



Original research article

A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data

János Podani^{a,*}, Carlo Ricotta^b, Dénes Schmera^{c,d}^a Department of Plant Systematics, Ecology and Theoretical Biology and Ecology Research Group of HAS, Institute of Biology, L. Eötvös University, Pázmány P. s. 1/C, H-1117 Budapest, Hungary^b Department of Environmental Biology, University of Rome “La Sapienza”, Piazzale Aldo Moro 5, I-00185 Rome, Italy^c Section of Conservation Biology, University of Basel, St. Johanns-Vorstadt 10, CH-4056 Basel, Switzerland^d Balaton Limnological Institute, Centre for Ecological Research, Hungarian Academy of Sciences, Klebelsberg K. u. 3, H-8237 Tihany, Hungary

ARTICLE INFO

Article history:

Received 17 October 2012

Received in revised form 11 January 2013

Accepted 1 March 2013

Available online 12 April 2013

Keywords:

Marczewski–Steinhaus dissimilarity

Jaccard index

Presence–absence

Ruzicka similarity

Simplex

Turnover

ABSTRACT

We describe a procedure for evaluating the relative importance of beta diversity, nestedness, and similarity properties of ecological data matrices containing density, cover or biomass scores of species. Our goals are achieved by extension of the simplex approach – originally proposed for presence–absence data – to abundances. Basically, the method involves decomposition of the Marczewski–Steinhaus coefficient of dissimilarity between pairs of sites into two fractions, one derived from differences between total abundance and the other from differences due to abundance replacement. These are contrasted by the similarity function counterpart, known as the Ruzicka coefficient, and are displayed graphically using ternary (or 2D simplex) plots. Interpretation is aided by calculating percentage contributions from these components to the (dis)similarity structure. Measures of replacement and nestedness are new for abundance data; these are considered complementary phenomena reflecting antithetic ecological processes that are analogous to those operating at the presence–absence level. The method is illustrated by artificial data and a range of actual ecological data sets representing different groups of organisms, different scales and different types of data. While the simplex diagrams and associated coefficients are meaningful by themselves, their comparison with presence–absence based results gives additional insight into data structure and background factors.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Ecological data matrices have long been known as carriers of information on numerous important ecological phenomena, including beta diversity and nestedness. Beta diversity was first defined by Whittaker (1960) as “the extent of change in community composition, or degree of community differentiation, in relation to a complex-gradient of environment, or a pattern of environments”. Whittaker (1960) proposed to quantify beta diversity with two broad categories of measures: beta as a pairwise dissimilarity coefficient between sites (Anderson et al., 2006, 2011; Tuomisto, 2010a,b) or as the ratio of two inventory diversities measured at different scales (i.e. gamma/alpha; Lande, 1996; Veech et al., 2002; Jost, 2007). Recently, Jurasinski et al. (2009) have named these measures ‘differentiation diversity’ and ‘proportional diversity’, respectively. The overwhelming majority of beta diversity functions from both groups apply to

presence–absence data (Vellend, 2001; Koleff et al., 2003; Tuomisto, 2010a,b), including the well-known Jaccard similarity index adapted by Whittaker (1960, p. 320) to this purpose. Much less attention is paid to abundance data in beta diversity analysis, although appropriate expressions are well-known (see e.g., Magurran, 2004).

Nestedness refers to the extent the species of smaller assemblages are a subset of larger assemblages (Atmar and Patterson, 1993). Similarly to beta diversity, earlier definitions of nestedness rely exclusively on presence/absence data (Ulrich et al., 2009) with measures falling into two broad categories: global coefficients such as the nestedness temperature, and averages of pairwise indices (see Podani and Schmera, 2012, for review). The issues of how nestedness may be understood for abundance data and how abundance estimates influence nestedness measurement have been raised only recently (Galeano et al., 2009; Araujo et al., 2010; Almeida-Neto and Ulrich, 2011), therefore our knowledge on nestedness properties of actual data is even more limited than on their beta diversity.

Beta diversity and nestedness are not independent features, and their joint evaluation is promising to reveal and explain ecological

* Corresponding author. Tel.: +36 1 3812 293.

E-mail address: podani@ludens.elte.hu (J. Podani).

factors influencing community composition, structure and functioning. Quantification of their relationship was first suggested by Baselga (2010) via decomposition of pairwise presence–absence based dissimilarity into two components. He used the Sørensen dissimilarity index to measure beta diversity, from which a “spatial turnover” component expressed by the Simpson dissimilarity function was distracted to yield a “nestedness resultant” fraction. Podani and Schmera (2011) and Carvalho et al. (2012, 2013) proposed an algebraic decomposition of Jaccard dissimilarity as a measure of beta diversity into a fraction due to species turnover (or replacement) and another due to difference in the number of species (richness difference). Podani and Schmera (2011) considered the second fraction as a contributor to nestedness, in contrast to replacement which indicates processes completely antithetic to nestedness. One advantage of the above manipulations with Jaccard formula is that they can be embedded with ease into a more general theoretical and methodological framework for analyzing pattern in presence–absence data. According to Podani and Schmera (2011), this involves calculating three complementary indices that measure similarity, relative species replacement, and relative richness difference for all pairs of sites via partitioning pairwise gamma diversity into three additive components, and by displaying the results in a two-dimensional simplex diagram, or ternary plot. In this diagram, a point corresponds to a pair of sites, and the shape and position of the point cloud is informative about community pattern. Percentages are especially useful to evaluate the relative importance of beta diversity, nestedness and agreement in species richness in presence–absence data matrices.

As mentioned earlier, evaluating beta diversity and nestedness in abundance data poses no methodological problems, but there is no general conceptual framework available which handles these aspects of abundance pattern simultaneously. The aim of this paper is thus to extend the simplex approach to abundances (cover, density, biomass, etc.) using the Marczewski–Steinhaus coefficient of dissimilarity and its similarity function counterpart, known as the Ruzicka coefficient. First, we present a summary of abbreviations and new definitions, and then present results for artificial and actual community data. These results demonstrate the utility of our approach in comparing features of presence–absence and abundance data for the same set of study sites.

2. Abbreviations, definitions and functions

Let the abundance data for two sites j and k be presented in vectors \mathbf{x}_j and \mathbf{x}_k . The number of species in the two sites is n , while the number of sites in the dataset is m . The description of different functions starts with those reflecting proportions, which are analogous to the indices described in Podani and Schmera (2011) for presence absence data. All functions listed below have a theoretical range of $[0, 1]$. We assume with good reason that no empty sites appear in the data, so that the denominators of functions that follow can never be zero.

