Introduction

What is Ordination?

Multivariate Analysis and Ordination

- Basic ordination methods to simplify multivariate data into low dimensional graphics
- Analysis of multivariate dependence and hypotheses
- Analyses can be performed in R statistical software using vegan package and allies
- Course homepage http://cc.oulu.fi/~jarioksa/opetus/metodi/
- Vegan homepage https://github.com/vegandevs/vegan/
Why Ordination?

- **Nobody** should want to make an ordination, but they are desperate with multivariate data
- Map multidimensional table into low-dimensional display
Gradient Analysis

- Gradient Analysis developed in 1950s in USA, with R. H. Whittaker as the main founding father
- Only two or three environmental variables, or *Gradients* needed to explain complicated community patterns
- Against classification: Species responses smooth along gradients
- Against organism analogies: Species responses individualistic
- The basis of modern theory and praxis: Ordination and Gradient modelling of communities
**Introduction to Gradient Analysis**

The Gradient Model


---

**Types of Gradients**

- **Direct gradients**: Influence organisms but are not consumed.
  - Correspond to conditions.

- **Resource gradients**: Consumed
  - Correspond to resources.

- **Complex gradients**: Covarying direct and/or resource gradients: Impossible to separate effects of single gradients.
  - Most observed gradients.
Species responses

- Species have non-linear responses along gradients.
- Often assumed to be Gaussian...
Gaussian Response Function has three interpretable parameters that define the expected response $\mu$ along the gradient $x$:

- Location of the optimum $u$ on the gradient $x$
- Width of the response $t$ in the units of gradient $x$
- Height of the response $h$ in the units of response height $\mu$

$$\mu = h \times \exp\left(\frac{(x-u)^2}{2t^2}\right)$$

Dream of species packing

Species have Gaussian responses and divide the gradient optimally:

- Equal heights $h$.
- Equal widths $t$.
- Evenly distributed optima $u$. 
Introduction

Gradient Analysis

Evidence for Gaussian Responses

- Whittaker reported a large number of different response types
- Only a small proportion were symmetric, bell shaped responses
- Still became the standard of our times
- Comparison of ordination methods based on simulation, and many of those use Gaussian responses
- We need to use simulation because then we know the truth that should be found

Ordination

- Ordination maps multivariate data onto low dimensional displays: “Most data sets have 2.5 dimensions”
- Gradients define vegetation: ordination tries to find the underlying gradients
- Basic ordination uses only community composition: Indirect Gradient Analysis
- Constrained ordination studies only the variation that can be explained by the available environmental variables: Often called Direct Gradient Analysis
- Distinct flavours of tools:
  - Nonmetric MDS the most robust method
  - PCA duly despised
  - Flavours of Correspondence Analysis popular
  - Canonical method: Constrained Correspondence Analysis
Unconstrained Ordination

Nonmetric Multidimensional Scaling

- Rank-order relation with (1) community dissimilarities and (2) ordination distances: No specified form of regression, but the best shape is found from the data.
- Non-linear regression can cope with non-linear species responses of various shapes: Not dependent on Gaussian model.
- Iterative solution: No guarantee of convergence.
- Must be solved separately for each number of dimensions: A lower dimensional solutions is not a subset of a higher, but each case is solved individually.
- A test winner, and a natural choice...

From Ranks of Dissimilarities to Ordination Distances

Observed dissimilarities:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.467</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.636</td>
<td>0.511</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.524</td>
<td>0.356</td>
<td>0.634</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.843</td>
<td>0.600</td>
<td>0.753</td>
<td>0.513</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.922</td>
<td>0.693</td>
<td>0.852</td>
<td>0.667</td>
<td>0.606</td>
</tr>
</tbody>
</table>

Ranks of observed dissimilarities:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>12</td>
<td>6</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>

Ordination distances:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.301</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.539</td>
<td>0.303</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.323</td>
<td>0.133</td>
<td>0.421</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0.615</td>
<td>0.414</td>
<td>0.612</td>
<td>0.307</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.922</td>
<td>0.636</td>
<td>0.636</td>
<td>0.605</td>
<td>0.416</td>
</tr>
</tbody>
</table>
MDS is a map

- MDS tries to draw a map using distance data.
- MDS tries to find an underlying configuration from dissimilarities.
- Only the configuration counts:
  - No origin, but only the constellations.
  - No axes or natural directions, but only a framework for points.

Map of Europe from road distances

![Map of Europe from road distances](image)

Shepard Diagram

![Shepard Diagram](image)

- Non-metric fit, $R^2 = 0.967$
- Linear fit, $R^2 = 0.835$
Recommended procedure

NMDS may be good, but its use needs special care: Not every NMDS automatically is good.

