A new method for indicator species analysis in the framework of multivariate analysis of variance

Carlo Ricotta, Sandrine Pavoine, Bruno E L Cerabolini, Valério D Pillar

DR. CARLO RICOTTA (Orcid ID : 0000-0003-0818-3959)

DR. VALERIO DEPATTA PILLAR (Orcid ID : 0000-0001-6408-2891)

A new method for indicator species analysis
in the framework of multivariate analysis of variance

Carlo Ricotta[a,*], Sandrine Pavoine[b], Bruno E.L. Cerabolini[c], Valério D. Pillar[d]

[a]Department of Environmental Biology, University of Rome, 'La Sapienza', Piazzale Aldo Moro 5, 00185 Rome, Italy. [b]Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, 43 Rue Cuvier, CP 135 75005 Paris, France. [c]Department of Biotechnology and Life Sciences, University of Insubria, Via J.H. Dunant 3, 21100 Varese, Italy. [d]Departamento de Ecologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 91540-000, Brazil. *Author for correspondence. E-mail address: carlo.ricotta@uniroma1.it (C. Ricotta).

Running Head: Relating indicator species to distance-based ANOVA

Abstract

Question: In vegetation science, the compositional dissimilarity among two or more groups of plots is usually tested with dissimilarity-based multivariate analysis of variance (db-MANOVA), whereas the compositional characterization of the different groups is performed by means of indicator species analysis. Although db-MANOVA and indicator species analysis are apparently very far from each other, the question we address here is: can we put both approaches under the same methodological umbrella?

Methods: We will show that for a specific class of dissimilarity measures, the partitioning of variation used in one-factor db-MANOVA can be additively decomposed into species-level values allowing us to identify the species that contribute most to the compositional differences among the

groups. The proposed method, for which we provide a simple R function, is illustrated with one small data set on Alpine vegetation sampled along a successional gradient.

Conclusion: The species that contribute most to the compositional differences among the groups are preferentially concentrated in particular group of plots. Therefore, they can be appropriately called indicator species. This connects multivariate analysis of variance with indicator species analysis.

Introduction

For a set of plots that are grouped according to some external criteria, such as selected environmental variables or different experimental treatments, two relevant questions usually asked by community ecologists are: 'Are these groups compositionally different from each other?' and 'Which species contribute most to the compositional differences among the groups?' These questions are generally answered with different tools: the compositional dissimilarity among groups of plots is usually tested with dissimilarity-based multivariate analysis of variance (Pillar & Orlóci 1996; Anderson 2001), whereas the compositional characterization of the different groups is performed by means of indicator species analysis (Dufrêne & Legendre 1997; Chytrý et al. 2002).

Dissimilarity-based multivariate analysis of variance (or shorter db-MANOVA) is applicable to any type of compositional data, irrespective of the number of species sampled and the way they are sampled (i.e. presence/absence scores, number of individuals, species cover, etc.) provided that a meaningful measure is used to adequately represent the dissimilarity between pairs of plots (or relevés, communities, assemblages, quadrats, sites, etc.).

Given a community composition matrix containing the presence/absence or the abundance values x_{sj} of S species $\left(s=1,2,...,S\right)$ in N plots $\left(j=1,2,...,N\right)$, let the number of plots in group k $\left(k=1,2,...,K\right)$ be N_k such that $\sum_{k=1}^{K} N_k = N$ and d_{ij} be the compositional dissimilarity between plot i and plot j. In the simplest situation of a single-factor analysis on a set of plots partitioned into K groups, the very essence of db-MANOVA is to compare the plot-to-plot dissimilarity within groups with the plot-to-plot dissimilarity among groups with an adequate test statistic

(Figure 1). The larger the value of between-group dissimilarity compared to within-group dissimilarity, the more likely it is that the groups are compositionally different from each other (Anderson 2001). For db-MANOVA, hypothesis testing is usually performed with randomization tests where the N plots are randomly permuted among the K groups, thus creating a reference distribution under the null hypothesis of no compositional difference among the different groups of plots.

Once the compositional differences among the groups of plots have been verified, the identification of diagnostic species for the different groups is a relevant step for their ecological characterization (Dufrêne & Legendre 1997). In vegetation science, the concept of indicator or diagnostic species which we use here as synonyms has been usually associated with the species concentration in particular groups of plots. According to Chytrý et al. (2002), diagnostic species "include species which preferably occur in a single vegetation unit (character species) or in a few vegetation units (differential species)".

