RDA, tb-RDA, CCA & db-RDA (constrained ordination)

Redundancy analysis (RDA) and transformation-based redundancy analysis (tb-RDA)

Linear constrained ordination methods implicitly based on Euclidean (RDA) or Hellinger/chord/other (tb-RDA) distances. The calculation (detailed below) can be simply described as a set of (multiple) linear regression analyses, where species abundances (for each species in the species composition matrix separately) are regressed against (one or several) environmental variable(s). The result is that variation in species composition is decomposed into variation related to environmental variables (represented by constrained/canonical axes) and not related to environmental variables (unconstrained axes). The number of constrained axes is equal or lower than the number of quantitative explanatory variables; in the case of a qualitative/categorical variable, the number of constrained axes is equal to the number of categories in that variable minus one. Each canonical axis is a linear combination of *all* explanatory variables.

The algorithm of RDA can be summarised as follows (Fig. 1 and Fig. 2). The matrix of species composition (sample x species) and the matrix of environmental variables (sample x env.variables, for simplicity containing only one env. variable in the illustration below) needs to be available.

- Abundances of the first species (*spe1*) are regressed against environmental variable (*env1*) by linear regression (or by multiple regression if more env. variables are available), with *spe1* as the dependent variable and *env1* (and other env. variables if available) as explanatory.
- The values of species abundances fitted by the regression model (i.e. located on the regression line) are stored in the matrix of fitted values, while residuals of species abundances (the difference between observed abundances and fitted abundances) are stored in the matrix of residuals.
- The same is repeated for all species in a matrix of species composition. Resulting matrices of predicted values and residual values have the same size (no. of samples x no. of species) as the original matrix of species composition.
- The matrix of predicted values is used in PCA to extract constrained ordination axes, while the matrix of residual values is used in PCA to extract unconstrained axes.
- In the example on Fig. 2 with only one explanatory variable there is only one constrained ordination axis (the second, vertical one in the ordination diagram is the first unconstrained axis).

