sequential test of variables

52 / 54

- ▶ Variables are tested in the order they were specified (first to last)
 - ▶ Test of additional variance explained by adding the variable to the model
 - ▶ Order matters!

```
# Tests RDA axes
anova(tbRDA, by = 'terms')

Permutation test for rda under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 999

Model: rda(formula = decostand(Dabu, "hellinger") ~ alt + oxy + pH + das + nit + pho, data =
      Df Variance F Pr(>F)
alt     1 0.153953 18.8903 0.001 ***
oxy     1 0.085314 10.4682 0.001 ***
pH      1 0.004285 0.5258 0.755
das     1 0.031761 3.8971 0.009 **
nit     1 0.007244 0.8889 0.433
pho     1 0.032338 3.9680 0.008 **
Residual 23 0.187447
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



- ▶ Test explained variance of variable when all other variable are included in the model
- ▶ Order has no influence

```
# Tests RDA axes
anova(tbRDA, by = 'margin')

Permutation test for rda under reduced model
Marginal effects of terms
Permutation: free
Number of permutations: 999

Model: rda(formula = decostand(Dabu, "hellinger") ~ alt + oxy + pH + das + nit + pho, data
          Df Variance   F Pr(>F)
alt      1 0.023369 2.8674 0.031 *
oxy      1 0.037891 4.6493 0.003 **
pH       1 0.004184 0.5134 0.773
das      1 0.027131 3.3290 0.020 *
nit      1 0.014190 1.7411 0.118
pho      1 0.032338 3.9680 0.010 **
Residual 23 0.187447
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Datasets
oooooooo

Unconstrained Ordination
oooooooooooooooooooo

Constrained Ordination
oooooooooooo

Model diagnostics / testing
oooooooo●○

Multivariate topics not covered here:

- ▶ Model selection (be careful with automatic methods!)
- ▶ Distance-based hypothesis testing ((PER-)MANOVA, SIMPER, ANOSIM)
- ▶ Dispersion measures (β -Diversity, Functional diversity)
- ▶ Consensus RDA, RLQ (traits), ...
- ▶ Model-based multivariate framework (See work of David Warton et al.)
- ▶ many more