The similarity of sites j and k based on abundances of n species is expressed as the Ruzicka (1958) index

$$S_{Ruz(j,k)} = \frac{\sum_{i=1}^n \min \{x_{ij}, x_{ik}\}}{\sum_{i=1}^n \max \{x_{ij}, x_{ik}\}}. \quad (1)$$

The numerator is the total amount of abundances in which the two sites agree, while the denominator is the possible maximum agreement, henceforth denoted by T_{jk} . The value of S_{Ruz} is 1 if the two sites have identical values for all species, and zero if a positive score in site j is associated with a zero score in site k , or vice versa, for every species. In the presence–absence case, S_{Ruz} simplifies to the Jaccard index of similarity. The complement of Eq. (1) is the

Marczewski–Steinhaus coefficient of dissimilarity, which is a metric (see e.g., Levandowsky and Winter, 1971) and is given by the formula

$$\beta_{MS(jk)} = \frac{\sum_{i=1}^n |x_{ij} - x_{ik}|}{T_{jk}}. \quad (2)$$

Set theoretically, the numerator is the symmetric difference of the abundance data representing the two sites (Orlóci, 1978), corresponding to the total amount of abundances in which they differ. β_{MS} reflects the *relativized abundance turnover* between the two sites, which conceptually corresponds to pairwise *beta diversity* for abundances in our framework.

The sum of absolute differences in the numerator of Eq. (2) can be decomposed into two fractions, which are of central importance in developing the new methodology in this paper. These are analogous to the two fractions obtained from the Jaccard dissimilarity coefficient for presence–absence data (Podani and Schmera, 2011; Carvalho et al., 2012). The first fraction is the absolute deviation between the site totals and is interpretable ecologically as a reflection of the difference between the carrying capacity of the two sites. This, divided by T_{jk} yields the following quantity

$$^aD_{rel(jk)} = \frac{|\sum_{i=1}^n x_{ij} - \sum_{i=1}^n x_{ik}|}{T_{jk}} \quad (3)$$

which is called the *relativized abundance difference* measure. (Superscript a distinguishes this function and the forthcoming equations from those applicable to presence–absence data as used in Podani and Schmera, 2011.) Minimum value, i.e., zero is obtained when the site totals are identical. In practice, it never takes the maximum value, that is 1, because this is possible only if one of the sites is completely empty. The complement of relativized abundance difference is the *relativized abundance agreement*, given by the following formula:

$$^aA_{rel(jk)} = \frac{T_{jk} - |\sum_{i=1}^n x_{ij} - \sum_{i=1}^n x_{ik}|}{T_{jk}} \quad (4)$$

The second fraction of the Marczewski–Steinhaus coefficient comes from the sum of abundances in site j that are replaced by the same amount of abundances in site k , pertaining to completely different species. This is called the absolute abundance replacement for the sites. Division by T_{jk} gives the *relativized abundance replacement* function

$$^aR_{rel(jk)} = \frac{\sum_{i=1}^n |x_{ij} - x_{ik}| - |\sum_{i=1}^n x_{ij} - \sum_{i=1}^n x_{ik}|}{T_{jk}} \quad (5)$$

for the two sites being compared. Its value is zero when the difference between site totals equals the sum of absolute differences between the abundances – which is possible only if the two sites can be labelled by j and k such that $x_{ij} \geq x_{ik}$ for all i . In words, abundances in one site can exceed the abundances in the other for none of the species. In this case, there are no abundances that are replaced, only surplus (or gain) on one side. The maximum value, 1, reflects a situation in which the sum of absolute differences equals the sum of maxima: it is possible only if site totals are the same and the two sites share no species at all. Ecologically, this value reflects agreement in the carrying capacity of the two sites while environmental conditions are completely different causing maximum floristic dissimilarity.

Now, we introduce a *nestedness concept* for abundance data and define a function for its quantification. Perfect nestedness will be understood as a situation when abundances in one site are not smaller than the abundances in the other for every species. For

example, if two sites are characterized by the following scores for four species

j	k
10	5
7	2
2	2
1	0

then we can say that site k is perfectly nested in site j , and a meaningful nestedness measure should yield its maximum value. Actually, this is the situation when $^aR_{rel}$ (Eq. (5)) is zero. Correspondingly, the complement of relativized abundance replacement serves as a measure of *relativized nestedness*:

$$^aN_{rel(jk)} = \frac{\sum_{i=1}^n \min\{x_{ij}, x_{ik}\} + |\sum_{i=1}^n x_{ij} - \sum_{i=1}^n x_{ik}|}{T_{jk}} \quad \text{if } \sum_{i=1}^n \min\{x_{ij}, x_{ik}\} > 0$$

$$^aN_{rel(jk)} = 0 \text{ otherwise.} \quad (6)$$

In other words, nestedness for abundance data is calculated as the sum of similarity and relativized abundance difference. The condition that the sum of minima is larger than zero is necessary, because without overlap there is no nestedness at all in the presence/absence case (Almeida-Neto and Ulrich, 2011) and logic dictates to maintain this condition here as well. If sites with equal totals (a rare coincidence in actual abundance data) are to be excluded from the comparison, the above equation – more precisely, the condition of positive support – modifies to

$$^aN'_{rel(jk)} = \frac{\sum_{i=1}^n \min\{x_{ij}, x_{ik}\} |\sum_{i=1}^n x_{ij} - \sum_{i=1}^n x_{ik}|}{T_{jk}} \quad \text{if } \sum_{i=1}^n \min\{x_{ij}, x_{ik}\} > 0 \text{ and } \sum_{i=1}^n x_{ij} \neq \sum_{i=1}^n x_{ik},$$

$$^aN'_{rel(jk)} = 0 \text{ otherwise.} \quad (7)$$

On the analogy of the presence/absence version (Podani and Schmera, 2011) this is called the *relativized strict nestedness* function. Again, identical site totals are exceptional in case of abundances, so that this function is likely to be only of theoretical interest.

3. The SDR simplex for abundances

The sum of three coefficients given above, namely S_{Ruz} , $^aD_{rel}$ and $^aR_{rel}$, is always 1, allowing the use of a 2D (two-dimensional) simplex (Podani and Schmera, 2011) which is routinely applied in science to summarize relationships for three quantities which add up to 1. The graphical illustration of the 2D simplex is an equilateral triangle, the so-called *ternary plot* (or triangle plot, simplex plot). In this, the vertices correspond to these three functions, and each pair of sites is represented by a point, with distances from the vertices inversely proportional to the corresponding coefficients. Prepared for all pairs of sites, the point scatter in the ternary plot will reveal the abundance structure in the data in terms of similarity, relativized abundance difference and relativized abundance replacement. Illustration is similar to that of the SDR-simplex proposed for presence–absence data, following the same standard for naming the corners (Fig. 1).