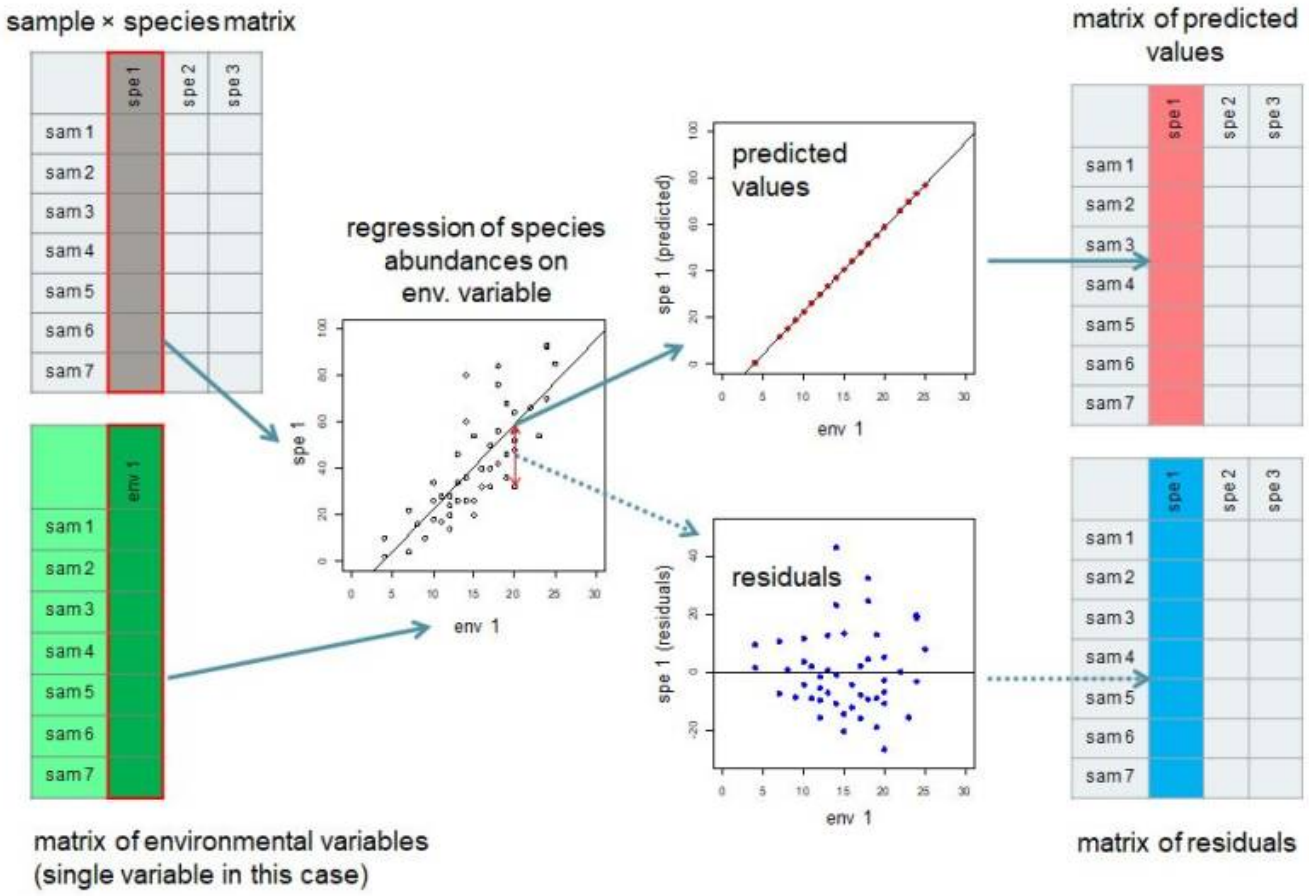


Figure 1: Schema of RDA algorithm.

