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

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

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

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

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

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

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

$$^{a}D_{rel(jk)} = \frac{\left| \sum_{i=1}^{n} x_{ij} - \sum_{i=1}^{n} x_{ik} \right|}{T_{jk}} \tag{3}$$

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:

$$^{a}A_{rel(jk)} = \frac{T_{jk} - \left| \sum_{i=1}^{n} x_{ij} - \sum_{i=1}^{n} x_{ik} \right|}{T_{jk}} \tag{4}$$

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

$$^{a}R_{rel(jk)} = \frac{\sum_{i=1}^{n} |x_{ij} - x_{ik}| - \left| \sum_{i=1}^{n} x_{ij} - \sum_{i=1}^{n} x_{ik} \right|}{T_{jk}} \tag{5}$$

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