Three 1D simplices can be derived from the 2D simplex, by summing two quantities at a time and using the third coefficient as a contrast. Graphically, it involves projection of points in the diagram to one of the medians of the triangle (Fig. 1). These 1D simplices are as follows:

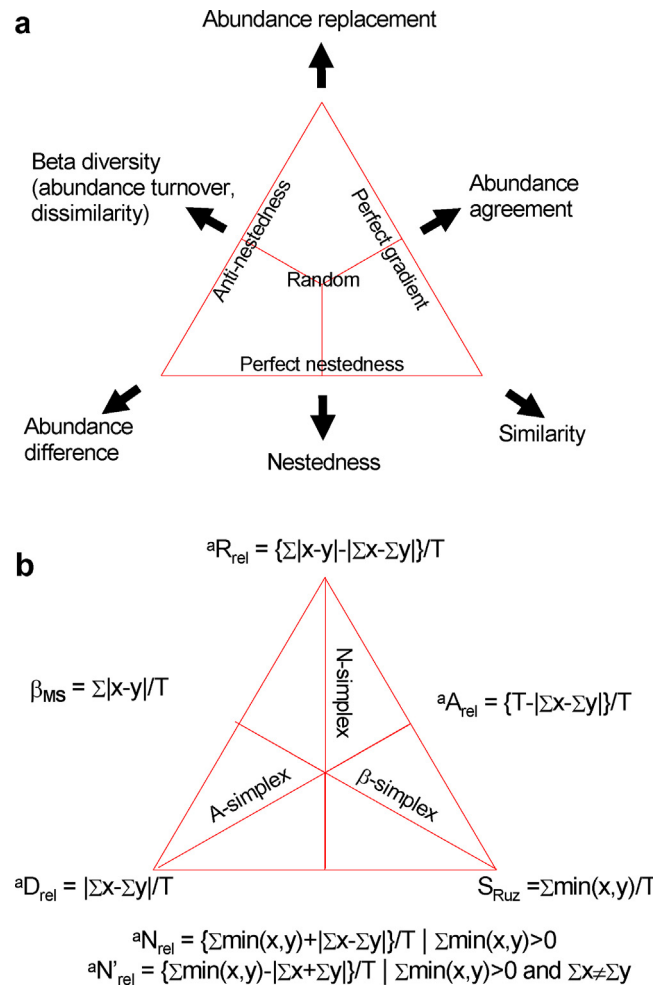


Fig. 1. Concepts (a) and calculations (b) associated with the SDR simplex approach for abundances. In (b), x, y, T and \sum refer to x_{ij}, x_{ik}, T_{jk} and \sum_i , respectively, to simplify illustration.

- 1) *Beta diversity*- or *β -simplex*; representing contrast between similarity (Eq. (1)) and dissimilarity (abundance turnover or beta diversity, Eq. (2)),
- 2) *Abundance agreement*- or *A-simplex*; illustrates contrast between abundance difference (Eq. (3)) and agreement (Eq. (4)),
- 3) *Nestedness*- or *N-simplex*, is an antithesis between abundance replacement (Eq. (5)) and nestedness (Eqs. (6) or (7)).

Recall that decomposition along each of these three simplices involves measurement of complementary terms, for example, similarity is one-complement to dissimilarity. Examining the position of a given point in these simplices offers evaluation of the relative importance of different fractions in determining abundance structure. For example, the closer a point on the N-simplex to the bottom edge, the more deeply nested is one site in the other in terms of abundance values. Closeness to the opposite corner R implies that both sites have large amounts of abundances carried by different species. It has been suggested that these two endpoints are on a gradient that shows the antagonistic relationship between nestedness and replacement. The beta diversity simplex is easy to interpret: closeness to the S vertex implies high similarity of the two sites, whereas closeness to the opposite edge means high dissimilarity, or beta diversity. The abundance agreement simplex shows how different the sites are in total abundances: closeness to the D corner means large differences,

whereas close proximity to the opposite edge shows that the sites in question have similar abundance totals.

3.1. Percentages

Whereas the simplex plots provide efficient graphical illustration of overall data structure, the picture may be incomplete if many points overlap in the diagram (Podani and Schmera, 2011), even though for abundance data the probability of overlaps is considerably smaller. Therefore, it is useful to examine percentage contributions based on the average values of the above functions for all possible pairs of m sites in the data.

On the 2D simplex, percentage mean similarity is

$$P\bar{S}_{Ruz} = \frac{200 \sum_{j < k} S_{Ruz(jk)} / T_{jk}}{m^2 - m} \quad (8)$$

percentage mean relativized abundance difference:

$$P^a \bar{D}_{rel} = \frac{200 \sum_{j < k} {}^a D_{rel(jk)} / T_{jk}}{m^2 - m} \quad (9)$$

percentage mean relativized abundance replacement

$$P^a \bar{R}_{rel} = \frac{200 \sum_{j < k} {}^a R_{rel(jk)} / T_{jk}}{m^2 - m} \quad (10)$$

These three values sum to 100 and, if scaled back to the unit range, determine the centroid of the point cloud in the ternary plot.

Contributions obtained by adding two of the above quantities at a time are percentage beta diversity for abundances

$$P\bar{\beta}_{MS} = P^a \bar{D}_{rel} + P^a \bar{R}_{rel} \quad (11)$$

the percentage mean abundance agreement

$$P^a \bar{A}_{rel} = P\bar{S}_{Ruz} + P^a \bar{R}_{rel}, \quad (12)$$

and the percentage mean nestedness

$$P^a \bar{N}_{rel} = P\bar{S}_{Ruz} + P^a \bar{D}_{rel} - \frac{200 \sum_{j < k} \{ {}^a D_{rel(jk)} / T_{jk} \text{ if } \sum_i \min \{ x_{ij}, x_{ik} \} = 0, \text{ otherwise } 0 \}}{m^2 - m} \quad (13)$$

in which the last term is the *anti-nestedness fraction* of mean relativized abundance difference which does not contribute to nestedness because intersection is 0. Finally, percentage mean strict nestedness for abundances is

$$P^a \bar{N}'_{rel} = P^a \bar{N}_{rel} - \frac{200 \sum_{j < k} \{ S_{Ruz(jk)} / T_{jk} \text{ if } \sum_i x_{ij} = \sum_i x_{ik}, \text{ otherwise } 0 \}}{m^2 - m} \quad (14)$$

in which the last term refers to the total richness identity fraction calculated from similarities between sites that have exactly the same site totals.

Computer program SDR-abunSimplex has been written to perform the calculations. The program and a short guide may be downloaded from <http://ramet.elte.hu/~podani>.

4. Artificial examples

The manner the positions of points are influenced by the data is demonstrated by artificial model matrices representing various data structures (insets in Fig. 2). The first example illustrates situations when the points are at the corners. Then, we examine cases in which the points are positioned on the edges of the ternary plot and the final example shows a relatively balanced situation when the points fall inside the triangle.

- Corner*: for complete identity of the two sites in question, the similarity is 1, and the corresponding point will be at the S corner of the triangle. When totals for the two sites are identical but the sites have no species in common, we have the situation that all abundances in site 1 are replaced by other species in site 2. Consequently, the point will be at the R corner. In actual data, the D corner is never taken, because it would mean that one site has no abundances at all. In practice, a point can be close to it on the left edge if, for example, the sites have no species in common, and one of them has much higher abundance values than the other;
- Difference-dominated perfect nestedness*: all points are on the bottom edge, most of them relatively close to the D corner;
- Overlap-dominated perfect nestedness*: all points are on the bottom edge, relatively close to the S corner;
- Anti-nestedness*: if all points are on the left edge, then we have the quantitative analogue of the anti-nestedness situation in presence-absences;
- Site total identity with irregularity*: when all site totals agree, then all points fall onto the right edge, even though there is no trend in the data;
- Site total identity with perfect gradient*: a special case of abundance agreement is when there is an underlying gradient with continuous partial replacement of abundances between the sites;
- Random*: in general, for randomly assigned scores the points will be scattered around the midregion of the triangle.