1. Use adequate dissimilarity indices: An adequate index gives a good rank-order relation between community dissimilarity and gradient distance.
2. No convergence guaranteed: Start with several random starts and inspect those with lowest stress.
3. Satisfied only if minimum stress configurations are similar.
metaMDS I

> vare.mds <- metaMDS(varespec)

Square root transformation
Wisconsin double standardization
Run 0 stress 0.184
Run 1 stress 0.196
Run 2 stress 0.185
... procrustes: rmse 0.0494  max resid 0.158
Run 3 stress 0.209
Run 4 stress 0.215
Run 5 stress 0.235
Run 6 stress 0.196
Run 7 stress 0.234
Run 8 stress 0.196
Run 9 stress 0.222
Run 10 stress 0.185
Run 11 stress 0.195
Run 12 stress 0.229
Run 13 stress 0.184
... New best solution
... procrustes: rmse 3.6e-05  max resid 0.000139
*** Solution reached

metaMDS II

> vare.mds

Call:
metaMDS(comm = varespec)

global Multidimensional Scaling using monoMDS

Data:     wisconsin(sqrt(varespec))
Distance: bray

Dimensions: 2
Stress:     0.184
Stress type 1, weak ties
Two convergent solutions found after 13 tries
Scaling: centring, PC rotation, halfchange scaling
Species: expanded scores based on ‘wisconsin(sqrt(varespec))’
Badness of fit measure **stress** is based on the residuals from the non-linear regression:

- A proportional measure in the range 0 (perfect) ... 1 (desperate) related to goodness of fit measure $1 - R^2$
- Random configuration typically $\approx 0.4$ and 0 degenerate
- Often given in percents (but omitting the percent sign: $15 = 0.15$, since cannot be $> 1$)

Orientation, rotation, scale and origin of the coordinates (scores) are indeterminate: only the constellation matters.

Vegan arbitrarily fixes some of these:

- Axes are centred, but the origin has no special meaning
- Axes are rotated so that the first is the longest (technically: rotated to principal components)
- Axes are scaled so that one unit corresponds to halving of similarity from the “replicate similarity”
- The sign (direction) of the axes still undefined
- Replicate similarity: dissimilarity at ordination distance = 0
- Maximum dissimilarity = 1: nothing in common
- Linear area of ordination distance – dissimilarity: 0 . . . 0.8

What happened in metaMDS?

- Square root transformation and Wisconsin double standardization
- Bray–Curtis dissimilarities
- monoMDS with several random starts and stopping after finding two identical minimum stress solutions
- Solution rotated to PCs
- Solution scaled to half-change units
- Species scores as weighted averages
Dissimilarity measures

- Use a dissimilarity that describes correctly gradient separation
- Bray–Curtis (Steinhaus), Jaccard, Kulczyński
- Wisconsin double standardization often helpful
- Should use dissimilarities which reach their maximum (1) when no species are shared (like those listed above)
- Indices with no bound maximum are usually bad (Euclidean distance etc.)

Procrustes rotation

- Procrustes rotation to maximal similarity between two configurations:
  - Translate the origin.
  - Rotate the axes.
  - Deflate or inflate the axis scale.
- Single points can move a lot, although the stress is fairly constant: Especially in large data sets.
> tmp <- wiscosin(sqrt(varespec))
> dis <- vegdist(tmp)
> vare.mds0 <- monomds(dis, trace = 0)
> pro <- procrustes(vare.mds, vare.mds0)
> pro

Call:
procrustes(X = vare.mds, Y = vare.mds0)

Procrustes sum of squares:
0.186
In NMDS, 2D solution is not a plane in 3D space
Solution must be found separately for each dimensionality
Some people very disturbed: how do they know the correct number
Answer is easy: there is no correct number, although some numbers may be worse than others
“Most data sets have 2.5 dimensions”
Typically you try with 2 and 3
Do you need more dimensions to explain species patterns and environmental data?
Is convergence very slow? Try another number of dimensions
Scree plot or stress against the number of dimensions often suggested but rarely works
Simplified mapping: Eigen analysis

- NMDS uses non-linear mapping for any dissimilarity measure: This is very difficult
- Things are much simpler if we accept only certain dissimilarity indices and map them linearly onto ordination
- Linear mapping is only a rotation, and can be solved using eigenvector techniques
- Sometimes said that certain methods are model-based (CA), but they also employ a distance

<table>
<thead>
<tr>
<th>method</th>
<th>metric</th>
<th>mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMDS</td>
<td>any</td>
<td>nonlinear</td>
</tr>
<tr>
<td>MDS</td>
<td>any</td>
<td>linear</td>
</tr>
<tr>
<td>PCA</td>
<td>Euclidean</td>
<td>linear</td>
</tr>
<tr>
<td>CA</td>
<td>Chi-square</td>
<td>weighted linear</td>
</tr>
</tbody>
</table>

Why Not PCA?

