Vegan: an introduction to ordination

Jari Oksanen

processed with vegan 2.4-3 in R version 3.3.1 (2016-06-21) on April 6, 2017

Abstract

The document describes typical, simple work pathways of vegetation ordination. Unconstrained ordination uses as examples detrended correspondence analysis and non-metric multidimensional scaling, and shows how to interpret their results by fitting environmental vectors and factors or smooth environmental surfaces to the graph. The basic plotting command, and more advanced plotting commands for congested plots are also discussed, as well as adding items such as ellipses, convex hulls, and other items for classes. The constrained ordination uses constrained (canonical) correspondence analysis as an example. It is first shown how a model is defined, then the document discusses model building and significance tests of the whole analysis, single constraints and axes.

Contents

1 Ordination 2
  1.1 Detrended correspondence analysis 2
  1.2 Non-metric multidimensional scaling 2

2 Ordination graphics 4
  2.1 Cluttered plots 5
  2.2 Adding items to ordination plots 5

3 Fitting environmental variables 6

4 Constrained ordination 7
  4.1 Significance tests 9
  4.2 Conditioned or partial ordination 11

Vegan is a package for community ecologists. This document explains how the commonly used ordination methods can be performed in vegan. The document only is a very basic introduction. Another document (vegan tutor) [http://cc.oulu.fi/~jarioksa/opetus/method/vegantutor.pdf] gives a longer and more detailed introduction to ordination. The current document only describes a small part of all vegan functions. For most functions, the canonical references are the vegan help pages, and some of the most important additional functions are listed at this document.
1 Ordination

The vegan package contains all common ordination methods: Principal component analysis (function rda, or prcomp in the base R), correspondence analysis (cca), detrended correspondence analysis (decorana) and a wrapper for non-metric multidimensional scaling (metaMDS). Functions rda and cca mainly are designed for constrained ordination, and will be discussed later. In this chapter I describe functions decorana and metaMDS.

1.1 Detrended correspondence analysis

Detrended correspondence analysis (DCA) is done like this:

```r
> library(vegan)
> data(dune)
> ord <- decorana(dune)
```

This saves ordination results in ord:

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

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

DCA1  DCA2  DCA3  DCA4

Eigenvalues  0.5117  0.3036  0.12125  0.14267
Decorana values  0.5360  0.2869  0.08136  0.04814
Axis lengths  3.7004  3.1166  1.30055  1.47888
```

The display of results is very brief: only eigenvalues and used options are listed. Actual ordination results are not shown, but you can see them with command summary(ord), or extract the scores with command scores. The plot function also automatically knows how to access the scores.

1.2 Non-metric multidimensional scaling

Function metaMDS is a bit special case. The actual ordination is performed by function vegan function monoMDS (or alternatively using isoMDS of the MASS package). Function metaMDS is a wrapper to perform non-metric multidimensional scaling (NMDS) like recommended in community ordination: it uses adequate dissimilarity measures (function vegdist), then it runs NMDS several times with random starting configurations, compares results (function procrustes), and stops after finding twice a similar minimum stress solution. Finally it scales and rotates the solution, and adds species scores to the configuration as weighted averages (function wascores):

```r
> ord <- metaMDS(dune)

Run 0 stress 0.1192678
Run 1 stress 0.1183186
... New best solution
... Procrustes: rmse 0.02026987 max resid 0.06495422
```
Run 2 stress 0.1183186
... New best solution
... Procrustes: rmse 4.224025e-06 max resid 1.244164e-05
... Similar to previous best
Run 3 stress 0.1808922
Run 4 stress 0.1192679
Run 5 stress 0.1183186
... New best solution
... Procrustes: rmse 5.93013e-06 max resid 1.880565e-05
... Similar to previous best
Run 6 stress 0.1183186
... Procrustes: rmse 5.934059e-05 max resid 0.0001551942
... Similar to previous best
Run 7 stress 0.1183186
... Procrustes: rmse 3.329493e-06 max resid 8.778026e-06
... Similar to previous best
Run 8 stress 0.1808912
Run 9 stress 0.1192684
Run 10 stress 0.1192682
Run 11 stress 0.1192678
Run 12 stress 0.1192678
Run 13 stress 0.119268
Run 14 stress 0.1812933
Run 15 stress 0.1808911
Run 16 stress 0.1183187
... Procrustes: rmse 0.0001324589 max resid 0.0003583602
... Similar to previous best
Run 17 stress 0.2042117
Run 18 stress 0.1183186
... Procrustes: rmse 5.580081e-06 max resid 1.877692e-05
... Similar to previous best
Run 19 stress 0.1192682
Run 20 stress 0.1808915
*** Solution reached

> ord

Call:
metaMDS(comm = dune)

global Multidimensional Scaling using monoMDS

Data:  dune
Distance:  bray

Dimensions:  2
Stress:  0.1183186
Stress type 1, weak ties
Two convergent solutions found after 20 tries
Scaling:  centring, PC rotation, halfchange scaling
Species:  expanded scores based on 'dune'
2 Ordination graphics

Ordination is nothing but a way of drawing graphs, and it is best to inspect ordinations only graphically (which also implies that they should not be taken too seriously).

All ordination results of vegan can be displayed with a plot command (Fig. 1):

