ECOGRAPHY

Research Article

Mathematically and biologically consistent framework for presence–absence pairwise indices

Arnošt L. Šizling[1], **Petr Keil**[2], **Even Tjørve**[3], **Kathleen M. C. Tjørve**[3], **Jakub D. Žárský**[4] and **David Storch**[1,4]

[1]Center for Theoretical Study, Charles University and Academy of Science of the Czech Republic, Prague, Czech Republic
[2]Department of Spatial Sciences, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Praha – Suchdol, Czech Republic
[3]Inland Norway University of Applied Sciences, Lillehammer, Norway
[4]Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic

Correspondence: Arnošt L. Šizling (sizling@cts.cuni.cz)

Ecography
2025: e07996
doi: 10.1002/ecog.07996
Subject Editor: Miguel Araújo
Editor-in-Chief: Miguel Araújo
Accepted 20 October 2025

Many indices based on presence–absence data that compare two communities have been proposed, with the aim to characterize community similarity, species turnover or beta-diversity, as well as other phenomena like community nestedness. These indices are often mathematically convertible to each other and are thus equivalent in terms of their information content. Based on this information equivalence, we classified all the indices to a few families, showing that only three families reflect ecologically relevant and directly interpretable phenomena, namely species turnover (family of Jaccard index that also includes Sørensen index of similarity), nestedness (the family of indices which compare species overlap with species richness of the species-poor community), and the uniformity of species richness (comparing species richness of the two communities). Importantly, our analysis shows that any attempt to partition indices, including Baselga's approach to partition turnover and nestedness (i.e. to control an index for an effect of a different phenomenon), leads either to an index belonging to one of the three abovementioned families, or produces indices that do not measure any ecologically relevant phenomenon. We provide guidance on how to apply pairwise indices to make proper inference about ecological phenomena.

Keywords: Baselga, beta diversity, interdependence, nestedness, index partitioning, community similarity, information content

Introduction

Differences in species composition between communities in space and time have been quantified by various indices, with a heated debate about which are better for different purposes (Brualdi and Sanderson 1999, Gotelli 2000, Vellend 2001, Koleff and Gaston 2002, Gaston et al. 2007, Jost 2007, Jurasinski et al. 2009, Ulrich et al.

NORDIC SOCIETY OIKOS
www.ecography.org

2009, 2017, Tuomisto 2010b, Baselga 2010, 2012, Podani and Schmera 2011, Almeida-Neto et al. 2012, Chao et al. 2012, Carvalho et al. 2012, Newbold et al. 2016, Matthews et al. 2019, Schmera et al. 2020). Substantial work has been done on abundance-based indices, and their inclusion in the framework of Hill numbers (Hill 1973, Jost 2007, Tuomisto 2010b, Chao et al. 2012, 2014a, 2014b, Chiu et al. 2014, Barwell et al. 2015). However, accurate data on abundances are still more difficult to obtain and are less common than simple presences and absences (Keil et al. 2021), particularly at large 'macroecological' spatial scales. Furthermore, abundance-based indices downweight rare species, which are often of conservation interest. Because of this, pairwise indices based on presences and absences, rather than abundances, continue to dominate global analyses of biodiversity (Dornelas et al. 2014, Blowes et al. 2019, Xu et al. 2023).

A particularly intensive debate has concerned the links between indices of species turnover and nestedness, and how to control them for the effect of each other and for species richness differences. This debate can be traced back to Simpson (1943), whose invention of the Simpson similarity (β'_{Sim}, Table 1) resolved the undesirable interdependence between the Jaccard index (*J*, Jaccard 1912, Table 1) and the difference in species richness of communities. Since then, a number of indices have been introduced that do not provide more information than Jaccard and Simpson indices (e.g. $\beta_{\text{sØr}}$, Sørensen 1948, and β_{jtu}, Baselga 2012, Table 1), or which are mathematically equivalent to earlier indices (e.g. $\beta_{\mathrm{t2}}=\beta_{\mathrm{SR}}=\beta_{\mathrm{HK}}$, defined by Wilson and Shmida 1984, Schluter and Ricklefs 1993, Harte and Kinzig 1997, respectively, Table 1, Supporting information). Yet, we lack a general framework for the relationships between indices, which would apply to any index already invented or introduced in the future.