Several methods have been developed to identify indicator species, (Dufrêne & Legendre 1997; Chytrý et al. 2002; Podani & Csányi 2010; Ricotta et al. 2015). Indicator species are fundamentally defined as those species that are more common in a given group of plots than expected by chance alone. Therefore, to determine if a given species is significantly associated with a target group of plots, most of these methods compare the actual abundance of that species in the target group of plots with a reference distribution obtained from a random null model in which the species occurrences or abundances are permuted among the N plots. The null hypothesis is that all plots have equal probability to host each species, irrespective of the species environmental preferences (De Cáceres & Legendre 2009).

Although dissimilarity-based MANOVA and indicator species analysis are apparently unrelated, in this paper we will show that for a particular group of dissimilarity measures the classical partitioning of variation used for hypothesis testing in one-factor db-MANOVA can be additively decomposed into species-level values, thus connecting indicator species analysis with multivariate analysis of variance.

Materials and methods

Several test statistics have been proposed for dissimilarity-based MANOVA. For review, see Pillar (2013) and references therein. For dissimilarity matrices with Euclidean properties (see Gower & Legendre 1986), Pillar & Orlóci (1996) proposed as test criterion the sum of squared

dissimilarities between groups. Given a symmetric $N \times N$ dissimilarity matrix whose elements d_{ij} represent the pairwise dissimilarities between plot i and j such that $d_{ij} = d_{ji}$ and $d_{ii} = 0$, the between-group sum of squared dissimilarities is:

$$Q_B = Q_T - Q_W \tag{1}$$

where

$$Q_T = \frac{1}{N} \sum_{i>j}^{N} d_{ij}^2 \tag{2}$$

is the total sum of squared dissimilarities. Hence, to calculate Q_T we have to add the $N(N-1)/2$ plot-to-plot squared dissimilarities d_{ij}^2 in the sub-diagonal half of the dissimilarity matrix and then divide by the total number of plots (N). The term Q_W is the sum of squares within the K groups, which is obtained as:

$$Q_W = \sum_{k=1}^{K} Q_{Wk} \tag{3}$$

where the sum of squares within group k is:

$$Q_{Wk} = \frac{1}{N_k} \sum_{i>j}^{N} d_{ij}^2 \varepsilon_{ijk} \tag{4}$$

The indicator variable ε_{ijk} takes the value one if plots i and j are both in group k. Otherwise it is zero. That is, Q_{Wk} is obtained by adding all squared dissimilarities d_{ij}^2 between all plots that occur in group k and then dividing the sum by the number of plots in that group N_k.

The sum of squares could be equally computed from deviations of single plots from group centroids, as described, among others, by Edgington (1987) or Manly (1991) but using plot-to-plot dissimilarities does not require the computation of group centroids. McArdle & Anderson (2001)

further showed that this partition of the total sum of squares is also valid for dissimilarity coefficients that do not have Euclidean properties, thus making dissimilarity-based MANOVA a very flexible component of the ecologist toolbox.

Since the distribution of Q_B under the null hypothesis of no compositional difference among the different groups of plots is generally unknown, a reference distribution of values of the test-statistic is then obtained by Mantel randomization of the dissimilarity matrix. For dissimilarity-based one-way MANOVA, this is equivalent to randomly permuting the N plots among the K groups. Note that in one-factor MANOVA, since the total sum of squares Q_T is invariant over permutations, some test statistics, such as Q_B, Q_W and the Anderson (2001) pseudo-F ratio Q_B/Q_W are equivalent in randomization testing. That is, for the same data and permutations they will lead to identical p-values (Pillar 2013).

A desirable property for a dissimilarity index is its ability to be decomposed into species-level values (Ricotta & Podani 2017). In this way, it is possible to see which species contribute most to plot-to-plot dissimilarity. In this framework, the advantage of adopting Q_W as a test statistic over Q_B or Q_B/Q_W is that it can be decomposed into intuitively simple species-level values, provided that the squared dissimilarities d_{ij}^2 are also decomposable in the same way.

