Table of Contents

Section: Ordination analysis

Supplementary variables (unconstrained ordination)

Theory R functions **Examples** Exercise 🔒

Example 1: tb-PCA on Vltava river valley data with passively projected environmental variables

Let's use data from river valley again and calculate tb-PCA on them, i.e. PCA on Hellinger standardised raw species composition data. Note that because the original data represent percentage cover, we will log-transform them first, before applying the Hellinger standardisation:

```
vltava.spe <- read.delim ('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/vltava-spe.txt', row.names = 1)
vltava.env <- read.delim ('https://raw.githubusercontent.com/zdealveindy/anadat-r/master/data/vltava-env.txt')

library (vegan)
PCA <- rda (X = decostand (log1p (vltava.spe), method = 'hellinger'))
```

The function `envfit` calculates multiple regression of environmental variable with ordination axes (environmental variable is used as dependent and selected ordination axes as explanatory variables). Significance is tested by permutation test. Vectors (for continual variables) and centroids (for categorical variables) can be projected onto ordination diagram using `plot` function.

```
ef <- envfit (PCA ~ ASPSSW + SOILDPT + pH, data = vltava.env, perm = 999)
ef
```

```
***VECTORS

             PC1      PC2     r2 Pr(>r)
ASPSSW  -0.87822  0.47826 0.4769  0.001 ***
SOILDPT  0.94893  0.31549 0.2891  0.001 ***
pH       0.38093  0.92460 0.3885  0.001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999
```

The variable which is the most strongly related to the first two ordination axes (judged by the value of r^2 in the `envfit` table) is the folded aspect (ASPSSW), followed by soil pH and soil depth. Soil pH is strongly related to the second ordination axis, while soil depth and folded aspect to the first one (according to the coefficients in PC1 and PC2 columns).

The result table contains the following columns:

- PC1, PC2 - the relationship of the environmental variable with the first and second PCA axis. These values are not correlation coefficients, but coordinates of the vector head on given ordination axes assuming that the vector is of a length = 1 unit[1].
- r2 - variation explained by the model of multiple regression; the square-root of this value is used to scale lengths of vectors (arrows) in the ordination diagrams (variables with higher sqrt (r^2) are represented by longer arrows).
- Pr(>r) - the significance of the multiple regression, calculated by permutation test with specified number of permutations. Indicates whether the variable is related to ordination axes more than if it is a randomly generated one.

The function `envfit` executed three permutation tests, all three with highly significant results. When the number of tested variables increases, correction for multiple testing is desirable. In the case above, we may extract the significance values from `ef` object and apply Bonferroni correction using the function `p.adjust`:

```
ef.adj <- ef
pvals.adj <- p.adjust (ef$vectors$pvals, method = 'bonferroni')
ef.adj$vectors$pvals <- pvals.adj
ef.adj
```

In the code above, I first copied the original `ef` object created by function `envfit` into new, `ef.adj` object. Then I extracted the significance values from `ef` (`ef$vectors$pvals`) and applied function `p.adjust` with Bonferroni's method of adjustment. By these adjusted values I replaced the original P-values in the copy of original results `ef.adj`. Newly printed results show adjusted P-values:

```
***VECTORS

             PC1      PC2     r2 Pr(>r)
ASPSSW  -0.87822  0.47826 0.4769  0.003 **
SOILDPT  0.94893  0.31549 0.2891  0.003 **
pH       0.38093  0.92460 0.3885  0.003 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 999
```

Alternatively, you can use custom-build function `p.adjust.envfit` (definition here), giving exactly the same results as above. First, source the definition of this custom made function from this website:

```
source
('http://www.davidzeleny.net/anadat-r/doku.php/en:customized_functions:p.adjust.envfit?do=export_code&codeblock=0')
```