Despite disagreements and redundancies, there has been progress. It is to Baselga's (2010) credit that the field has refocused from the classification of indices to their meaning. Chao et al. (2012) pointed out that the dependence of index formulas is more important than the dependence (or independence) of index values. Lennon et al. (2001), Tuomisto (2010b) and Newbold et al. (2016) showed that the Simpson index is in fact Sørensen similarity relativized to the contrast in species richness. Koleff et al. (2003) and Legendre and De Cáceres (2013) pointed out that Simpson and Jaccard represent two different groups of indices, which were labelled as broad-sense and narrow-sense turnover. This classification resonates with Almeida-Neto et al. (2008) and Šizling et al. (2009) who realized that the Simpson similarity (β'_{Sim}, Table 1) quantifies the phenomenon of nestedness, while the Jaccard index quantifies species turnover. Lastly and importantly, Ulrich et al. (2017, 2018) pioneered null models that disentangle mathematical and biological drivers of the dependence between various indices and species richness.

The realization that pairwise indices are simultaneously affected by different phenomena (e.g. by proportion of shared species as well as species richness difference) has recently led to an effort to control one effect from another by means of additive partitioning. Baselga (2010, 2012), Podani and Schmera (2011) and Schmera et al. (2020) presented two ways to mathematically partition the indices. Baselga (2012) based his reasoning on the observation that some re-arrangements of communities affect both turnover and nestedness simultaneously, and the respective indices are thus 'related' (sensu Chao et al. 2012), which means that their values are typically correlated (Box 1). To address this dependence, Baselga (2012) defined the 'nestedness-resultant component' (β_{sne}), which is $\beta_{\mathrm{sne}} \overset{\mathrm{def}}{=} \beta_{\mathrm{t2}} - \beta_{\mathrm{Sim}}$, where β_{t2} is Sørensen dissimilarity and β_{sim} is Simpson dissimilarity (definitions in Table 1). The subtraction supposedly removes the effect of the turnover component represented by β_{sim} from the total dissimilarity represented by β_{t2} (Baselga's (2012) interpretation), and what remains is considered to be the 'nestedness resultant component'. In contrast, Podani and Schmera (2011) see nestedness as antithetic to replacement (their synonym for turnover) and so indices of nestedness (β_{nps}, Podani and Schmera 2011) and replacement (β_{rps}, Podani and Schmera 2011) (Table 1 for definitions) together sum to one. Their index of nestedness is thus $\beta_{\mathrm{nps}} \overset{\mathrm{def}}{=} 1 - \beta_{\mathrm{rps}}$.

We argue that the abovementioned approaches have several problems which has led to a situation where researchers are typically not sure which indices can be used in different occasions and what the indices actually measure. To sort out the confusion, we present a unified framework for incidence-based indices that addresses these issues. In our framework, the indices are tools to capture ecological phenomena, and we show that these phenomena are already mutually dependent by their very nature, regardless of the indices that quantify them. This makes statistically independent or unrelated indices impossible – they cannot be 'purified' from the effects of other phenomena.

Unlike Baselga (2010) or Chao et al. (2012) we are primarily not focused on the statistical dependence of indices nor 'relatedness' of the formulas of the indices that constrain their possible values (as defined by Chao et al. 2012). Instead, we advocate for accounting for the dependence of indices in terms of their information content (which we call i-dependence, in contrast to statistical dependence, s-dependence), i.e. those that can be directly mathematically derived from each other (one index can be thus determined by the other and vice versa, see Box 1). We show that only two i-independent indices, together with species richness, carry enough information to compute any (even not yet introduced) index, and we provide a user-friendly tool to perform this conversion. We also develop guidelines to attribute the indices to different phenomena. Finally, we argue that all the mathematically and ecologically meaningful indices have already been described by Jaccard (1912), Simpson (1943) and others in the early 20th century (Tjørve et al. 2018).