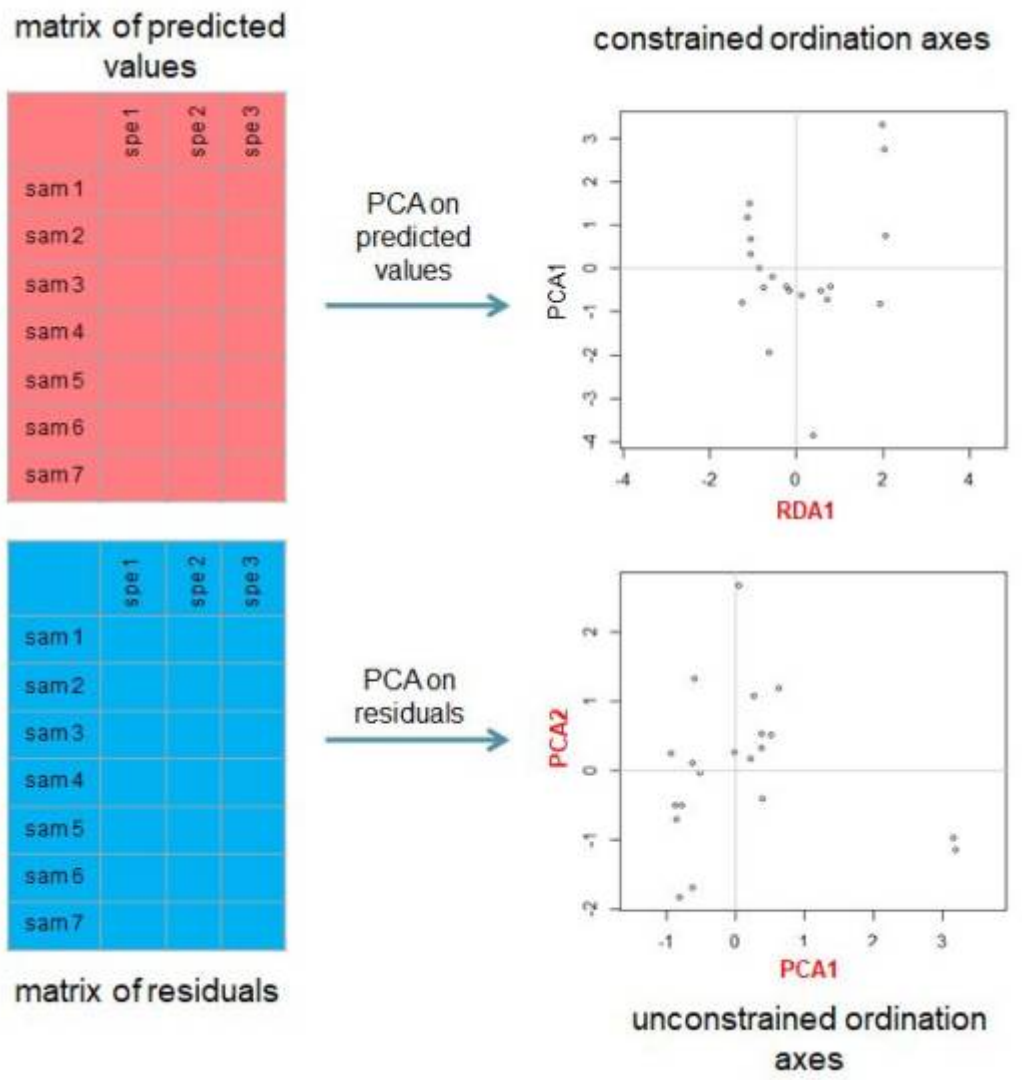


Figure 2: Schema of RDA algorithm - continuation.

While in the case of unconstrained ordination the information we are interested is mostly about the configuration of samples and species in the ordination diagram, the relative importance of individual ordination axes (measured by their eigenvalues) and ecological interpretation of ordination axes, in the case of constrained ordinations we are more interested in the effect of environmental variables on species composition, namely in the amount of variation these variables explain and whether this variation is significant or not (see Explained variation and Monte Carlo permutation test), which of the available environmental variables are important to explain the variation of studies community (Forward selection) and how to partition the variation explained by different variables or different sets of variables (Variation partitioning).

Canonical correspondence analysis (CCA)

Unimodal constrained ordination method, related to correspondence analysis (CA), with an algorithm derived from redundancy analysis (RDA). The algorithm of RDA is modified in the way that instead of raw species composition data, the set of regressions is done on the Q matrix, and the weighted multiple regression is used instead of simple multiple regression, where weights are row sums, i.e. the sums of species abundances in individual samples. The requirement for input data is the same as for correspondence analysis - the data must be non-negative integers or presences-absences.

Note that CCA calculates **two sets of sample scores**: LC scores, and WA scores. **LC scores** are linear combinations of the columns in the environmental matrix, while **WA scores** are weighted averages of the species scores. Default plotting of ordination diagrams differ between programs; e.g. in R (library *vegan*), the

samples in CCA ordination plots are using WA scores, while in CANOCO 5 they are plotted using LC scores. Use of each scoring method has its proponents and opponents. The difference when plotted onto the ordination diagram is rather obvious when explanatory (environmental) variables are factors with several levels, or quantitative variables with evenly spaced values (Fig. 3). Remember to report which scores you have chosen to display, whether LC or WA.