- We admit that PCA is just a rotation, but it is a linear method
- PCA works with species space, but we boldly go to gradient space
- CA is an optimal scaling method
  - Sites with similar species composition packed close to each other
  - Species that occur together simultaneously packed close to each other
- CA can handle unimodal species responses, even approximate one dimensional species packing model
Graphical presentations of data matrix: Species are axes and span the space where sites are points.

Some species show more of the configuration than others.

What is the ideal viewing angle to the species space?

Shows as much as possible of all species in just two or three dimensions.

---

Rotation in species space

1. Put sites into species space.
2. Move the origin to the centroid.
3. Rotate the axes so that the first axis (1) is as close to all points as possible, and (2) explains as much of the variance as possible.
Goodness of Fit

- The total variation ($\Lambda$) is the sum of squared distances of points from the origin.
- $\Lambda$ can be expressed as the sum of squares (SS) or variance ($SS/n$ or $SS/(n-1)$).
- The points are projected on the axis, and the sum of projected squared distances is the eigenvalue of the axis ($\lambda_i$).
- The eigenvalues are ordered and non-negative $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and sum up to total variance $\Lambda = \sum_{i=1}^{p} \lambda_i$.
- $\lambda_i/\Lambda$ gives the proportion that an axis explains of the total variance, and $\lambda_1$ explains the largest proportion.
- Cumulative sum gives the proportion of variance explained by the first axes: often emphasized but rather useless statistic.
- PCA is often used to reduce data into a few linearly independent components that explain the most of the original variables.

---

Euclidean Metric of PCA

\[
\begin{align*}
  d_{jk} &= \sum (x_{ij} - x_{ik})^2 \\
  s_j &= \sum (x_{ij})^2 \\
  s_k &= \sum (x_{ik})^2 \\
  \theta_{jk} &= \frac{\sum x_{ij} x_{ik}}{\sqrt{(\sum x_{ij}^2)(\sum x_{ik}^2)}}
\end{align*}
\]
Running PCA I

> (ord <- rda(dune))
Call: rda(X = dune)

Inertia Rank
Total 84.1
Unconstrained 84.1 19
Inertia is variance

Eigenvalues for unconstrained axes:
   PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8
  24.80 18.15  7.63  7.15  5.70  4.33  3.20  2.78
(Showed only 8 of all 19 unconstrained eigenvalues)

> head(summary(ord), 3, 1)

Running PCA II

Call:
  rda(X = dune)

Partitioning of variance:
Inertia Proportion
Total 84.1 1
Unconstrained 84.1 1

Eigenvalues, and their contribution to the variance

Importance of components:

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>24.795</td>
<td>18.147</td>
<td>7.6291</td>
<td>7.153</td>
<td>5.6950</td>
<td>4.3333</td>
</tr>
<tr>
<td>Proportion Explained</td>
<td>0.295</td>
<td>0.216</td>
<td>0.0907</td>
<td>0.085</td>
<td>0.0677</td>
<td>0.0515</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.295</td>
<td>0.510</td>
<td>0.6011</td>
<td>0.686</td>
<td>0.7539</td>
<td>0.8054</td>
</tr>
<tr>
<td>PC7</td>
<td>PC8</td>
<td>PC9</td>
<td>PC10</td>
<td>PC11</td>
<td>PC12</td>
<td></td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>3.199</td>
<td>2.7819</td>
<td>2.4820</td>
<td>1.854</td>
<td>1.7471</td>
<td>1.3136</td>
</tr>
<tr>
<td>Proportion Explained</td>
<td>0.038</td>
<td>0.0331</td>
<td>0.0295</td>
<td>0.022</td>
<td>0.0208</td>
<td>0.0156</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.843</td>
<td>0.8765</td>
<td>0.9060</td>
<td>0.928</td>
<td>0.9488</td>
<td>0.9644</td>
</tr>
<tr>
<td>PC13</td>
<td>PC14</td>
<td>PC15</td>
<td>PC16</td>
<td>PC17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>0.9905</td>
<td>0.63779</td>
<td>0.55083</td>
<td>0.35058</td>
<td>0.19956</td>
<td></td>
</tr>
<tr>
<td>Proportion Explained</td>
<td>0.0118</td>
<td>0.00758</td>
<td>0.00655</td>
<td>0.00417</td>
<td>0.00237</td>
<td></td>
</tr>
</tbody>
</table>
Running PCA III