```r
> plot(ord)
```

Default plot command uses either black circles for sites and red pluses for species, or black and red text for sites and species, resp. The choices depend on the number of items in the plot and ordination method. You can override the default choice by setting `type = "p"` for points, or `type = "t"` for text. For a better control of ordination graphics you can first draw an empty plot (`type = "n"`) and then add species and sites separately using `points` or `text` functions. In this way you can combine points and text, and you can select colours and character sizes freely (Fig. 2):

```r
> plot(ord, type = "n")
> points(ord, display = "sites", cex = 0.8, pch=21, col="red", bg="yellow")
> text(ord, display = "spec", cex=0.7, col="blue")
```

All vegan ordination methods have a specific plot function. In addition, vegan has an alternative plotting function ordiplot that also knows many non-vegan ordination methods, such as prcomp, cmdscale and isoMDS. All vegan plot functions return invisibly an ordiplot object, so that you can use ordiplot support functions with the results (`points, text, identify`).

Function ordirgl (requires rgl package) provides dynamic three-dimensional graphics that can be spun around or zoomed into with your mouse. Function ordiplot3d (requires package scatterplot3d) displays simple three-dimensional scatterplots.
2.1 Cluttered plots

Ordination plots are often congested: there is a large number of sites and species, and it may be impossible to display all clearly. In particular, two or more species may have identical scores and are plotted over each other. **Vegan** does not have (yet?) automatic tools for clean plotting in these cases, but here some methods you can try:

- Zoom into graph setting axis limits `xlim` and `ylim`. You must typically set both, because **vegan** will maintain equal aspect ratio of axes.
- Use points and add labels only some points with `identify` command.
- Use `select` argument in ordination `text` and `points` functions to only show the specified items.
- Use `ordilabel` function that uses opaque background to the text: some text labels will be covered, but the uppermost are readable.
- Use automatic `orditorp` function that uses text only if this can be done without overwriting previous labels, but points in other cases.
- Use automatic `ordipointlabel` function that uses both points and text labels, and tries to optimize the location of the text to avoid overwriting.
- Use interactive `orditkplot` function that draws both points and labels for ordination scores, and allows you to drag labels to better positions. You can export the results of the edited graph to encapsulated `postscript`, `pdf`, `png` or `jpeg` files, or copy directly to encapsulated `postscript`, or return the edited positions to **R** for further processing.

2.2 Adding items to ordination plots

**Vegan** has a group of functions for adding information about classification or grouping of points onto ordination diagrams. Function `ordihull` adds convex
hulls, ordihull adds ellipses enclosing all points in the group (ellipsoid hulls) or ellipses of standard deviation, standard error or confidence areas, and ordispider combines items to their centroid (Fig. 3):

```r
> data(dune.env)
> attach(dune.env)
> plot(ord, disp="sites", type="n")
> ordihull(ord, Management, col=1:4, lwd=3)
> ordiellipse(ord, Management, col=1:4, kind = "ehull", lwd=3)
> ordiellipse(ord, Management, col=1:4, draw="polygon")
> ordispider(ord, Management, col=1:4, label = TRUE)
> points(ord, disp="sites", pch=21, col="red", bg="yellow", cex=1.3)
```

In addition, you can overlay a cluster dendrogram from hclust using ordicluster or a minimum spanning tree from spantree with its lines function. Segmented arrows can be added with ordiarrows, lines with ordisegments and regular grids with ordigrid.

### 3 Fitting environmental variables

**Vegan** provides two functions for fitting environmental variables onto ordination:

- **envfit** fits vectors of continuous variables and centroids of levels of class variables (defined as factor in R). The arrow shows the direction of the (increasing) gradient, and the length of the arrow is proportional to the correlation between the variable and the ordination.

- **ordisurf** (which requires package mgcv) fits smooth surfaces for continuous variables onto ordination using thinplate splines with cross-validatory selection of smoothness.

Function envfit can be called with a formula interface, and it optionally can assess the “significance” of the variables using permutation tests:


```r
> ord.fit <- envfit(ord ~ A1 + Management, data=dune.env, perm=999)
> ord.fit