Take for example the Euclidean distance $\sqrt{\sum_{s=1}^{S}\left(x_{si}-x_{sj}\right)^2}$. It is easily shown that the squared Euclidean distance $\sum_{s=1}^{S}\left(x_{si}-x_{sj}\right)^2$ is additively decomposable into the species-level contributions $d_{ijs}^2=\left(x_{si}-x_{sj}\right)^2$ such that $d_{ij}^2=\sum_{s=1}^{S}d_{ijs}^2$. Therefore, in a sum of squares based test, if the plot-to-plot compositional dissimilarity is calculated with the Euclidean distance, the within-group sum of squares becomes:

$$Q_W = \sum_{s=1}^{S}Q_{Ws} = \sum_{s=1}^{S}\sum_{k=1}^{K}\frac{1}{N_k}\sum_{i>j}^{N}\left(x_{si}-x_{sj}\right)^2\varepsilon_{ijk} \tag{5}$$

where the contribution of single species to overall Q_W is:

$$Q_{Ws} = \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i>j}^{N} \left(x_{si} - x_{sj} \right)^2 \varepsilon_{ijk} \qquad (6)$$

Geometrically, a smaller value of Q_W denotes higher within-group concentration and larger between-group differences (Cai 2006). To test for significant compositional differences among the groups, the actual value of Q_W can be compared with a reference distribution of null values obtained by Mantel randomization of the dissimilarity matrix. P-values can be computed as the proportion of null values that are as low or lower than the actual value of the test statistic. The quantities Q_{Ws} can then be used as in classical post-hoc analysis: if the dissimilarity-based MANOVA shows significant overall compositional difference among groups, the analysis can proceed by exploring differences in single-species occurrences (or abundances) among the groups. This can be done by using the same permutation procedures used for testing the significance of Q_W.

The contribution of a particular species to the compositional differences among the *K* groups can be summarized as standardized effect size (*SES*; Gotelli & McCabe 2002):

$$SES = \frac{Obs(I) - Mean(I)}{SD(I)} \qquad (7)$$

where $Obs(I)$ is the observed value of a given variable *I*, $Mean(I)$ is the mean of the null distribution of the variable in random assemblages, and $SD(I)$ is the standard deviation of the null distribution. For the pros and cons of SES, see Botta-Dukát (2018). Put simply, *SES* represents the departure of the observed variable from the mean of the null distribution, expressed in standard deviation units. However, since the observed Q_{Ws} values are generally lower than the corresponding null values, we typically have: $Mean(Q_{Ws}) > Obs(Q_{Ws})$. Therefore, in this paper, the standardized effect size of species *s* is formulated in such a way that the resulting values are mainly positive:

$$SES_s = \frac{Mean(Q_{Ws}) - Obs(Q_{Ws})}{SD(Q_{Ws})} \qquad (8)$$

Note that in one-factor MANOVA, Eq. 8 (i.e. the standardized effect size of Q_{Ws} changed in sign) is identical to the standardized effect size of the species-level contribution to the between-group sum of squared dissimilarities $Q_{Bs} = Q_{Ts} - Q_{Ws}$, with $Q_{Ts} = \frac{1}{N}\sum_{i>j}^{N} d_{ijs}^2$, thus reinforcing the generality of the proposed approach. For details, see Appendix S1.

Worked example
Data

We used data on plant communities along a primary succession on the foreland of the Rutor Glacier (northern Italy). The data were sampled by Caccianiga et al. (2006) and have been used in many previous papers for exploring community assembly rules along ecological gradients (e.g. Caccianiga et al. 2006; Ricotta et al. 2016; Ricotta et al. 2018). The data set can be found in Appendix S2 and S5 and contains the abundances of 45 Alpine plant species collected in 59 vegetation plots of approximately 25 m². All species abundances were measured with a five-point ordinal scale transformed to ranks. The plots were divided into three successional stages based on the age of the moraine deposits: 17 early-successional plots, 32 mid-successional plots and 10 late-successional plots. For additional details, see Caccianiga et al. (2006).

Based on the species abundances in each plot, we calculated the Euclidean distance between all pairs of plots. Next, using Q_W and Q_{Ws} as test statistics, we tested for overall and species-wise compositional differences among the three successional stages with dissimilarity-based MANOVA. All calculations were done with a new R function available in Appendix S3 and S6. Significance is computed by randomly permuting the 59 plots (data vectors) among the three successional stages with 9999 replicates. For all species that showed significant compositional difference among the three groups of plots, we also tested for pairwise differences among all pairs of groups (9999 permutations) using the same approach described in Pillar & Orlóci (1996) for multivariate contrasts.