Cumulative Proportion 0.9762 0.98377 0.99032 0.99448 0.99686
                        PC18   PC19
Eigenvalue          0.14880 0.11575
Proportion Explained 0.00177 0.00138
Cumulative Proportion 0.99862 1.00000

Scaling 2 for species and site scores
* Species are scaled proportional to eigenvalues
* Sites are unscaled: weighted dispersion equal on all dimensions
* General scaling constant of scores: 6.3229

Species scores

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achimill</td>
<td>-0.6038</td>
<td>0.124</td>
<td>0.00846</td>
<td>0.160</td>
<td>0.4087</td>
</tr>
<tr>
<td>Agrostoi</td>
<td>1.3740</td>
<td>-0.964</td>
<td>0.16691</td>
<td>0.266</td>
<td>-0.0877</td>
</tr>
<tr>
<td>Airaprae</td>
<td>0.0234</td>
<td>0.251</td>
<td>-0.19477</td>
<td>-0.326</td>
<td>0.0557</td>
</tr>
<tr>
<td>Callcusp</td>
<td>0.5385</td>
<td>0.180</td>
<td>0.17509</td>
<td>0.239</td>
<td>0.2553</td>
</tr>
</tbody>
</table>

Running PCA IV

Site scores (weighted sums of species scores)

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.857</td>
<td>-0.172</td>
<td>2.608</td>
<td>-1.130</td>
<td>0.4507</td>
</tr>
<tr>
<td>2</td>
<td>-1.645</td>
<td>-1.230</td>
<td>0.887</td>
<td>-0.986</td>
<td>2.0346</td>
</tr>
<tr>
<td>3</td>
<td>-0.440</td>
<td>-2.383</td>
<td>0.930</td>
<td>-0.460</td>
<td>-1.0278</td>
</tr>
<tr>
<td>20</td>
<td>2.341</td>
<td>1.299</td>
<td>0.903</td>
<td>0.718</td>
<td>-0.0757</td>
</tr>
</tbody>
</table>
The scores are centred (= their mean is zero) and either normalized (= all have equal spread) or proportional to eigenvalues (= spread is higher when eigenvalue is high)

Normalized scores give the regression coefficients between the axis and the variables: often used for species

Scores proportional to the eigenvalue give the true configuration of points in the space defined by normalized scores: often used for sites (hence in species space)

Together these scores give a linear least square approximation of the data

Graphical presentation called **biplot**

However, there are many alternative scaling systems
Reading the Plot

- Origin: all species (variables) at their average values
- The *distance* from the origin for a row (site) implies how much the point differs from the average
- The *distance* from the origin for a column (species, variable) implies how much the point increases to that *direction*
- The change is measured in absolute scale: big changes, long distances from the origin
- Implies a linear model of species response against axes
- The *angle* between two points implies correlations
  - $90^\circ$ means zero correlation, $< 90^\circ$ positive correlation, $> 90^\circ$ negative correlation, $0^\circ$ implies $r = 1$
- Arrow biplots often used instead of point biplot
Variance and Correlations

- Analysis of raw data explains variances: variables with high variance are most important.
- If the variables are standardized to unit variance before analysis, $z = (x - \bar{x})/s_x$ all variables are equally important and the analysis explains correlations among variables.
- Standardization can be used when we want all variables to have equal weights.
- Standardization must be used when variables are measured in different scales, such as for environmental measurements.
Reducing the Number of Correlated Environmental Variables I

> (pc <- rda(varechem, scale=TRUE))

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

<table>
<thead>
<tr>
<th>Inertia Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Unconstrained</td>
</tr>
</tbody>
</table>

Inertia is correlations

Eigenvalues for unconstrained axes:

<table>
<thead>
<tr>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
<th>PC8</th>
<th>PC9</th>
<th>PC10</th>
<th>PC11</th>
<th>PC12</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.19</td>
<td>3.19</td>
<td>1.69</td>
<td>1.07</td>
<td>0.82</td>
<td>0.71</td>
<td>0.44</td>
<td>0.37</td>
<td>0.17</td>
<td>0.15</td>
<td>0.09</td>
<td>0.07</td>
</tr>
</tbody>
</table>