***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>A1</td>
<td>0.96473</td>
<td>0.26324</td>
<td>0.3649</td>
<td>0.014 *</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>NMDS1</th>
<th>NMDS2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>-0.4534</td>
<td>-0.0102</td>
</tr>
<tr>
<td>HF</td>
<td>-0.2636</td>
<td>-0.1282</td>
</tr>
<tr>
<td>NM</td>
<td>0.2957</td>
<td>0.5790</td>
</tr>
<tr>
<td>SF</td>
<td>0.1506</td>
<td>-0.4670</td>
</tr>
</tbody>
</table>

Goodness of fit:

<table>
<thead>
<tr>
<th></th>
<th>r²</th>
<th>Pr(&gt;r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>0.4134</td>
<td>0.009 **</td>
</tr>
</tbody>
</table>

---

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

The result can be drawn directly or added to an ordination diagram (Fig. 4):

> plot(ord, dis="site")
> plot(ord.fit)

Function `ordisurf` directly adds a fitted surface onto ordination, but it returns the result of the fitted thinplate spline `gam` (Fig. 4):

> ordisurf(ord, A1, add=TRUE)

Family: gaussian
Link function: identity

Formula:

\( y \sim s(x_1, x_2, k = 10, bs = "tp", fx = \text{FALSE}) \)

Estimated degrees of freedom:

1.59  total = 2.59

REML score: 41.58727

4 Constrained ordination

Vegan has three methods of constrained ordination: constrained or “canonical” correspondence analysis (function `cca`), redundancy analysis (function `rda`) and distance-based redundancy analysis (function `capscale`). All these functions

7
can have a conditioning term that is “partialed out”. I only demonstrate \texttt{cca}, but all functions accept similar commands and can be used in the same way.

The preferred way is to use \texttt{formula} interface, where the left hand side gives the community data frame and the right hand side lists the constraining variables:

```r
> ord <- cca(dune ~ A1 + Management, data=dune.env)
> ord
```

Call: \texttt{cca(formula = dune ~ A1 + Management, data = dune.env)}

<table>
<thead>
<tr>
<th></th>
<th>Inertia</th>
<th>Proportion</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2.1153</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Constrained</td>
<td>0.7798</td>
<td>0.3686</td>
<td>4</td>
</tr>
<tr>
<td>Unconstrained</td>
<td>1.3355</td>
<td>0.6314</td>
<td>15</td>
</tr>
</tbody>
</table>

Inertia is mean squared contingency coefficient

Eigenvalues for constrained axes:

\begin{verbatim}
CCA1  CCA2  CCA3  CCA4
0.3187 0.2372 0.1322 0.0917
\end{verbatim}

Eigenvalues for unconstrained axes:

\begin{verbatim}
CA1   CA2   CA3   CA4   CA5   CA6   CA7   CA8   CA9   CA10
0.3620 0.2029 0.1527 0.1345 0.1110 0.0800 0.0767 0.0553 0.0444 0.0415
CA11  CA12  CA13  CA14  CA15
0.0317 0.0178 0.0116 0.0087 0.0047
\end{verbatim}

The results can be plotted with (Fig. 5):

```r
> plot(ord)
```

There are three groups of items: sites, species and centroids (and biplot arrows) of environmental variables. All these can be added individually to an empty plot, and all previously explained tricks of controlling graphics still apply.

It is not recommended to perform constrained ordination with all environmental variables you happen to have: adding the number of constraints means slacker constraint, and you finally end up with solution similar to unconstrained
ordination. In that case it is better to use unconstrained ordination with environmental fitting. However, if you really want to do so, it is possible with the following shortcut in formula:

```r
> cca(dune ~ ., data=dune.env)
```

Call: `cca(formula = dune ~ A1 + Moisture + Management + Use + Manure, data = dune.env)`

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

Inertia is mean squared contingency coefficient

Some constraints were aliased because they were collinear (redundant)

Eigenvalues for constrained axes:

<table>
<thead>
<tr>
<th>CCA1</th>
<th>CCA2</th>
<th>CCA3</th>
<th>CCA4</th>
<th>CCA5</th>
<th>CCA6</th>
<th>CCA7</th>
<th>CCA8</th>
<th>CCA9</th>
<th>CCA10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4671</td>
<td>0.3410</td>
<td>0.1761</td>
<td>0.1532</td>
<td>0.0953</td>
<td>0.0703</td>
<td>0.0589</td>
<td>0.0499</td>
<td>0.0318</td>
<td>0.0260</td>
</tr>
</tbody>
</table>

CCA11  CCA12  0.0228  0.0108

Eigenvalues for unconstrained axes:

<table>
<thead>
<tr>
<th>CA1</th>
<th>CA2</th>
<th>CA3</th>
<th>CA4</th>
<th>CA5</th>
<th>CA6</th>
<th>CA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.27237</td>
<td>0.10876</td>
<td>0.08975</td>
<td>0.06305</td>
<td>0.03489</td>
<td>0.02529</td>
<td>0.01798</td>
</tr>
</tbody>
</table>