Results

The overall compositional differences among the three groups of plots and between all pairs of groups proved highly significant with the minimum possible p-value for 9999 randomizations (p = 0.0001) showing that the three successional stages are compositionally well distinct. The results of

the single-species MANOVA are shown in Figure 2, together with the species abundances in each group and the contributions of single species to the compositional differences among the groups (SES_s).

From a total of 45 species, we identified 32 with a clear preference for certain successional stages ($P < 0.05$). It therefore appears that more than two thirds of the species sampled have rather narrow ecological requirements allowing them to colonize only the most ecologically suitable habitats along the primary succession. Caccianiga et al. (2006) showed that the establishment of the first pioneer species is associated with random dispersal mechanisms that drive the colonization of the glacial deposits by early-successional ruderal forbs, such as *Cerastium uniflorum*, *Oxyria digyna*, or *Tussilago farfara*, whereas the late successional stages are preferentially colonized by stress-tolerant graminoids (*sensu* Grime 1974), such as *Carex curvula*, or *Carex sempervirens*. Mid-successional plots include species with intermediate (ruderal/stress-tolerant) strategies, such as *Achillea moschata* or *Trifolium pallescens*, which are among the species that contribute most to the compositional differences among the successional stages (Figure 2).

Discussion

In this paper, we showed that when we test for compositional differences among two or more groups of plots with one-factor db-MANOVA, the within-group sum of squares Q_W can be additively decomposed into species-level values, provided that the squared dissimilarities d_{ij}^2 can be also additively decomposed into their species-level contributions d_{ijs}^2 such that $d_{ij}^2 = \sum_{s=1}^{S} d_{ijs}^2$. In the remainder we will call a dissimilarity coefficient d_{ij} that conforms to this property 'squared decomposable' (S-decomposable).

The species that contribute most to the compositional differences among the groups can be appropriately called indicator species as they are preferentially concentrated in particular group of plots. Note however that, while classical indicator species analysis usually identifies indicators of single groups of plots, with the proposed method some species may be related to more than one group. By taking into account combinations of groups of plots and adopting a multiple contrast approach (Pillar & Orlóci 1996), this method provides an extra flexibility to model the habitat preferences of species with varying niche breadths (De Cáceres et al. 2010). Readers interested in

comparing the results of the proposed method with the results of classical indicator species analysis using the same data of our worked example are addressed to Ricotta et al. (2020).

A desirable aspect of this method for identifying indicator species is that it is based on the same dissimilarity measure used for multivariate analysis of variance (see Pavoine et al. 2013). This puts indicator species analysis and multivariate analysis of variance under the same mathematical umbrella. Regarding the most appropriate dissimilarity coefficient for db-MANOVA, being S-decomposable, the Euclidean distance appears to be a natural choice. Likewise, all Euclidean distances computed on transformed species abundances, such as the Chord or the Hellinger distance may be equally adequate (see Legendre & Gallagher 2001). Note that, according to Pillar (2013), a comprehensive evaluation of the power and accuracy of one-factor db-MANOVA showed that the Euclidean distance gave better results compared to other dissimilarity coefficients in spite of its well-known limitations for summarizing compositional dissimilarity among plots (see Orlóci 1978; Legendre & Gallagher 2001).

However, the same species-level decomposition can be also extended to virtually any dissimilarity measure of the form $\sqrt{\delta_{ij}}$ provided that δ_{ij} is additively decomposable (A-decomposable) such that $\delta_{ij} = \sum_{s=1}^{S} \delta_{ijs}$. A-decomposable measures include among others the Manhattan distance $\sum_{s=1}^{S} |x_{is} - x_{js}|$, the Canberra distance $\sum_{s=1}^{S} \frac{|x_{is} - x_{js}|}{(x_{is} + x_{js})}$, or the Bray-Curtis dissimilarity $\sum_{s=1}^{S} \frac{|x_{is} - x_{js}|}{\sum_{s=1}^{S}(x_{is} + x_{js})}$. Note however that since the Bray-Curtis dissimilarity summarizes the difference in species abundances between plot i and j compared to the total species abundance in both plots (see Ricotta & Podani 2017), the contribution of each species depends on the values of other species. This is conceptually distinct from classical indicator species analysis which is usually species-specific.