PC13 | PC14
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The Number of Components

- PCA is a rotation in species (character) space and retains the original configuration
- The number of PC’s is \( \min(N, S) \), and all together give the original data
- First axes are most important and we may ignore the minor axes
- We can either use the axes as variables in other models, or use them to identify major (almost) independent variables
- Often we want to retain a certain proportion of the variance, say 50 %
- Sometimes we would like to retain “significant” axes
- There really is no way of doing this, but some people suggest comparing eigenvalues against broken stick distribution
**Unconstrained Ordination**

**PCA**

**Broken Stick and Eigenvalues**

Inertia

0 1 2 3 4 5

```
●
●
●
●
●
●
●
●
●
●
```

**Two Dimensions, but which?**

```
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9
```

```
pc
```

```
Baresoil Humdepth Baresoil Mn N
```

```
P K Zn P S Mo
```

```
Al Fe
```

```
http://cc.oulu.fi/ jarioksa/ (Oulu) Multivariate Analysis in Ecology January 2016 51 / 103
```

```
http://cc.oulu.fi/ jarioksa/ (Oulu) Multivariate Analysis in Ecology January 2016 52 / 103
```
**Metric Scaling** a.k.a. **Principal Coordinates Analysis**
- Used dissimilarities instead of raw data
- With Euclidean distances equal to PCA, but can use other dissimilarities

**Factor Analysis**
- A *statistical* method that makes a difference between systematic components and random error
- In PCA we just ignore latter components, but here we really identify the real components
- Much used in human sciences and often referred to in ecology (but usually misunderstood)
Minor variant of PCA: Weighted Principal Components with Chi-square metric

All sites should have all species in in the same proportions as in the whole data

Site and species marginal profiles define the expected abundances

Null model: Species composition is identical in all sampling units

Chi-square transformation tells how much the observed proportions $f_{ij}$ differ from the expected proportions $e_{ij}$:

$$\chi_{ij} = \frac{f_{ij} - e_{ij}}{\sqrt{e_{ij}}}$$
Running CA I

> (ord <- cca(dune))

Call: cca(X = dune)

Inertia Rank
Total  2.12
Unconstrained  2.12  19

Eigenvalues for unconstrained axes:
  CA1  CA2  CA3  CA4  CA5  CA6  CA7  CA8
  0.536  0.400  0.260  0.176  0.145  0.108  0.092  0.081
(Showed only 8 of all 19 unconstrained eigenvalues)

> head(summary(ord), 2, 1)
### Running CA II

Call:
```
cca(X = dune)
```

Partitioning of mean squared contingency coefficient:

<table>
<thead>
<tr>
<th>Inertia Proportion</th>
<th>Total</th>
<th>2.12</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unconstrained</td>
<td>2.12</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Eigenvalues, and their contribution to the mean squared contingency coefficient

<table>
<thead>
<tr>
<th>Importance of components:</th>
<th>CA1</th>
<th>CA2</th>
<th>CA3</th>
<th>CA4</th>
<th>CA5</th>
<th>CA6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>0.536</td>
<td>0.400</td>
<td>0.260</td>
<td>0.176</td>
<td>0.1448</td>
<td>0.108</td>
</tr>
<tr>
<td>Proportion Explained</td>
<td>0.253</td>
<td>0.189</td>
<td>0.123</td>
<td>0.0832</td>
<td>0.0684</td>
<td>0.051</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.253</td>
<td>0.443</td>
<td>0.565</td>
<td>0.6486</td>
<td>0.7170</td>
<td>0.768</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Importance of components:</th>
<th>CA7</th>
<th>CA8</th>
<th>CA9</th>
<th>CA10</th>
<th>CA11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>0.0925</td>
<td>0.0809</td>
<td>0.0733</td>
<td>0.0563</td>
<td>0.0483</td>
</tr>
<tr>
<td>Proportion Explained</td>
<td>0.0437</td>
<td>0.0382</td>
<td>0.0347</td>
<td>0.0266</td>
<td>0.0228</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.8117</td>
<td>0.8500</td>
<td>0.8847</td>
<td>0.9113</td>
<td>0.9341</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Importance of components:</th>
<th>CA12</th>
<th>CA13</th>
<th>CA14</th>
<th>CA15</th>
<th>CA16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>0.0412</td>
<td>0.0352</td>
<td>0.02053</td>
<td>0.01491</td>
<td>0.00907</td>
</tr>
<tr>
<td>Proportion Explained</td>
<td>0.0195</td>
<td>0.0167</td>
<td>0.00971</td>
<td>0.00705</td>
<td>0.00429</td>
</tr>
</tbody>
</table>