An important aspect of the simplex method, namely how the results change when interest is shifted from abundances to presence/absence is also illustrated by artificial examples. It is easy to see that, if a pair of sites is positioned in the R or the D corner, it remains there no matter which type of data is used. Otherwise, the points may move along a particular edge or, in general, in an unpredictable manner when data type is modified (Fig. 3).

- On the right edge at which sites with identical totals are positioned, two types of change may happen. The points may be on the edge for the presence/absence data, but then move into the interior in the abundance version, but change in the other direction is also possible (Fig. 3a).
- On the left edge, at which site pairs fitting the anti-nestedness model are positioned, moves along the edge are possible only. That is, anti-nestedness for the presence/absence case implies anti-nestedness for the abundances automatically, and vice versa (Fig. 3b).
- For perfect nestedness, with points on the bottom edge, there is only one constraint. Perfect nestedness for abundances implies that the sites remain perfectly nested even if we reduce the data to presence/absences. However, perfect nestedness in the presence/absence case does not necessarily mean the same relationship for abundances (Fig. 3c). This latter case exemplifies the unsymmetric relationship between the nestedness measures for the two data types. For example, for the following pair of sites

j	k
1	10
1	1
1	0
1	0

site k is perfectly nested in j in terms of presence/absence ($N_{rel} = 1$), whereas site j is nested, although not perfectly, in k for abundances (${}^a N_{rel} = 0.69$).

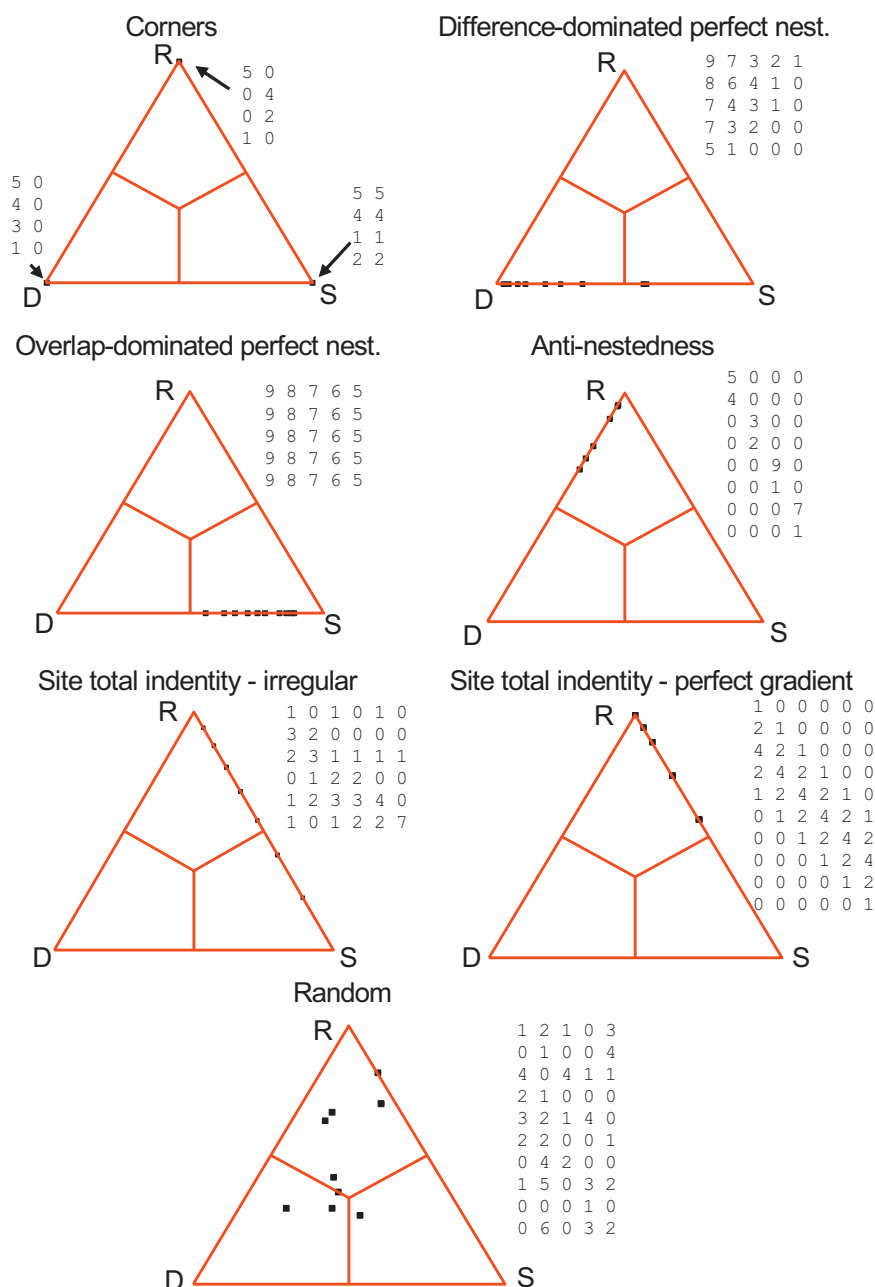


Fig. 2. The abundance simplex diagram for model matrices (insets), with species as rows and sites as columns.

5. Actual examples

5.1. Rock grassland

This data set comes from an extensive study of rock grasslands on the dolomite bedrock of Sas-hill, lying within the city limits of Budapest, Hungary (Podani, 1998). Eighty sample units were selected in the grasslands, representing open rock grassland, closed grassland and slope steppe. Each sample unit consisted of a series of 8 nested quadrats with a common corner, the smallest being $0.5 \text{ m} \times 0.5 \text{ m}$, and the largest $4 \text{ m} \times 4 \text{ m}$, with 0.5 m side increments in between. Percentage cover of vascular plants was recorded within each plot for each size. For the present study, we used only the smallest and the largest quadrat sizes: $0.5 \text{ m} \times 0.5 \text{ m}$, and $4 \text{ m} \times 4 \text{ m}$ to demonstrate the effect of extreme quadrat size

changes upon the decomposition of beta diversity. For more details, see Electronic supplement.