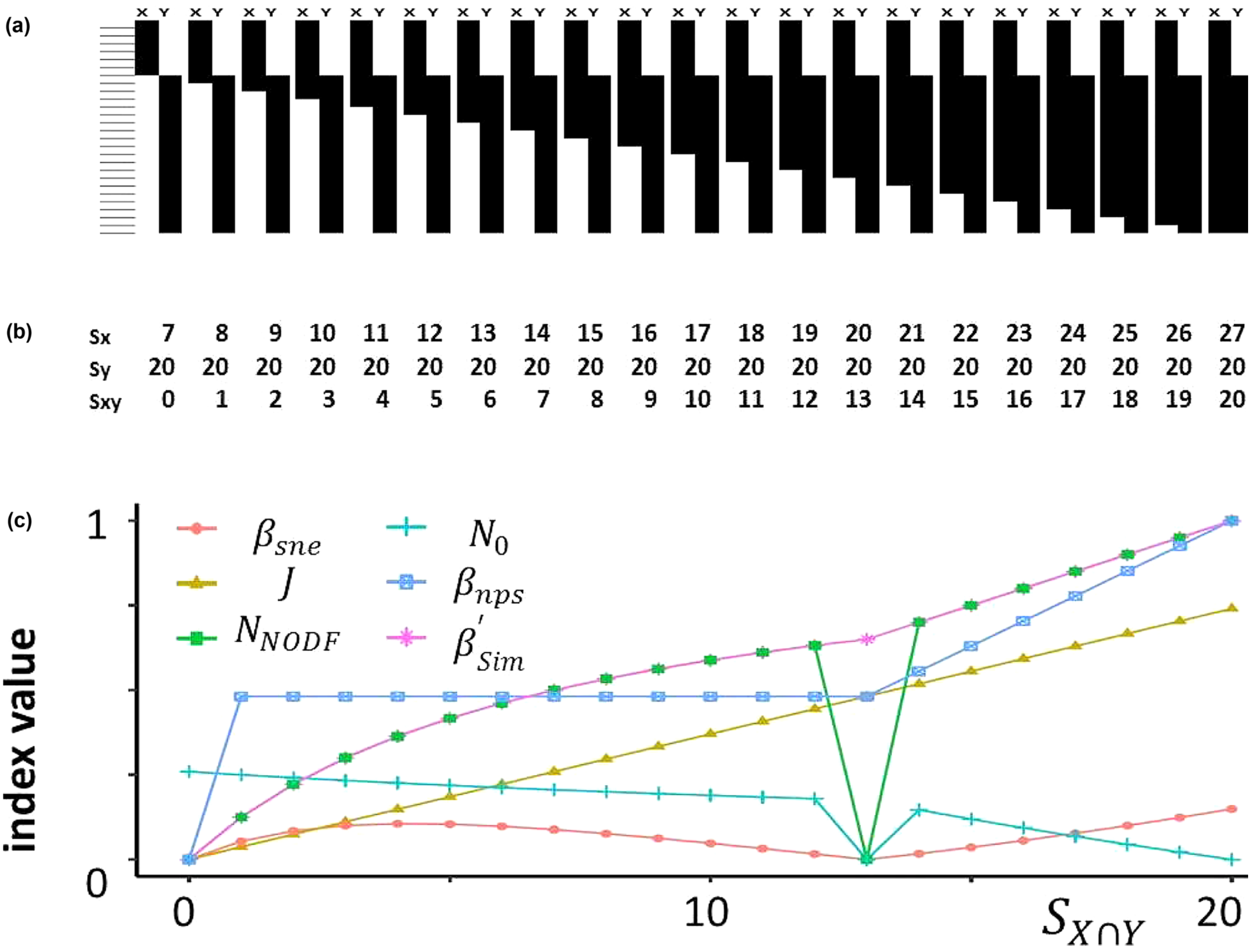
Table 1. The frequently used pairwise indices. S_X and S_Y are species richness of the communities to compare, and $S_{X\cap Y}$ is the number of shared species between them. For index definitions using the a, b, c notation ($a = S_{X\cap Y}$, $b = S_X - a$, $c = S_Y - a$) see the Supporting information. The labelling follows Gaston et al. (2007) and the symbol '$\overset{\text{def}}{=}$' represents definitions. For a complete list of indices and their mutual relationships see the Supporting information. Asterix (*) labels indices that capture none of the defined phenomena. Numbers in brackets '[]' refer to problems from the section 'Problems with earlier concepts'.