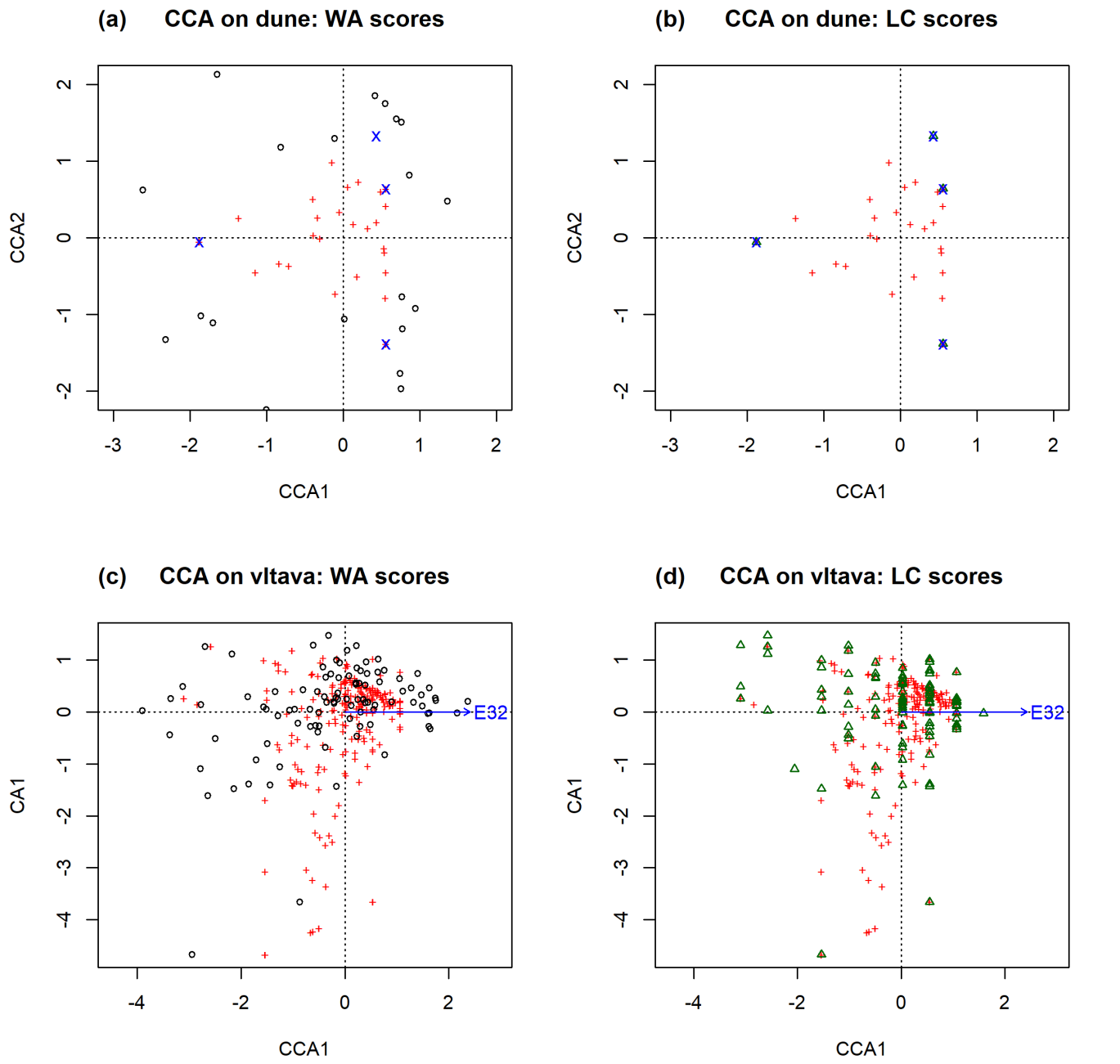


Figure 3: Difference between WA and LC scores in CCA. The upper row of figures is CCA calculated using dune data with a factor (Management with four levels) as an explanatory variable. The lower row of figures is CCA calculated using vltava data with quantitative variable (cover of tree and shrub layer, E32) as explanatory; the values in E32 were rounded into nearest tens (ie it contains values like 20, 30, 40, ...). Diagrams in the left column are using WA scores, those in the right column are using LC scores. In Figure (b), the sample scores are all hidden behind the centroids of the management factor. Note that species scores (red plus symbols) are not influenced by the choice of sample scores.

Distance based RDA (db-RDA)

This is RDA applied to the matrix of sample scores calculated by principal coordinate analysis (PCoA). The raw species data are first converted into a dissimilarity matrix using a selected dissimilarity metric, and this matrix is submitted to PCoA. The matrix of site scores on all PCoA ordination axes is then used in RDA instead of the raw species data together with explanatory variables. The benefit of db-RDA is that any distance metric can be applied on the data (i.e. not only Euclidean as in RDA, Hellinger (or few others) as in tb-RDA, or chi-square as in CCA). Care must be applied to avoid negative eigenvalues in PCoA, which would be omitted from the analyses; the solution is to either use only metric (Euclidean) dissimilarities, or to apply transformation which will turn non-metric dissimilarity into metric one (e.g. square root transformation applied on Bray-Curtis dissimilarity), or to use some of the available corrections. Although species information is lost during the calculation of the dissimilarity matrix, if the original matrix of species composition matrix is available, the species scores can be added into the final ordination diagram as weighted means of site score in which they occur or as vectors fitted onto the ordination space.

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

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

Last update: **2021/04/10 15:58**