The influence of measurement and spatial scales on data pattern is obvious (Fig. 4). The position of the point cloud relative to the left edge of the triangle, as well as the percentage contributions (79% and 87.5%, see Table 1) reflect that beta diversity is the highest for small quadrats for both data types. Increase in quadrat size leads to a more balanced situation, with similarity being the highest for presence-absence data. High beta diversity for $0.5 \text{ m} \times 0.5 \text{ m}$ quadrats is attributable to the fact that the grassland has a mosaic-like spatial structure, which becomes obscured upon quadrat size increases. Nevertheless, there is considerable robustness in the data because shift from presence-absence data to abundances provides the same trends for both quadrat sizes: similarity, richness agreement and

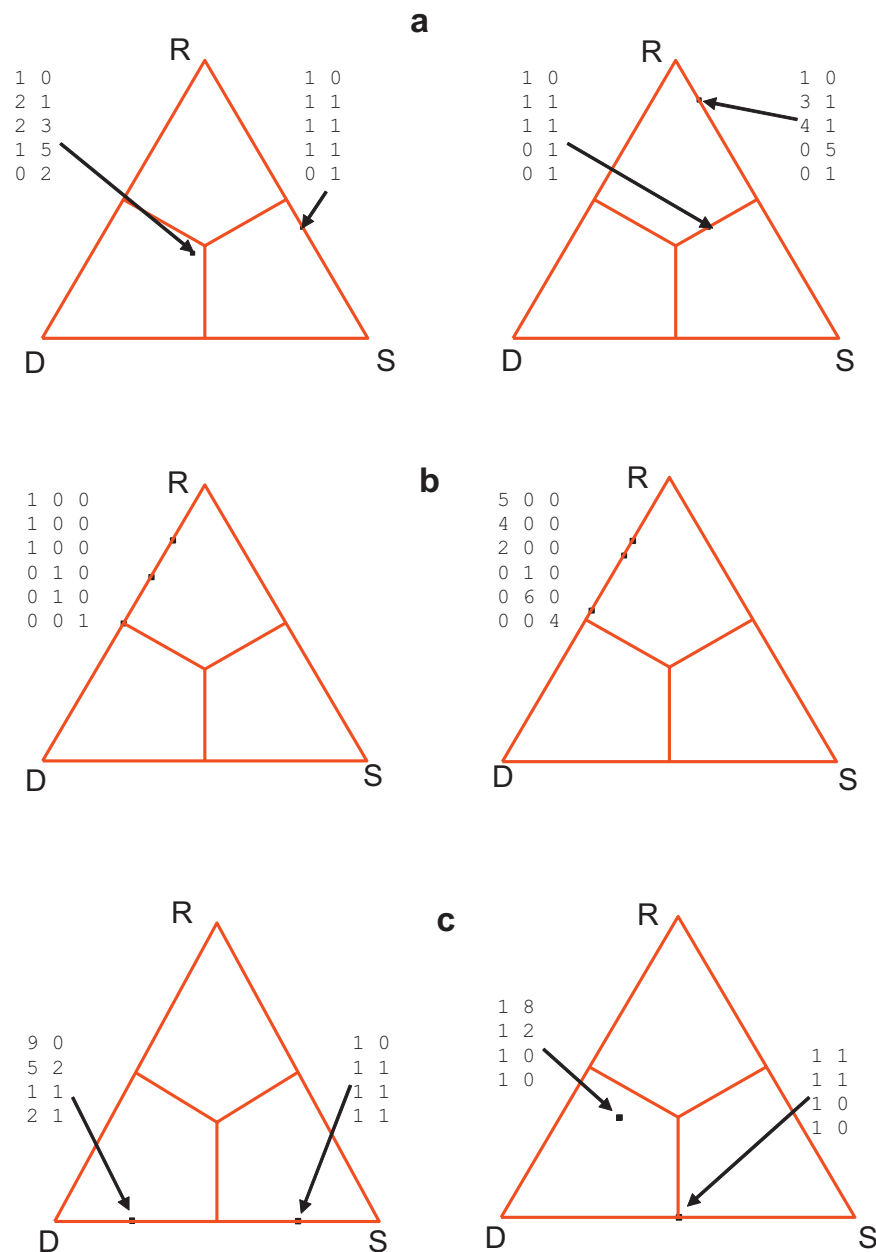


Fig. 3. Potential effects of changing data type on the position of points on the edges of the simplex plot. (a) identical vs different totals, (b) anti-nestedness for both data types, (c) nestedness as influenced by abundances. In the data sets species are rows and sites are columns.

nestedness decrease, whereas beta diversity, richness difference and replacement increase. A practical conclusion from these observations is that the simplex approach is useful to evaluate scale dependence in the real topographical space (quadrat size) as well as in the conceptual data space (data type).

5.2. Epiphyte moss assemblages

The dataset derives from a sampling study of bryophyte assemblages on 90 tree individuals in the Őrség National Park, southwestern Hungary. The trees represented 3 species, *Fagus sylvatica*, *Carpinus betulus* and *Quercus petraea*, 30 individuals each. The number of moss species detected was 26, each value in the data matrix representing the absolute cover (dm^2) of a given species from the soil surface up to 1.5 m height of a tree (Király and Ódor, 2010).

The three components of the 2D simplex, *S*, *D* and *R* are fairly balanced in the presence absence case, suggesting that the spatial distribution of mosses on trees follows no particular pattern (Fig. 5, Table 1). When absolute cover of species is considered, abundance differences are larger while replacement becomes much smaller than in the p–a case. This mutual compensation causes that beta diversity is not influenced much by data type, whilst this is the only case study in this paper in which beta is larger, although slightly, for presence–absences than abundances. Abundance agreement strongly decreases and nestedness greatly increases when data type is changed from presence–absence to cover. These are explained by the dominance of a very few common species in the moss assemblages (mainly *Hypnum cupressiforme*). The overall picture does not change if data are split into three subsets according to the host tree species (results not shown), suggesting that tree species identity is not influential in determining bryophyte distributional pattern in the park.

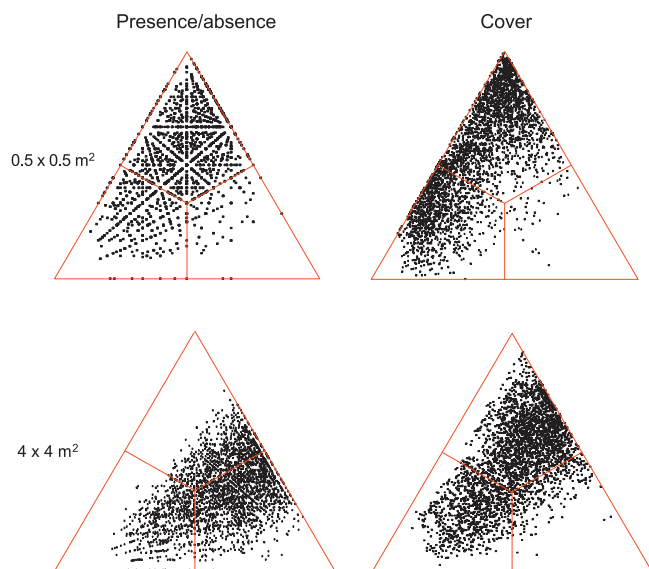


Fig. 4. Comparison of results of the SDR simplex method as applied to the Sashegy data for two types of data and two quadrat sizes.