Definition	Name and source	Phenomenon
$J \overset{\text{def}}{=} \frac{S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	Jaccard index, Jaccard similarity, Jaccard 1912	Turnover (a reversed measure)
$\beta_G \overset{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}} = 1 - J$	Jaccard dissimilarity, Gaston et al. 2001, labeled as in Gaston et al. 2007	Turnover
$\beta_{\text{Sør}} \overset{\text{def}}{=} \frac{2S_{X\cap Y}}{S_X + S_Y} = 2\frac{J}{1+J}$	Sørensen similarity, Sørensen 1948	Turnover (a reversed measure)
$\beta_{t2} \overset{\text{def}}{=} \frac{S_X + S_Y - 2S_{X\cap Y}}{S_X + S_Y} = 1 - \beta_{\text{Sør}}$	Wilson and Shmida 1984, for two sites, labelled as in Gaston et al. 2007, Baselga 2010 calls this index Sørensen dissimilarity	Turnover
$\beta_{\text{Sim}} \overset{\text{def}}{=} \frac{\min(S_X, S_Y) - S_{X\cap Y}}{\min(S_X, S_Y)}$	Simpson 1943, Koleff et. al. 2003 Turnover component of β_{t2}, Baselga 2012	Nestedness (a reversed measure)
$\beta'_{\text{Sim}} \overset{\text{def}}{=} \frac{S_{X\cap Y}}{\min(S_X, S_Y)} = 1 - \beta_{\text{Sim}}$	Simpson similarity, Simpson nestedness, Simpson 1943, Lennon et al. 2001	Nestedness
$\beta_{\text{jtu}} \overset{\text{def}}{=} \frac{2\min(S_X, S_Y) - 2S_{X\cap Y}}{2\min(S_X, S_Y) - S_{X\cap Y}} =$ $= 2\beta_{\text{Sim}} / (1 + \beta_{\text{Sim}})$	Turnover component of Jaccard dissimilarity, Baselga 2012	Nestedness (a reversed measure)
$\beta_{\text{gl}} \overset{\text{def}}{=} 2\frac{\lvert S_X - S_Y \rvert}{S_X + S_Y}$	Gaston and Lennon index of species richness gradient, Lennon et al. 2001	Species richness uniformity (a reversed measure)
$R \overset{\text{def}}{=} \frac{\min(S_X, S_Y)}{\max(S_X, S_Y)} = \frac{2 - \beta_{\text{gl}}}{2 + \beta_{\text{gl}}}$	Newbold et al. 2016	Species richness uniformity
$\beta_{\text{sne}} \overset{\text{def}}{=} \beta_{t2} - \beta_{\text{Sim}}$	Nestedness resultant component of β_{t2}, Baselga 2012, labeled β_{NES} in Baselga 2010	(*) [1,2,3,4]
$\beta_{\text{jne}} \overset{\text{def}}{=} \beta_G - \beta_{\text{jtu}}$	Nestedness resultant component of Jaccard dissimilarity, Baselga 2012	(*) [1,2,3,4]
$\beta_{\text{rps}} \overset{\text{def}}{=} 2\frac{\min(S_X, S_Y) - S_{X\cap Y}}{S_X + S_Y - S_{X\cap Y}}$	(one-for-one) replacement, labeled as R_{rel} in Podani and Schmera 2011	(*) [1,2]
$\beta_{nps} \overset{\text{def}}{=} \frac{S_{X\cap Y} + \lvert S_X - S_Y \rvert}{S_X + S_Y - S_{X\cap Y}}$ if $S_{X\cap Y} \neq 0$ else $\beta_{\text{nps}} \overset{\text{def}}{=} 0$.	Nestedness, labeled as N_{rel} in Podani and Schmera 2011	(*) [1,2]
$N_{\text{NODF}} \overset{\text{def}}{=} \beta'_{\text{Sim}}$ if $S_X \neq S_Y$, else $N_{\text{NODF}} = 0$	Nestedness by Almeida-Neto et al. 2008	(*) [1]
$N_0 \overset{\text{def}}{=} \frac{\min(S_X, S_Y) - S_{X\cap Y}}{S_X + S_Y}$ if $S_X \neq S_Y$ else $N_0 \overset{\text{def}}{=} 0$.	Nestedness by Patterson and Atmar 1986	(*) [1]

Problems with earlier concepts

Our basic assumption is that an index should consistently and unambiguously reflect the intensity of an ecological phenomenon (Ulrich et al. 2017, 2018). We have identified four key problems that introduce inconsistencies between formal definitions of indices and the phenomena they are intended to measure. Specifically:

Problem 1. Some commonly used indices do not satisfy the "requirement of continuity" (Neumann and Morgenstern 1953). This requirement, presented for ecologists by Podani and Schmera (2012), states that a small re-arrangement of