If the square root of an A-decomposable dissimilarity coefficient $\sqrt{\delta_{ij}}$ is used in one-factor db-MANOVA, we obtain:

$$Q_W = \sum_{k=1}^{K} Q_{Wk} = \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i>j}^{N} \delta_{ij} \varepsilon_{ijk} = \sum_{s=1}^{S} \sum_{k=1}^{K} \frac{1}{N_k} \sum_{i>j}^{N} \delta_{ijs} \varepsilon_{ijk} \qquad (9)$$

Therefore, dealing with A-decomposable dissimilarities, Eq. 9 tells us that the within-group sum of squares of any dissimilarity measure of the form $\sqrt{\delta_{ij}}$ can be additively decomposed into species-level values. At the same time, the term Q_W in Eq. 9 can be also interpreted as the within-group sum of *non-squared* A-decomposable dissimilarities δ_{ij}. As such, Q_W is a special case of the 'average dissimilarity within groups' used by Mielke & Berry (2001) in multiresponse permutation procedures (MRPP), which is a class of multivariate permutation methods for testing for compositional differences among a-priori defined groups of plots. This makes the differences between randomization tests based on squared and non-squared dissimilarities much fuzzier than one might think. A short overview of MRPP and its relationship with db-MANOVA is shown in Appendix S4.

Although randomization tests based on A-decomposable non-squared dissimilarities have been rarely used in ecology (partly because they are not included in common statistical packages), they can be equally adopted in one-factor designs for testing for compositional differences among a-priori defined groups of plots *and* for identifying the indicator species that contribute most to the compositional differences among the groups. This allows to expand the number of suitable dissimilarities to the whole arsenal of A-decomposable coefficients already available in the ecological literature (Legendre & De Cáceres 2013). Also, the partitioning of other test statistics could be evaluated, such as the between-groups sum of squares (Q_B) defined in Eq. 1. In this case, the proposed approach could be further extended to multi-factor group comparisons (see Pillar & Orlóci 1996), when it may be relevant to identify indicator species of each group considering the factors separately and their interactions.

Which of these dissimilarities will show the best performance in statistical and biological terms (i.e. accuracy and power vs. biological interpretability)? Can the proposed approach be further extended to functional and phylogenetic dissimilarities, such as the large class of A-decomposable coefficients introduced by Pavoine & Ricotta (2019)? These are relevant questions, and their answers can help shed light on the effects of ecological processes on community composition.

References

Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32–46.

Benjamini, Y. & Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 29: 1165–1188.

Botta-Dukát, Z. 2018. Cautionary note on calculating standardized effect size (SES) in randomization test. *Community Ecology* 19: 77–83. doi: 10.1556/168.2018.19.1.8

Caccianiga, M., Luzzaro, A., Pierce, S., Ceriani, R.M. & Cerabolini, B. 2006. The functional basis of a primary succession resolved by CSR classification. *Oikos* 112: 10–20.

Cai, L. 2006. Multi-response permutation procedure as an alternative to the analysis of variance: An SPSS implementation. Behavior Research Methods 38: 51-59.

Chytrý, M., Tichý, L., Holt, J. & Botta-Dukát, Z. 2002. Determination of diagnostic species with statistical fidelity measures. *Journal of Vegetation Science* 13: 79–90.

De Cáceres, M. & Legendre, P. 2009. Associations between species and groups of sites: indices and statistical inference. *Ecology* 90: 3566–3574.

De Cáceres, M., Legendre, P. & Moretti, M. 2010. Improving indicator species analysis by combining groups of sites. *Oikos* 119: 1674–1684. doi: 10.1111/j.1600-0706.2010.18334.x

Dufrêne, M. & Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345–366.

Edgington, E.S. 1987. *Randomization tests*. Marcel Dekker, New York, NY.

Fortin, M.J. & Jacquez, G.M. 2000. Randomization tests and spatially autocorrelated data. *Bulletin of the Ecological Society of America* 81: 201–205.

Gotelli, N.J. & McCabe, D.J. 2002. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* 83:2091–2096.

Gower, J.C. & Legendre, P. 1986. Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* 3: 5–48.