5.3. Danube river macroinvertebrates

The Danube River was sampled at its full length for macroinvertebrate species during the Second Joint Danube Survey in 2007 (see Podani and Csányi, 2010). The number of sample sites selected fairly regularly along the river was 74. A total of 173 species were detected, and their density values were recorded from benthic samples. See electronic supplement for list of species.

It is remarkable that replacement is almost the same for presence–absences and abundances so its complement, nestedness is also invariant for data type (Table 1). For both data types, differences in site totals are fairly large, although replacement contributes even more to beta diversity, reflecting continuous faunistic turnover along the Danube river. Rearrangement due to a shift in data type is manifested in larger beta diversity (and smaller agreement) for the abundances, easily recognizable on the simplex diagrams (Fig. 6). Contrary to the previous example, however, this can only be explained by extreme abundance variances for several species.

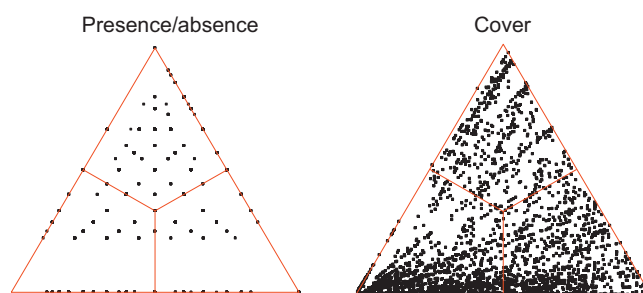


Fig. 5. Comparison of results of the SDR simplex for the epiphytic moss data.

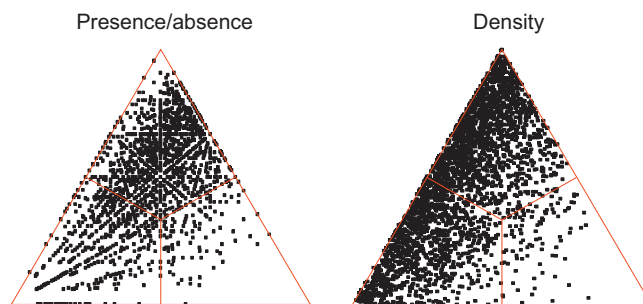


Fig. 6. Comparison of results of the SDR simplex for the Danube invertebrates data.

5.4. Raba river macroinvertebrates

Density data of benthic macroinvertebrates, i.e., counts, were recorded at 18 sites along the Raba River, Eastern Austria and Western Hungary (Szekeres et al., 2011), yielding a total of 196 taxa, most of them identified at the species level (see Electronic Supplement). Depending on taxa, individual scores may be as high as several thousand.

The simplex diagrams (Fig. 7) clearly illustrate that the macroinvertebrates in the Raba have a completely different presence–absence pattern than in the Danube. Replacement is the dominant process over the entire length of the river, whereas site totals are much more similar than in the Danube (Table 1). For density, the overall picture is similar to the Danube, while the scarcity of points is explained by the much smaller sample size in the Raba study. Beta diversity and nestedness both increase

Table 1
Percentage contributions of various components of pairwise gamma diversity for various data sets for abundances and presence–absences. Nestedness is calculated according to Eq. (13).

Dataset		S	D	R	Beta	Agreement	Antinest.	Nest.
Rock grassland 0.5 m × 0.5 m	P/A	21.0	26.6	52.4	79.0	73.4	0.9	46.7
	Cover	12.5	33.5	54	87.5	66.5	1.1	44.9
Rock grassland 4 m × 4 m	P/A	43.2	23.1	33.7	56.8	76.9	0	66.3
	Cover	27.5	27	45.5	72.5	73	0	54.5
Epiphytic mosses	P/A	30.7	35.6	33.7	69.3	66.2	0.9	63.5
	Abs. cover	34.2	52.5	13.3	65.8	47.4	1.7	85.0
Danube invertebrates	P/A	22.0	30.4	47.6	78	69.6	0.8	51.7
	Density	13.4	39.5	47.1	86.6	60.5	0.5	52.4
Raba invertebrates	P/A	22	13	64	77	87	0	36
	Density	12	35	53	88	65	0	47
Italian shrubland	P/A	47.1	16.0	36.9	52.9	83.9	0	63.1
	Cover	26.1	24.0	49.9	73.9	76.0	0	50.1
Oribatids	P/A	54	27	19	46	81	0	73
	Abund.	37	39	24	63	61	0	76

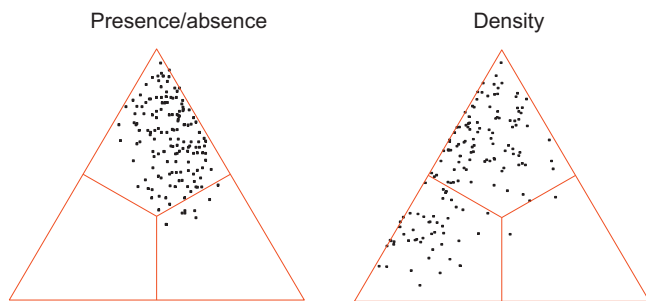


Fig. 7. Comparison of results of the SDR simplex method as applied to the Raba invertebrates data.

considerably when switching from presence–absences to abundances.

5.5. Italian shrubland data

Thirty-five square plots of 1 m × 1 m in size were sampled in spring 1994 in a species-poor garrigue community on serpentine soils south of Siena, Tuscany (Chiarucci et al., 1998). The plots were sampled with the point quadrat method (Moore and Chapman, 1986) with a density of 441 pins/m². Species present in a plot but not touched by any pins were recorded with an arbitrary cover of 0.1%. As a result of field work, the total number of species was 31.

The high overall similarity for the presence absence case, and the considerably higher beta diversity (especially the replacement part) for cover (Table 1, Fig. 8) are striking. This is probably because, due to the peculiar chemical properties of serpentine soils with low nutrient levels and high concentrations of potentially toxic elements (such as magnesium, chromium, nickel and cobalt, Chiarucci et al., 1998), these garrigues host a very distinctive flora with relatively stable species composition. Nonetheless, while plant growth is generally limited by the infertility of such soils, the abundance of the dominant species may vary substantially according to the local availability of nutrients and water. The arrangement of points for the cover data is superficially similar to that observed for the Sashegy grasslands at much larger quadrat sizes, which has to do with the fewer number of species and the overall homogeneity in the garrigue vegetation.