4.1 Significance tests

`vegan` provides permutation tests for the significance of constraints. The test mimics standard analysis of variance function (`anova`), and the default test analyses all constraints simultaneously:

```r
> anova(ord)
```

Permutation test for cca under reduced model
Permutation: free
Number of permutations: 999
Model: cca(formula = dune ~ A1 + Management, data = dune.env)  
  Df ChiSquare  F Pr(>F)  
Model 4 0.77978 2.1896 0.001 ***  
Residual 15 1.33549  
---  
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
The function actually used was anova.cca, but you do not need to give its name in full, because R automatically chooses the correct anova variant for the result of constrained ordination.

It is also possible to analyse terms separately:

```r
> anova(ord, by="term", permutations=199)
Permutation test for cca under reduced model  
Terms added sequentially (first to last)  
Permutation: free  
Number of permutations: 199
```

Model: cca(formula = dune ~ A1 + Management, data = dune.env)  
  Df ChiSquare  F Pr(>F)  
A1 1 0.22476 2.5245 0.005 **  
Management 3 0.55502 2.0780 0.005 **  
Residual 15 1.33549  
---  
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
This test is sequential: the terms are analysed in the order they happen to be in the model. You can also analyse significances of marginal effects (“Type III effects”):

```r
> anova(ord, by="mar", permutations=199)
Permutation test for cca under reduced model  
Marginal effects of terms  
Permutation: free  
Number of permutations: 199
```

Model: cca(formula = dune ~ A1 + Management, data = dune.env)  
  Df ChiSquare  F Pr(>F)  
A1 1 0.17594 1.9761 0.050 *  
Management 3 0.55502 2.0780 0.005 **  
Residual 15 1.33549  
---  
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1  
Moreover, it is possible to analyse significance of each axis:

```r
> anova(ord, by="axis", permutations=499)
Permutation test for cca under reduced model  
Marginal tests for axes  
Permutation: free  
Number of permutations: 499
```

Model: cca(formula = dune ~ A1 + Management, data = dune.env)  
  Df ChiSquare  F Pr(>F)  
---  

4.2 Conditioned or partial ordination

All constrained ordination methods can have terms that are partialled out from the analysis before constraints:

```r
> ord <- cca(dune ~ A1 + Management + Condition(Moisture), data=dune.env)
> ord
```

```
Call: cca(formula = dune ~ A1 + Management + Condition(Moisture), data = dune.env)

Inertia Proportion Rank
Total 2.1153 1.0000
Conditional 0.6283 0.2970 3
Constrained 0.5109 0.2415 4
Unconstrained 0.9761 0.4615 12
Inertia is mean squared contingency coefficient

Eigenvalues for constrained axes:
CCA1 CCA2 CCA3 CCA4
0.24932 0.12090 0.08160 0.05904

Eigenvalues for unconstrained axes:
CA1 CA2 CA3 CA4 CA5 CA6 CA7 CA8 CA9
0.30637 0.13191 0.11516 0.10947 0.07724 0.07875 0.04871 0.03758 0.03106
CA10 CA11 CA12
0.02102 0.01254 0.00928
```

This partials out the effect of Moisture before analysing the effects of A1 and Management. This also influences the significances of the terms:

```r
> anova(ord, by="term", permutations=499)
```

```
Permutation test for cca under reduced model
Terms added sequentially (first to last)
Permutation: free
Number of permutations: 499

Model: cca(formula = dune ~ A1 + Management + Condition(Moisture), data = dune.env)

                  Df  ChiSquare   F Pr(>F)
A1                   1 0.11543 1.4190 0.118
Management           3 0.39543 1.6205 0.008 **
Residual             12 0.97610

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

If we had a designed experiment, we may wish to restrict the permutations so that the observations only are permuted within levels of Moisture. Restricted
permutation is based on the powerful `permute` package. Function `how()` can be used to define permutation schemes. In the following, we set the levels with `plots` argument:

```r
> how <- how(nperm=499, plots = Plots(strata=dune.env$Moisture))
> anova(ord, by="term", permutations = how)
```

Permutation test for cca under reduced model
Terms added sequentially (first to last)
Plots: dune.env$Moisture, plot permutation: none
Permutation: free
Number of permutations: 499

Model: cca(formula = dune ~ A1 + Management + Condition(Moisture), data = dune.env)

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>ChiSquare</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>0.11543</td>
<td>1.4190</td>
<td>0.286</td>
</tr>
<tr>
<td>Management</td>
<td>3</td>
<td>0.39543</td>
<td>1.6205</td>
<td>0.002 **</td>
</tr>
<tr>
<td>Residual</td>
<td>12</td>
<td>0.97610</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1