Grime, J.P. 1974. Vegetation classification by reference to strategies. *Nature* 250: 26–31.

Legendre, P. & De Cáceres, M. 2013. Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters* 16: 951–963. doi: 10.1111/ele.12141

Legendre, P. & Gallagher, E.D. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271–280.

Manly, B.F.J. 1991. *Randomization and Monte Carlo methods in biology*. Chapman and Hall, London.

McArdle, B.H. & Anderson, M.J. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82: 290–297.

Mielke, P.W. & Berry, J.A. 2001. *Permutation Methods: A Distance Approach*. Springer-Verlag, New York.

Orlóci, L. 1978. *Multivariate Analysis in Vegetation Research*. Junk, The Hague, NL.

Pavoine, S. & Ricotta, C. 2019. Measuring functional dissimilarity among plots: Adapting old methods to new questions. *Ecological Indicators* 97: 67-72. doi: 10.1016/j.ecolind.2018.09.048

Pavoine, S., Gasc, A., Bonsall, M.B. & Mason, N.W.H. 2013. Correlations between phylogenetic and functional diversity: mathematical artefacts or true ecological and evolutionary processes? *Journal of Vegetation Science* 24: 781–793. doi: 10.1111/jvs.12051

Pillar, V.D. 2013. How accurate and powerful are randomization tests in multivariate analysis of variance? *Community Ecology* 14: 153-163. doi: 10.1556/ComEc.14.2013.2.5

Pillar, V.D. & Orlóci, L. 1996. On randomization testing in vegetation science: multifactor comparisons of relevé groups. *Journal of Vegetation Science* 7: 585-592.

Podani, J. & Csányi, B. 2010. Detecting indicator species: some extensions of the IndVal measure. *Ecological Indicators* 10: 1119–1124. doi: 10.1016/j.ecolind.2010.03.010

Ricotta, C., Acosta, A.T.R., Caccianiga, M., Cerabolini, B.E.L., Godefroid, S. & Carboni, M. 2020. From abundance-based to functional-based indicator species, *Ecological Indicators* 118: 106761. doi: 10.1016/j.ecolind.2020.106761

Ricotta, C., Bacaro, G., Caccianiga, M., Cerabolini, B. & Pavoine, S. 2018. A new method for quantifying the phylogenetic redundancy of biological communities. *Oecologia* 186: 339–346. doi: 10.1007/s00442-017-4026-x

Ricotta, C., Carboni, M. & Acosta, A.T.R. 2015. Let the concept of indicator species be functional! *Journal of Vegetation Science* 26: 839–847. doi: 10.1111/jvs.12291

Ricotta, C., de Bello, F., Moretti, M., Caccianiga, M., Cerabolini, B. & Pavoine, S. 2016. Measuring the functional redundancy of biological communities: a quantitative guide. *Methods in Ecology and Evolution* 7: 1386–1395. doi: 10.1111/2041-210X.12604

Ricotta, C. & Podani, J. 2017. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. *Ecological Complexity* 31: 201–205. doi: 10.1016/j.ecocom.2017.07.003

Supplementary Material

Appendix S1. On the relationship between $SES(Q_{Ws})$ and $SES(Q_{Bs})$

Appendix S2. Community composition matrix of the Alpine vegetation used in the worked example

Appendix S3. R scripts

Appendix S4. Overview of MRPP and its relationship with db-MANOVA

Appendix S5. Community composition matrix of the Alpine vegetation in text format

Appendix S6. R scripts in text format

Authors contribution statement. CR conceived the idea. CR, SP and VP developed the methodology. CR, SP and BC analyzed the data. SP wrote the R script. CR took the lead in writing the main text; all authors revised the manuscript critically and approved the final version.

Conflicts of Interest. The authors declare no conflict of interest.

Data accessibility statement. Data on the Alpine vegetation used in the worked example are available in Appendix S2 and S5; the R scripts for the dissimilarity-based MANOVA decomposition are available in Appendix S3 and S6.

Figure Captions

Figure 1. Schematic example of a semimatrix of pairwise dissimilarities among 9 plots clustered into two groups. The very essence of dissimilarity-based MANOVA is to compare the within-group dissimilarity with the between-group dissimilarity with an adequate test statistic. The larger the value of between-group dissimilarity compared to within-group dissimilarity, the more likely it is that the groups are compositionally different from each other.