5.6. Oribatid mites data

A single site in a dry oak forest in Hungary was sampled 27 times in the same year (i.e., every two weeks in 2009). In the foena layer (the thin horizon between litter and topsoil), oribatids were collected and were identified at the genus level (Gergőcs et al., 2011). The data set contains the number of individuals for each

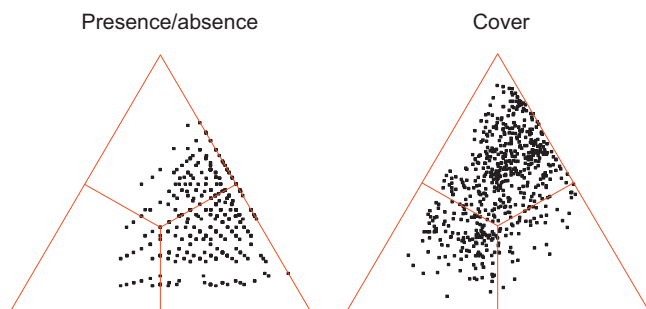


Fig. 8. Comparison of results of the SDR simplex method as applied to the Italian shrubland (garrigue) data set.

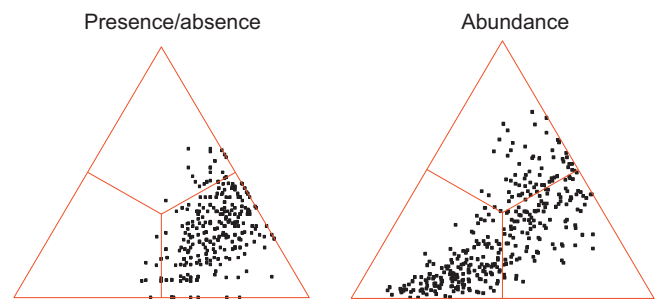


Fig. 9. The SDR simplices showing temporal data structure in the oribatid assemblage of the foena level in the understorey of an oak forest.

genus (52) at each sampling date and comparisons were made in all possible pairs of sampling dates.

The relatively large overall similarity of sample dates for the presence–absence data (the maximum in this study, Table 1, Fig. 9) is not surprising, and shows that in this single site the fauna is remarkably stable over the year both in structure and in site totals. Nevertheless, the temporal pattern differs with data type because abundance differences are more influential, suggesting temporal fluctuations in the size of oribatid populations.

6. Discussion

Structural features in community level data are routinely expressed in terms of characteristic values, such as beta diversity, nestedness, mean similarity and so on. While the literature abounds in coefficients developed for or adjusted to the presence–absence case, less attention has been paid to abundances. Furthermore, no method was available as yet for comparing the p–a and abundance pattern in the same set of sites in an algebraically and ecologically logical way. The present paper was written to fill this methodological gap. It is the first attempt to define a conceptual and methodological framework for evaluating structure in abundance data by considering comparable and additive components of beta diversity, nestedness and other phenomena of pattern simultaneously. The innovation involves the decomposition of the Marczewski–Steinhaus coefficient of dissimilarity into two fractions and the use of its complement, the Ruzicka index. Since these formulae correspond to functions derived from the Jaccard coefficient used in the presence absence case, the approach is a logical extension of the simplex approach developed by Podani and Schmera (2011).

The essence of the method is to partition the abundance data for a pair of sites into three components: similarity (species abundances in which the two sites agree), difference (the amount by which the total in either site exceeds the total of the other) and replacement (species abundances in either site replaced by abundances pertaining to *different* species in the other site). After normalization to unit sum, these quantities determine the position of the given site pair in a two-dimensional SDR simplex plot. In this, the scatter of points for all the possible pairs of sites demonstrates data structure in terms of the above three quantities, plus beta diversity (difference + replacement = Marczewski–Steinhaus dissimilarity), nestedness (difference + similarity, with conditions) and abundance agreement (similarity + replacement). In addition to graphical display, percentage contributions provide numerical results for a more quantitative analysis.

To the best of our knowledge, the Marczewski–Steinhaus coefficient, β_{MS} has been suggested newly in this paper for calculating beta diversity in cover, density or biomass data. Of pairwise dissimilarity coefficients, the Bray–Curtis (Magurran, 2004, p. 174) and the Morisita–Horn (Wolda, 1983) indices have

received applications to date. Obviously, many other forms of abundance-based dissimilarities are also conceivable as measures of beta diversity, including more complex formulae based on phylogenetic relationships of the constituting species (Lozupone et al., 2007). Historically, abundance data were first used as the starting basis for calculating beta diversity only indirectly, by doing an ordination first and then calculating gradient length as the measure sought (Wilson and Mohler, 1983) – an approach not functional when no obvious gradients exist (but see Anderson et al., 2006, for a multivariate solution). However, the advantage of using β_{MS} against other proposals is that one of its fractions is conceived as a contributor to nestedness as well, establishing a mathematical link between two important ecological concepts.

Our measure of nestedness in abundance data, $^a N_{rel}$ is derived as the sum of Ruzicka similarity and relativized abundance difference component of β_{MS} , provided that the first component is nonzero. Previously, only two other approaches had been used for measuring nestedness based on abundances. In a pioneering study, Galeano et al. (2009) proposed the WINE statistic which is essentially the mean of weighted Manhattan distances calculated over the cells of the data matrix. The result is sensitive to the absolute magnitude of scores, and therefore the data require standardization. Almeida-Neto and Ulrich (2011) suggested a weighted extension of the NODF measure (which is the average of pairwise Simpson similarities, see Podani and Schmera, 2012) to abundances (WNODF). In fact, their method derives directly from the original presence–absence version of the coefficient, that is, nestedness for abundances strongly depends on nestedness in presence–absences. As a result, if nestedness is lacking for the NODF measure, then there is no nestedness for abundances either, as obvious from the example in Fig. 2.C in Almeida-Neto and Ulrich (2011):

9	8	7	6	5
8	7	6	5	4
7	6	5	4	3
6	5	4	3	2
5	4	3	2	1

Here neither NODF (if data are reduced to presence–absence) nor WNODF will detect any nested pattern, because sites have equal number of species (i.e. species presences). For the above example, $^a N_{rel}$ is 1 for all pairs sites (no matter if they are rows or columns) and therefore the matrix proves to be completely nested in terms of abundances ($^a P\tilde{N}_{rel} = 100\%$). This agrees well with intuitive expectations, because there is a strictly monotonous change of scores in all columns and all rows of the matrix. Therefore, we feel that rigorous demand of richness difference as a prerequisite to positive nestedness (as expressed by Ulrich and Almeida-Neto, 2012) may prevent us to recognize truly nested patterns in abundance data. Further advantage of our approach is its order-invariance: the same result obtains no matter how the rows and the columns of the matrix are arranged, which is not so for WINE and WNODF. This is important when there is no unique solution for ordering the species and sites in the data.

The third statistic derived from combining two simplex components is abundance agreement, $^a A_{rel(jk)}$. Such an expression has never been used in ecology, even though it reflects an important ecological property of the community: the similarity of sites in their carrying capacity. The positions of points projected on the A simplex are informative on the distribution of the site totals; the more scattered the points the higher the variance of sites in total abundances.

