Table of Contents

Ordination analysis

PCoA & NMDS (distance-based unconstrained ordination)

Theory R functions Examples Exercise 🔒

Principal Correspondence Analysis (PCoA)

This method is also known as MDS (Metric Multidimensional Scaling). While PCA preserves Euclidean distances among samples and CA chi-square distances, PCoA provides Euclidean representation of a set of objects whose relationship is measured by any dissimilarity index. As well as PCA and CA, PCoA returns a set of orthogonal axes whose importance is measured by eigenvalues. This means that calculating PCoA on Euclidean distances among samples yields the same results as PCA calculated on the covariance matrix of the same dataset (if scaling 1 is used), and PCoA on Chi-square distances similar results to CA (but not identical, because CA is applying the weights in the calculation). In case of using non-metric (non-Euclidean) distance index, the PCoA may produce axes with negative eigenvalues which cannot be plotted. Solution to this is to either convert the non-metric dissimilarity index to metric one (e.g. Bray-Curtis dissimilarity is non-metric, but after square-root transformation becomes metric) or using specific corrections (Lingoes or Cailliez). Since the PCoA algorithm is based on the matrix of dissimilarities between samples, the species scores are not calculated; however, the species can be projected to the ordination diagram by weighted averaging or correlations, similarly as supplementary environmental variables.

Non-metric Multidimensional Scaling (NMDS)

Non-metric Multidimensional Scaling is a non-metric alternative of PCoA analysis. It can use any dissimilarity measure among samples, and the main aim is to locate samples in low dimensional ordination space (two or three axes) so as the Euclidean distances between these samples correspond to the dissimilarities represented by the original dissimilarity index. The method is non-metric, because it does not use the raw dissimilarity values, but converts them into the ranks and use these ranks in the calculation. The algorithm is iterative - it starts from the initial distribution of samples in the ordination space, and by the iterative reshuffling of samples it searches for optimal final distribution. Due to the iterative nature of the algorithm, each run may result in a different solution.

The algorithm goes like this (simplified):

1. Specify the number of dimensions m you want to use (into which you want to *scale* down the distribution of samples in multidimensional space - that's why it's scaling).
2. Construct initial configuration of all samples in m dimensions as a starting point of the iterative process. The result of the whole iteration procedure may depend on this step, so it's somehow crucial - the initial configuration can be generated randomly, but a better way is to help it a bit, e.g. by using as starting positions results of PCoA ordination on the same dissimilarity matrix.
3. An iterative procedure tries to reshuffle the objects in a given number of dimension in such a way that the real (Euclidean) distances among samples in the ordination spaces reflect best their compositional dissimilarity measured by used dissimilarity index. The fit between these two parameters is expressed by so-called *stress value* - the lower stress value the better.

4. The algorithm stops when the new iteration cannot lower the stress value - the solution has been reached.
5. After the algorithm is finished, the final solution is rotated using PCA to ease its interpretation (that's why the final ordination diagram has ordination axes, even if the original algorithm doesn't produce any).

Similarly to PCoA, NMDS solution does not have species scores, which need to be added to the final configuration of samples using weighted averaging.

Considering the algorithm, NMDS and PCoA have close to nothing in common. NMDS is an iterative method which may return different solution on re-analysis of the same data, while PCoA has a unique analytical solution. The number of ordination axes (dimensions) in NMDS can be fixed by the user, while in PCoA the number of axes is given by the dataset properties (number of samples). If the initial configuration of samples in NMDS algorithm is produced by PCoA on the same matrix, then the iterative NMDS algorithm may be seen as a method how to further optimize the sample distribution so as more variation in species composition is represented by fewer ordination axes.
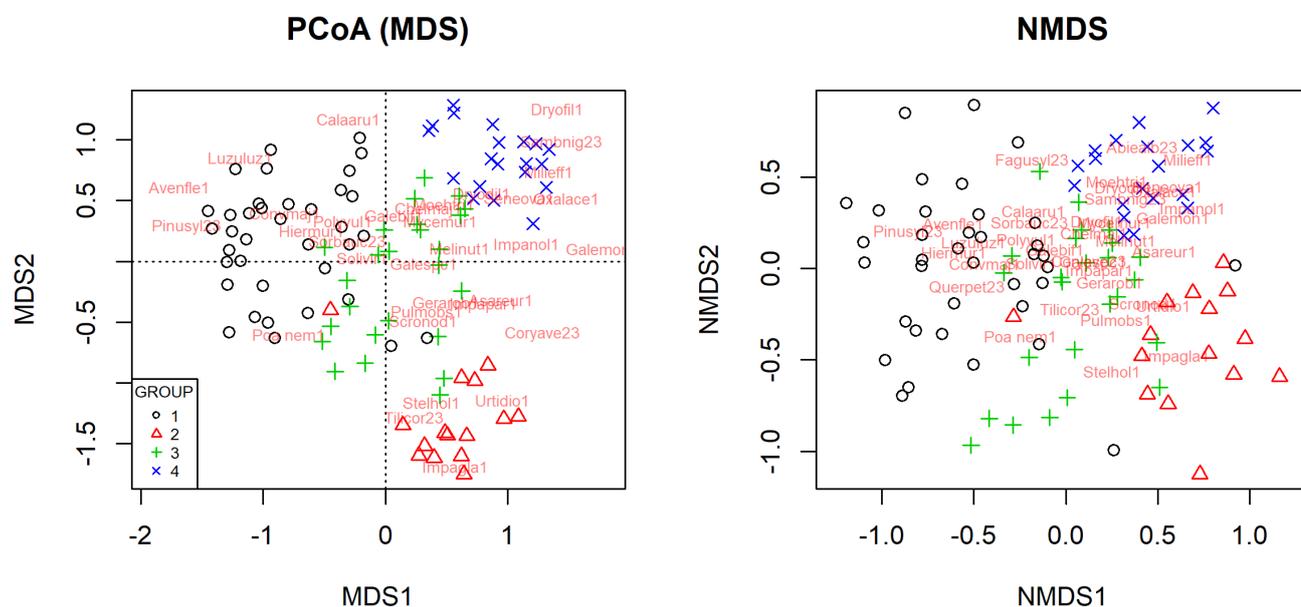

Figure 1: Ordination diagrams of PCoA (left) and NMDS (right) calculated on Bray-Curtis dissimilarity index (square-rooted to made metric) using data from Vltava river valley dataset. The classification of samples into one of the four vegetation groups (GROUP 1-4) is displayed by different colour and symbol of individual site scores. Species are added to the ordination diagrams as weighted averages of their abundances in the sites; only species occurring in at least 20 sites are displayed.