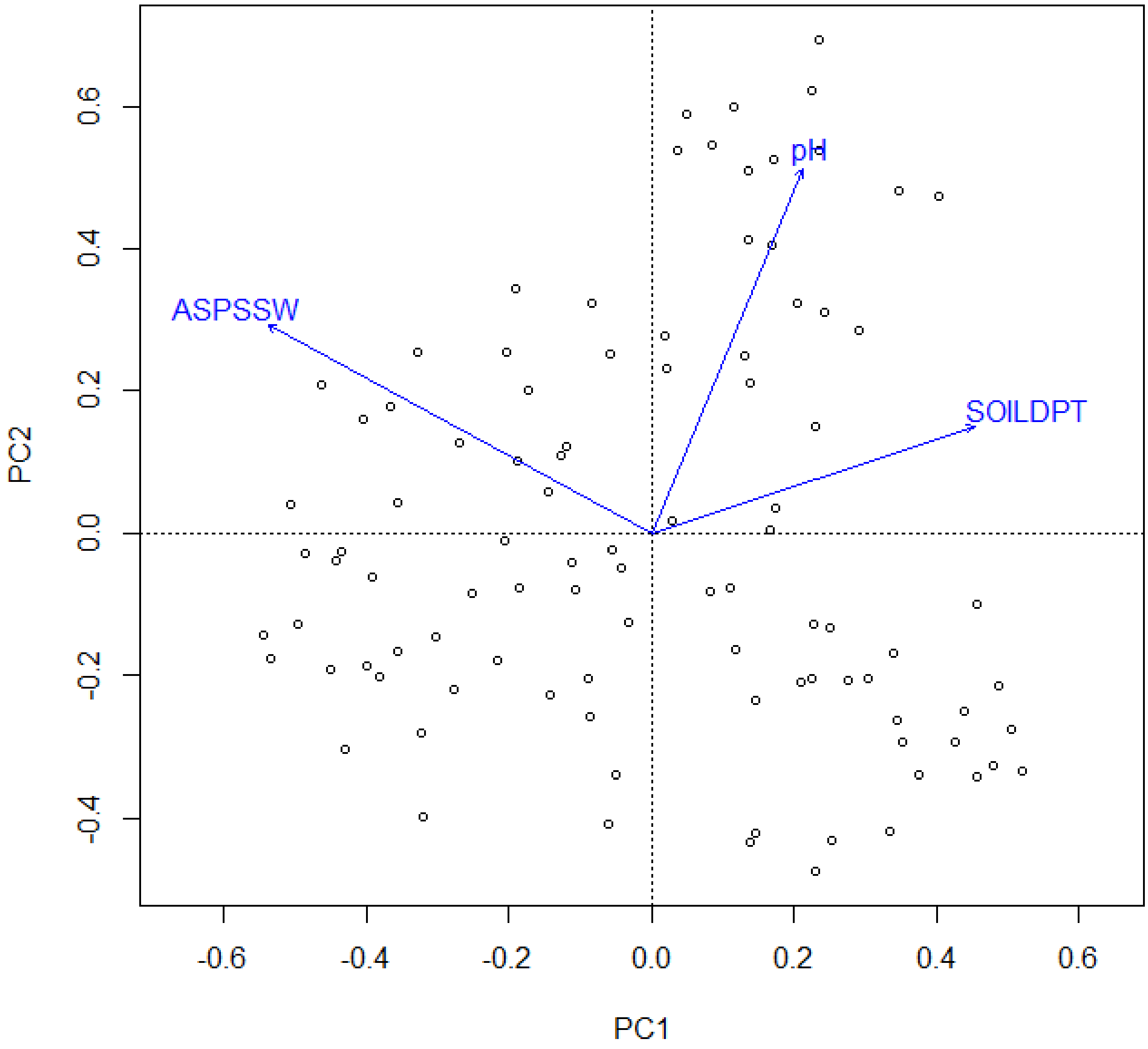
... and just use it (the results should be the adjusted results above):

```
ef.adj <- p.adjust.envfit (ef)
ef.adj
```

To draw the variables onto the ordination diagram, use the function `plot` on the object returned by the function `envfit` (`ef` and `ef.adj` above, both being of the class `envfit`). `plot.envfit` (the

.envfit does not need to be typed) is a low-level graphical function, which adds the vectors (for quantitative variables) or centroids (for qualitative variables) of the environmental variables into already existing ordination diagram:

```
ordiplot (PCA, display = 'sites')
plot (ef)
```



In the case of PCA, the coordinates of the arrow heads on the ordination axes represent the values of Pearson's correlation coefficient between the variable and given ordination axis (in the case of other unconstrained ordination - CA, DCA or NMDS - the situation is more complicated is not that straightforward, since the regression is weighted). To confirm this, let's calculate the correlation of each environmental variable with each ordination axis.