### Running CA III

Cumulative Proportion 0.9536 0.9702 0.97995 0.98700 0.99129

<table>
<thead>
<tr>
<th>CA17</th>
<th>CA18</th>
<th>CA19</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00794</td>
<td>0.00700</td>
<td>0.00348</td>
</tr>
</tbody>
</table>

Scaling 2 for species and site scores

* Species are scaled proportional to eigenvalues
* Sites are unscaled: weighted dispersion equal on all dimensions

Species scores

<table>
<thead>
<tr>
<th>CA1</th>
<th>CA2</th>
<th>CA3</th>
<th>CA4</th>
<th>CA5</th>
<th>CA6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achimill -0.909</td>
<td>0.0846</td>
<td>-0.586</td>
<td>-0.00892</td>
<td>-0.660</td>
<td>-0.1888</td>
</tr>
<tr>
<td>Agrostol 0.934</td>
<td>-0.2065</td>
<td>0.282</td>
<td>0.02429</td>
<td>-0.139</td>
<td>-0.0226</td>
</tr>
</tbody>
</table>

Callcusp 1.952 0.5674 -0.859 -0.09897 -0.557 0.2328

Site scores (weighted averages of species scores)
Goodness of Fit of Scores

- Inertia is “mean square contingency coefficient”: Chi-squared of a matrix standardized to unit sum, or Chi-square of \( \frac{x}{\sum x} \)
- Eigenvalues are non-negative and ordered like in PCA, but they are bound to maximum 1
- The origin gives the expected abundances for all species and all sites
- The deviant species and deviant sites are far away from the origin
- CA is weighted analysis, and the weighted sum of squared scores is the eigenvalue
- The species and site scores are (scaled) weighted averages of each other: proximity matters
- Rare species have low weights: they are further away from the origin

<table>
<thead>
<tr>
<th>CA1</th>
<th>CA2</th>
<th>CA3</th>
<th>CA4</th>
<th>CA5</th>
<th>CA6</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.812</td>
<td>-1.083</td>
<td>-0.1448</td>
<td>-2.107</td>
<td>-0.393</td>
<td>-1.8346</td>
</tr>
<tr>
<td>-0.633</td>
<td>-0.696</td>
<td>-0.0971</td>
<td>-1.187</td>
<td>-0.977</td>
<td>0.0658</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.944</td>
<td>1.069</td>
<td>-0.6660</td>
<td>-0.553</td>
<td>1.596</td>
<td>-1.7029</td>
</tr>
</tbody>
</table>
Weighted Average?

- For presence/absence data: weighted average of a species is in the middle ("barycentre") of plots where the species occurs
- For quantitative data: plots where species is abundant are heavier and the weighted average is closer to them
- Sampling units (SU) are close to species that occur on them
- CA is a weighted average method: it tries to put SUs close to the species that occur in them, and all SUs with similar species composition close to each other: Unimodal response

Default Plot and Effect of Scaling

scaling = 1

- For presence/absence data: weighted average of a species is in the middle ("barycentre") of plots where the species occurs
- For quantitative data: plots where species is abundant are heavier and the weighted average is closer to them
- Sampling units (SU) are close to species that occur on them
- CA is a weighted average method: it tries to put SUs close to the species that occur in them, and all SUs with similar species composition close to each other: Unimodal response
Weighted Averages

- Species scores are proportional to weighted averages of site scores, and simultaneously
- Site scores are proportional to weighted averages of species scores
- Either one (but not both) of these can be a direct weighted average of other
- If site scores are weighted averages of species scores, site point is in the middle of points of species that occurs in the site
- The location of the point is meaningful whereas in PCA the main things were distance and direction from the origin (but these, too, matter)
- Can approximate unimodal response model and therefore CA is much better for community ordination than PCA

Linear and Unimodal Models

- PCA implies linear relations between axes and species abundances
- CA packs species and approximates a unimodal model
Optimal Scaling

- The locations of species optima (tops) should be widespread: spread is measured as $SS_B$
- The species responses should be narrow: width is measured as $SS_w$
- The total variance is their sum $SS_T = SS_B + SS_w$
- High $SS_B$ means that species have different optima, and low $SS_w$ means that species have narrow tolerance
- Scaling is optimal if most of variance is between species and $SS_B$ is high
- The criterion of variance is the eigenvalue maximized in CA: $\lambda = SS_B / SS_T$