Figure 2. Results of the distance-based MANOVA for the 45 species of the Rutor data set. The species abundances in each group of plots on a five-point ordinal scale and the contribution of each species to the compositional differences among the groups (SES_s) are also shown. The species showing significant compositional differences among the three groups of plots ($P < 0.05$; 9999 permutations) are marked with an asterisk. For those species, we also tested for significant pairwise differences among all pairs of groups. For each species, letters indicate individually significant tests ($P < 0.05$; 9999 permutations) after correction for multiple tests according to Benjamini and Yekutieli (2001).
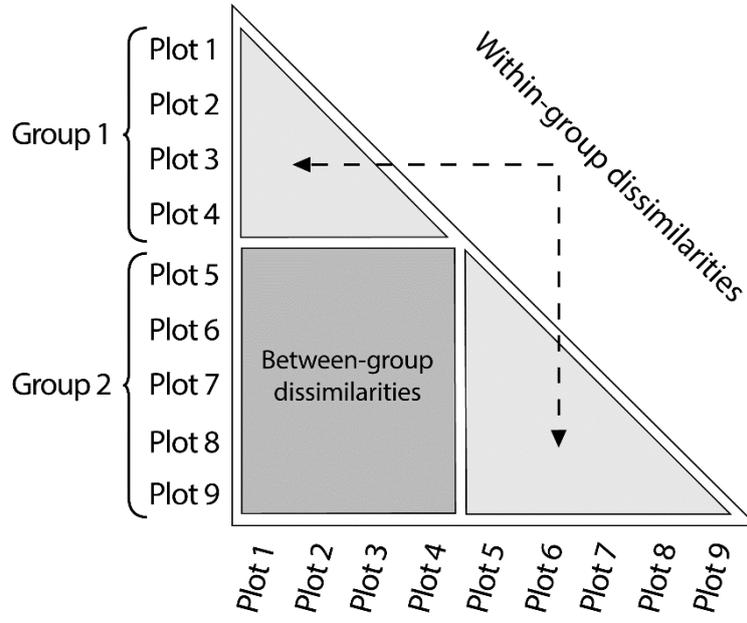
Figure 1.

Figure 2.

	Early successional plots (17)	Mid successional plots (32)	Late successional plots (10)	SES_S
Achillea moschata*	a	b	a	14.319
Adenostyles leucophylla*	a	b		6.860
Agrostis rupestris				0.439
Agrostis schraderiana				0.250
Antennaria dioica*				1.917
Anthoxanthum odoratum*	a	b		5.000
Arabis alpina*	a	b		6.426
Avenula versicolor*	a	b	b	7.111
Cardamine bellidifolia				0.326
Cardamine resedifolia				-0.495
Carex curvula*	a	a	b	13.421
Carex sempervirens*	a	a	b	12.367
Cerastium uniflorum*	a	b	b	11.759
Epilobium fleischeri				0.540
Erigeron uniflorus*				3.734
Euphrasia minima				1.640
Festuca halleri*	a	b	b	3.379
Gnaphalium supinum*				3.410
Hieracium angustifolium*	a	b		9.767
Homogyne alpina*	a	a	b	9.017
Juncus trifidus*	a		b	5.721
Leontodon helveticus*				2.768
Leucanthemopsis alpina*	a	a	b	6.355
Linaria alpina*	a	b	b	8.289
Lotus corniculatus				-0.382
Luzula alpinopilosa				-1.075
Luzula lutea*				6.488
Luzula spicata				1.342
Minuartia recurva*	a	b	b	8.712
Myosotis alpestris*	a	b		5.184
Oxyria digyna*	a	b	b	12.624
Pedicularis kerneri				1.394
Phleum rhaeticum*				2.743
Phyteuma hemisphaericum*	a		b	3.135
Poa alpina*	a	a	b	9.958
Ranunculus glacialis				0.517
Ranunculus kuepferi				1.152
Saxifraga aizoides*	a	b		5.614
Saxifraga bryoides*	a	a	b	5.051
Senecio incanus				-0.186
Silene acaulis*				2.879
Trifolium badium*	a	a	b	5.438
Trifolium pallescens*	a	b	a	13.733
Tussilago farfara*	a	b	b	9.013
Veronica bellidioides*	a	b		3.028