To get scores of samples (or species) on ordination axes, we need function scores (applicable on object created by vegan ordination functions:

```
scores.pca <- scores (PCA, display = 'sites', choices = 1:2)
scores.pca
```

```
            PC1         PC2
1  -0.36770382  0.178399044
2  -0.46395458  0.209974936
3  -0.26899925  0.127517062
4  -0.32793000  0.254319805
5   0.03691218  0.536838689
6   0.22455684  0.622463000
...
```

(argument `display` specifies that we want scores of `sites`, not `species`, and argument `choices` points to ordination axes we want to extract - the first two).

Correlation between each environmental variable and each axis scores can be calculated as follows:

```
cor (vltava.env [, c('ASPSSW', 'SOILDPT', 'pH')], scores.pca)
```

```
               PC1       PC2
ASPSSW  -0.6065085 0.3302958
SOILDPT  0.5101798 0.1696164
pH       0.2374247 0.5762849
```

We get exactly the same values if we take results of the function `envfit` and multiply each value in the column PC1 and PC2[2] by it's appropriate sqrt (r2):

```
arrow_heads <- ef$vectors$arrows  # extracts matrix of coordinates of arrow
heads from ef
r2 <- ef$vectors$r                # extracts vector of r2 for each env.
variable
arrow_heads * sqrt (r2)
```

```
               PC1       PC2
ASPSSW  -0.6065085 0.3302958
SOILDPT  0.5101798 0.1696164
pH       0.2374247 0.5762849
attr(,"decostand")
[1] "normalize"
```

Note that these values are identical to Pearson's correlation coefficients in the table above. But this is true only for PCA or tb-PCA, in which ordination scores are standardized to unit variance. In case of other unconstrained ordinations (CA, DCA), standard deviations of individual axes differ, and regression coefficients in multiple regression cannot be directly converted into Pearson's correlation coefficients. Additionally, unimodal methods (DCA, CA) are using weighted multiple regression, with weights of samples reflecting their importance in analysis[3].