The utility of our approach is demonstrated by artificial and actual ecological examples. In general, the comparison of results of the analyses based on p–a and abundance data suggests that actual

data sets weighted by species abundances may show a drastically different picture about community organization when compared to analyses of p–a matrices. While it is often argued that abundances do not express much more than presence–absence data in a multivariate context (e.g., Wilson, 2012), it may very well be true that low beta diversity in presence/absences is associated with large differences in abundances, or vice versa. It is clear, therefore, that the abundance-based analyses are not simple mathematical extensions of p–a based analyses. While p–a measures are most informative when communities differ primarily by their species composition (e.g. in presence of strong environmental filters), quantitative measures may reveal more subtle differences that are due to the presence of a limiting nutrient source or to responses to a pollutant or other impact. Our examples illustrated only a few cases, while the interpretation of beta diversity, nestedness and other measures for the two different data types for the same set of study objects still remains an interesting challenge in quantitative ecology.

The percentage statistics may represent the basis for significance testing, a topic which was beyond the scope of the present communication. A good starting point may be the study of Ulrich and Gotelli (2010) who propose 14 different null model algorithms for randomizing abundance data matrices. We feel, however, that methods for statistical evaluation of simplex arrangements should first be developed for the presence/absence case, which may then be extended to abundances. We plan to elaborate this topic in the near future.

Acknowledgments

We are grateful to B. Csányi, W. Graf, I. Király, P. Ódor, V. Gergőcs and A. Chiarucci for placing their data at our disposal.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecocom.2013.03.002>.

References

- Almeida-Neto, M., Ulrich, W., 2011. A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling and Software* 26, 173–178.
- Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders, N.J., Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C., Swenson, N.G., 2011. Navigating the multiple meanings of β -diversity: a roadmap for the practicing ecologist. *Ecology Letters* 14, 19–28.
- Anderson, M.J., Ellingsen, K.E., McArdle, B.H., 2006. Multivariate dispersion as a measure of beta diversity. *Ecology Letters* 9, 683–693.
- Araújo, A.L.L., de Almeida, A.M., Cardoso, M.Z., Corso, G., 2010. Abundance and nestedness in interaction networks. *Ecological Complexity* 7, 494–499.
- Atmar, W., Patterson, B.D., 1993. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96, 373–382.
- Baselga, A., 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* 19, 134–143.
- Carvalho, J.C., Cardoso, P., Gomes, P., 2012. Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography* 21, 760–771.
- Carvalho, J.C., Cardoso, P., Borges, P.A.V., Schmera, D., Podani, J., 2013. Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos*, <http://dx.doi.org/10.1111/j.1600-0706.2012.20980.x>.
- Chiarucci, A., Maccherini, S., Bonini, I., De Dominicis, V., 1998. Effects of nutrient addition on species diversity and ground cover of serpentine vegetation. *Plant Biosystems* 132, 143–150.
- Galeano, J., Pastor, J.M., Iriando, J.M., 2009. Weighted-interaction nestedness estimator (WINE): a new estimator to calculate over frequency matrices. *Environmental Modelling and Software* 24, 1342–1346.
- Gergőcs, V., Garamvölgyi, Á., Homoródi, R., Hufnagel, L., 2011. Seasonal change of oribatid mite communities (Acari, Oribatida) in three different types of

- micrihabitats in an oak forest. *Applied Ecology and Environmental Research* 9, 181–195.
- Jost, L., 2007. Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439.
- Jurasinski, G., Retzer, V., Beierkuhnlein, C., 2009. Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. *Oecologia* 159, 15–26.
- Király, I., Ódor, P., 2010. The effect of stand structure and tree species composition on epiphytic bryophytes in mixed deciduous–coniferous forests of Western Hungary. *Biological Conservation* 143, 2063–2069.
- Koleff, P., Gaston, K.J., Lennon, J.J., 2003. Measuring beta diversity for presence–absence data. *Journal of Animal Ecology* 72, 367–382.
- Lande, R., 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* 76, 5–13.
- Levandosky, Winter, M.D., 1971. Distance between sets. *Nature* 234, 34–35.
- Lozupone, C.A., Hamady, M., Kelley, S.T., Knight, R., 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology* 73 (5) 1576–1585.
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Oxford Univ. Press, Oxford.
- Moore, P.D., Chapman, S.B., 1986. *Methods in Plant Ecology*. Oxford Univ. Press, Oxford.
- Orlói, L., 1978. *Multivariate Analysis in Vegetation Research*. Junk, The Hague.
- Podani, J., 1998. Numerikus cönológiai vizsgálatok a Sas-hegy (Budai hg.) dolo-mitsziklagyepjeiben. In: Csontos, P. (Ed.), *Sziklagyepék szünbotanikai kutatása*. Scientia, Budapest, (in Hungarian), pp. 213–229.
- Podani, J., Csányi, B., 2010. Detecting indicator species: some extensions of the IndVal measure. *Ecological Indicators* 10, 1119–1124.
- Podani, J., Schmera, D., 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos* 120, 1625–1638.
- Podani, J., Schmera, D., 2012. A comparative evaluation of pairwise nestedness measures. *Ecography* 35, 889–900.
- Ruzicka, M., 1958. *Anwendung mathematisch-statistischer Methoden in der Geobotanik (synthetische Bearbeitung von Aufnahmen)*. Biologia, Bratislava 13, 647–661.
- Szekeres, J., Csányi, B., Kovács, K., Podani, J., 2011. A 2009. évi Rába-vizsgálat vízi makrogerinctelenekre vonatkozó eredményei II. *Ökológiai állapotértékelés. Acta Biologica Debrecina Supplementum Oecologica Hungarica* 26, 191–201.
- Tuomisto, H., 2010a. A diversity of beta diversities: straightening up a concept gone awry. Part 1: defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33, 2–22.
- Tuomisto, H., 2010b. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography* 33, 23–45.
- Ulrich, W., Almeida-Neto, M., 2012. On the meanings of nestedness: back to the basics. *Ecography* 35, 865–871.
- Ulrich, W., Almeida-Neto, M., Gotelli, N.J., 2009. A consumer's guide to nestedness analysis. *Oikos* 118, 3–17.
- Ulrich, W., Gotelli, N., 2010. Null model analysis of species associations using abundance data. *Ecology* 91, 3384–3397.
- Veech, J.A., Summerville, K.S., Crist, T.O., Gering, J.C., 2002. The additive partitioning of species diversity: recent revival of an old idea. *Oikos* 99, 3–9.
- Vellend, M., 2001. Do commonly used indices of β -diversity measure species turnover? *Journal of Vegetation Science* 12, 545–552.
- Whittaker, R.H., 1960. Vegetation of the Siskiyou mountains, Oregon and California. *Ecological Monographs* 30, 279–338.
- Wilson, J.B., 2012. Species presence/absence sometimes represents a plant community as well as species abundances do, or better. *Journal of Vegetation Science* 23, 1013–1023.
- Wilson, M.V., Mohler, C.L., 1983. Measuring compositional change along gradients. *Vegetatio* 54, 129–141.
- Wolda, H., 1983. Diversity, diversity indices and tropical cockroaches. *Oecologia* 58, 290–298.