Goodness of Fit Statistics: Repetition

- **NMDS**: stress of nonlinear transformation from observed dissimilarities to ordination distances
  - In range 0...1 (0...100%), but in practice 0.4 for random configuration
  - 0.1 is good, and 0.2 is not bad, 0 is suspect
- **PCA**: sum of eigenvalues is variance (or SS)
  - Upper limit is total variance, large is good
- **CA**: sum of all eigenvalues is (scaled) Chi-square
  - Single eigenvalue maximum 1
  - high is good, but $\lambda < 0.2$ may not be bad
  - Eigenvalues $\lambda > 0.7$ are suspect: disjunct or very heterogeneous data
Nonlinear and Linear Mapping: A Difficult Case

NMDS
Non-metric fit, R² = 0.98
Linear fit, R² = 0.92

PCA

CA

Dune Meadow data


http://cc.oulu.fi/~jarioksa/ (Oulu)  Multivariate Analysis in Ecology  January 2016  70 / 103
Anatomy of a Plot

Plotting functions

- All vegan ordination functions have a plot function, and ordiplot can be used for other functions as well
- For full control, use first plot(x, type="n") and then add configurable points or text
- Congested plots can be displayed with orditorp or edited with orditkplot
- Lattice graphics can be made with ordixyplot, ordicloud or ordisplom
- Dynamic, spinnable 3D plots can be made with ordirgl function in the vegan3d package
- Items can be added to the plots with ordiarrow, ordihull, ordispider, ordihull, ordiellipse, ordisegments, or ordigrid
We take granted that vegetation is controlled by environment, so

1. Two sites close to each other in ordination have similar vegetation
2. If two sites have similar vegetation, they have similar environment
3. Two sites far away from each other in ordination have dissimilar vegetation, and perhaps
4. If two sites have different vegetation, they have different environment

**Fitted Vectors**

- **Direction** of fitted vector shows the gradient of the environmental variable, **length** shows its importance.
- For every arrow, there is an equally long arrow into opposite direction: Decreasing direction of the gradient.
- Implies a linear model: Project sample plots onto the vector for expected value.
Alternatives to Vectors

- Fitted vectors natural in constrained ordination, since these have linear constraints.
- Distant sites are different, but may be different in various ways: Environmental variables may have a non-linear relation to ordination.
> (ef <- envfit(vare.mds, varechem, permu = 999))

***VECTORS

<table>
<thead>
<tr>
<th></th>
<th>NMDS1</th>
<th>NMDS2</th>
<th>r2</th>
<th>Pr(&gt;r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>-0.050</td>
<td>-0.999</td>
<td>0.21</td>
<td>0.098</td>
</tr>
<tr>
<td>P</td>
<td>0.687</td>
<td>0.727</td>
<td>0.18</td>
<td>0.135</td>
</tr>
<tr>
<td>K</td>
<td>0.827</td>
<td>0.562</td>
<td>0.17</td>
<td>0.147</td>
</tr>
<tr>
<td>Ca</td>
<td>0.750</td>
<td>0.661</td>
<td>0.28</td>
<td>0.029 *</td>
</tr>
<tr>
<td>Mg</td>
<td>0.697</td>
<td>0.717</td>
<td>0.35</td>
<td>0.015 *</td>
</tr>
<tr>
<td>S</td>
<td>0.276</td>
<td>0.961</td>
<td>0.18</td>
<td>0.143</td>
</tr>
<tr>
<td>Al</td>
<td>-0.838</td>
<td>0.546</td>
<td>0.52</td>
<td>0.002 **</td>
</tr>
<tr>
<td>Fe</td>
<td>-0.862</td>
<td>0.507</td>
<td>0.40</td>
<td>0.013 *</td>
</tr>
<tr>
<td>Mn</td>
<td>0.802</td>
<td>-0.597</td>
<td>0.53</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>Zn</td>
<td>0.665</td>
<td>0.747</td>
<td>0.18</td>
<td>0.146</td>
</tr>
<tr>
<td>Mo</td>
<td>-0.849</td>
<td>0.529</td>
<td>0.05</td>
<td>0.581</td>
</tr>
<tr>
<td>Baresoil</td>
<td>0.872</td>
<td>-0.490</td>
<td>0.25</td>
<td>0.035 *</td>
</tr>
<tr>
<td>Humdepth</td>
<td>0.926</td>
<td>-0.377</td>
<td>0.56</td>
<td>0.001 ***</td>
</tr>
<tr>
<td>pH</td>
<td>-0.799</td>
<td>0.601</td>
<td>0.26</td>
<td>0.042 *</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999