Example 2: Projecting environmental variable onto ordination as nonlinear

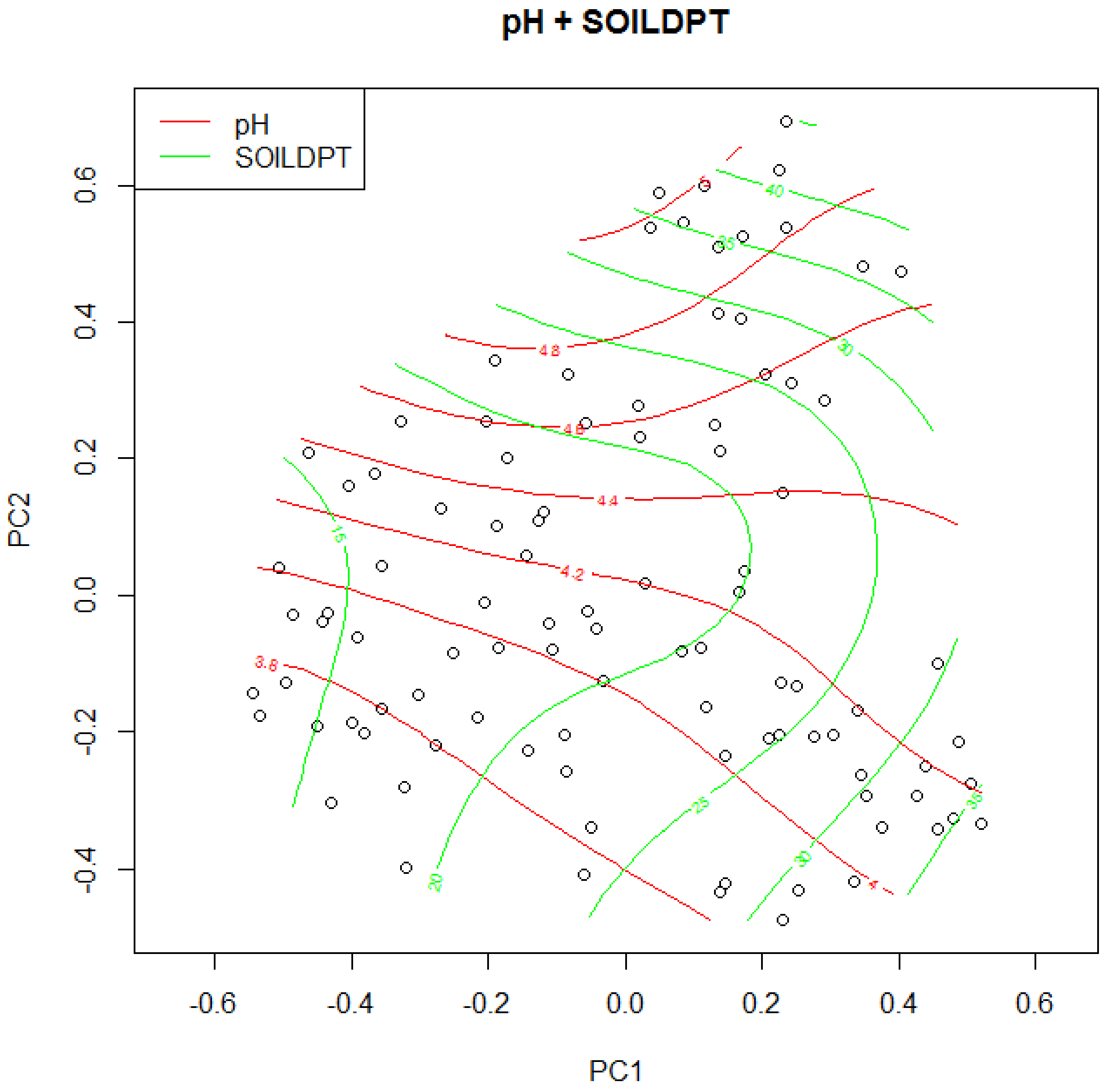
surface

This can be done using function `ordisurf` from `vegan`. It draws the surface of fitted environmental variable into ordination diagram using GAM models. This is fancy option in case that you don't expect linear relationship between ordination axis and environmental variable, or you don't want to relate the environmental variables directly to ordination axes (the latter may seem as relevant argument in case of NMDS, which is basically a method free of ordination axes[4)]).

To add e.g. pH into PCA ordination diagram using data from Vltava as in the example above, use the function `ordisurf`. This function is both high- and low-level graphical function, meaning that it can either created the whole new ordination diagram and add the surface in it, or it can only add the surface into already existing ordination diagram (use argument `add = TRUE`):

```
ordisurf (PCA, vltava.env[, 'pH'], main = 'pH + SOILDPT')
ordisurf (PCA, vltava.env[, 'SOILDPT'], add = T, col = 'green')
legend ('topleft', col = c('red', 'green'), lty = 1, legend = c('pH',
'SOILDPT'))
```

In this script, the first `ordisurf` draws new ordination diagram and adds pH values in it (including the title using argument `main`). The second application of `ordisurf` adds the surface of soil depth variable, using different color (argument `col`). We may also add the `legend` to ease the interpretation of the surface colors.



[1] The help function to `envfit` calls this '*direction cosines which are the coordinates of the heads of unit length vectors*', see `?envfit`.

[2] To see the structure of the object returned by function `envfit`, use the function `str` - e.g. `str (ef)` in this case. The object is a list of lists, containing all relevant information. The printed version which is printed into command line when we type `ef` is the product of the function `print.envfit`, which is automatically called when the `envfit` object is evaluated.

[3] Weights are basically sums of species presences-absences/abundances in samples, meaning that samples with more species (in case of presence-absence data) or higher overall species abundances (in case of abundance data) have higher weight in the regression.

[4] There is persistent discussion whether it is logical/possible/allowed to project supplementary environmental variables onto NMDS ordination diagram. See e.g. opinion of Jari Oksanen on this topic in vegan's FAQ list, which clarifies some misunderstanding related to NMDS non-metricity.

From:
https://anadat-r.davidzeleny.net/ - **Analysis of community ecology data in R**

Permanent link:
https://anadat-r.davidzeleny.net/doku.php/en:suppl_vars_examples

Last update: **2019/02/26 23:35**