Plotting Environmental Vectors
Limit \( p < 0.1 \)
> ef <- envfit(vare.mds ~ Al + Ca, varechem)
> plot(vare.mds, display = "sites")
> plot(ef)
> tmp <- with(varechem, ordisurf(vare.mds, Al, add = TRUE))
> tmp <- with(varechem, ordisurf(vare.mds, Ca, add = TRUE, col = "green4"))
Factor Fitting I

> dune.ca <- cca(dune)
> ef <- envfit(dune.ca ~ A1 + Management, data=dune.env, perm=999)
> ef

***VECTORS

<table>
<thead>
<tr>
<th></th>
<th>CA1</th>
<th>CA2</th>
<th>r2 Pr(&gt;r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.9980</td>
<td>0.0606</td>
<td>0.31 0.052</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999

***FACTORS:

Centroids:

<table>
<thead>
<tr>
<th></th>
<th>CA1</th>
<th>CA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ManagementBF</td>
<td>-0.73</td>
<td>-0.14</td>
</tr>
<tr>
<td>ManagementHF</td>
<td>-0.39</td>
<td>-0.30</td>
</tr>
<tr>
<td>ManagementNM</td>
<td>0.65</td>
<td>1.44</td>
</tr>
<tr>
<td>ManagementSF</td>
<td>0.34</td>
<td>-0.68</td>
</tr>
</tbody>
</table>

Factor Fitting II

Goodness of fit:

<table>
<thead>
<tr>
<th></th>
<th>r2 Pr(&gt;r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management</td>
<td>0.44 0.003 **</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999
Environmental variables need not be parallel to ordination axes.

Axes cannot be taken as gradients, but gradients are oblique to axes: You cannot tear off an axis from an ordination.

Never calculate a correlation between an axis and an environmental variable.

Environmental variables need not be linearly correlated with the ordination, but locations in ordination can be exceptional.
Single gradients appear as curves in linear ordination methods

PCA *horseshoe*: curve bends inward and gives wrong ordering of points on axis 1

CA *arch*: axis 1 retains the correct ordering of sites despite the curve

Environmental interpretation by vector fitting or surface bound to be biased

Axes cannot be interpreted as “gradients”

---

**The birth of the curve**

- There is a curve in the species space and PCA shows it correctly
- CA deals better with unimodal responses, but the second optimal scaling axis is folded first axis
Detrended Correspondence Analysis (DCA)

- CA axis retains the correct ordering: keep that, but instead of orthogonal axes, use detrended axes
- Programme DECORANA additionally rescales axes to $sd$ units approximating $t$ parameter of the Gaussian model
- Distorts space, introduces new artefacts and probably should be avoided

Nonmetric Multidimensional Scaling (NMDS) should be able to cope with moderately long gradients

Constrained ordination may linearize the responses

Running Detrended Correspondence Analysis

```r
> (ord <- decorana(dune))
Call:
decorana(veg = dune)

Detrended correspondence analysis with 26 segments.
Rescaling of axes with 4 iterations.

                          DCA1  DCA2  DCA3  DCA4
Eigenvalues  0.512 0.304 0.1213 0.1427
Decorana values 0.536 0.287 0.0814 0.0481
Axis lengths     3.700 3.117 1.3005 1.4789
```
Community Pattern Simulation
Short Gradients: Is There a Niche for PCA?

- Folklore: PCA with short gradients ($\leq 2t$).
- Not based on research, but simulation finds PCA uniformly worse than CA: With short gradients about as good as CA, but usually worse.
- There should be no species optimum within gradient: Shortness alone not sufficient.
- PCA best used for really linear cases (environment) or for reduction of variables into principal components (but see FA).
- Noise dominates over signal in homogeneous data.

Long Gradients: DCA or NMDS

- Curvature with long gradients: Need either DCA or NMDS.
- NMDS is a test winner: More robust than DCA.
- DCA more popular.
- DCA may produce new artefacts, since it twists the space.
Extended Dissimilarities and Step-across

- How different are sites that have nothing in common?
- Use step-across points to estimate their distance
- Flexible shortest path or their approximations, extended dissimilarities
- Extended dissimilarity: use only one-site steps, do not update dissimilarities below a threshold
- No shared species since rare species were not observed: Swan transformation estimates the probability of finding an unobserved species

Strong and Weak Ties

- Maximum dissimilarities (no shared species) are tied
- Strong tie treatment tries to keep tied values together and puts maximum dissimilarites to a circle
- Weak tie treatment allows breaking ties and straightens the axes: now the default in vegan, whereas earlier was impossible