(b)																					
Sx	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Sy	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
Sxy	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Figure 1. The behavior of various indices responding to a gradual change of community arrangements. The top plot (a) shows a sequence of 21 pairs of communities X and Y, with a continuous re-arrangement of communities, increasing similarity in species composition and nestedness from left to right, and constant total species richness. Columns show the communities, lines species, and blackened cells presences of species within communities. The first two lines in (b) show numbers of species in the left (Sx) and right (Sy) communities for each situation along the gradient. The third line in (b) shows $S_{X\cap Y}$ labeled as Sxy. The most bottom plot (c) shows the values of β_{sne} (red circles), J (olive triangles), N_{NODF} (green squares), N_0 (green-blue crosses), β_{nps} (blue squares), and β'_{Sim} (purple asterisks). While J and β'_{Sim} respond strictly monotonically to the gradual change of community rearrangement, the others reveal discontinuity (problem 1) and the index of nestedness β_{nps} does not distinguish different arrangements (problem 2).

species communities should lead to only a small change in the index value and conversely, a small change in an index value should indicate only a small re-arrangement of the communities. This is important for two reasons: an index that changes considerably with a negligible re-arrangement of communities is 1) sensitive to errors in observed data, and 2) leads to misleading inference about the pattern and/or ecological phenomenon. For example, common indices of nestedness as N_0 (Patterson and Atmar 1986) and N_{NODF} (Almeida-Neto et al. 2008) have been defined as zero when species richness of one site (S_X) is equal to species richness of the other site (S_Y). This violates the requirement of continuity (Fig. 1). To satisfy the requirement when $S_X = S_Y$ (S_X and S_Y stand for the species richness of respective communities), a value of nestedness should instead fall as close as possible to the nestedness values for $S_X = S_Y - 1$ and $S_X = S_Y + 1$.

The problematic claim of zero nestedness when $S_X = S_Y$ also affects the internal consistency of Baselga's (2010, 2012) framework. Baselga (2012) defines nestedness conceptually as "a pattern characterised by the poorest site being a strict subset of the richest site". At the same time, Baselga's (2012) theory proposes that "in the absence of nestedness, the dissimilarity index [β_G in Table 1] is equal to its turnover component [β_{jtu} in Table 1]". However, this equality also holds for two identical communities (Supporting information). Therefore, according to the conceptual definition, two nearly identical, fully nested communities should have maximum nestedness, and according to the proposition, two exactly identical communities should have minimum nestedness. This logic violates the requirement of continuity.

Problem 2. The relationships between indices and ecological phenomena that the indices aim to capture (e.g.

Box 1. Key concept: the informational dependence of indices.

Information content of two indices is identical when each value of the first index uniquely determines the value of the second index, and each value of the second index uniquely determines the value of the first index (Orlitsky 2003). We thus call two variables informationally dependent (i-dependent) when they are strictly monotonic transformations of each other (Fig. 2a). A strict monotony of two variables implies equality of their informational values, and when two indices have equal information content, they cannot capture different phenomena.

In contrast, the kind of dependence that is used in ecological literature is almost always of statistical nature, and we call it s-dependence. While i-dependence refers to a mathematical formula, s-dependence refers to data values (i.e. a conditional probability of measuring a particular value of one variable, given a value of another variable) and is thus linked to correlation.

Chao et al. (2012) distinguished statistical dependence of index values from a dependence of mathematical formulas of indices, and uses the term relatedness for this dependence of formulas. Relatedness refers to the situation where the minimum or maximum of one index is affected by the variation of another index. Relatedness, however, is one of the mechanisms that induces s-dependence, since a change in the minimum and maximum possible value of an index has an effect on its probability density, which is the definition of statistical dependence (Fig. 2). Importantly, relatedness does not identify whether two or more indices carry identical information. For example, J and β'_{Sim} are i-independent, while their values are s-dependent (their values are correlated) because their definitions are related sensu Chao et al. (2012) (Fig. 2b).

Tuomisto (2012) defines the independence of indices based on different defining formulas. According to Tuomisto (2012), two different formulas , e.g. $I_1 = (A - B)/B$ and $I_2 = A/B$, are conceptually independent of each other and quantify two different phenomena. The reason is that I_1 reads as 'relative difference' but I_2 reads as 'proportion'. However, I_1 and I_2 are i-dependent and define different measures of the same phenomenon, because $I_1 = (I_2 - 1)$, and whatever can be inferred from I_1 can also be inferred from I_2.

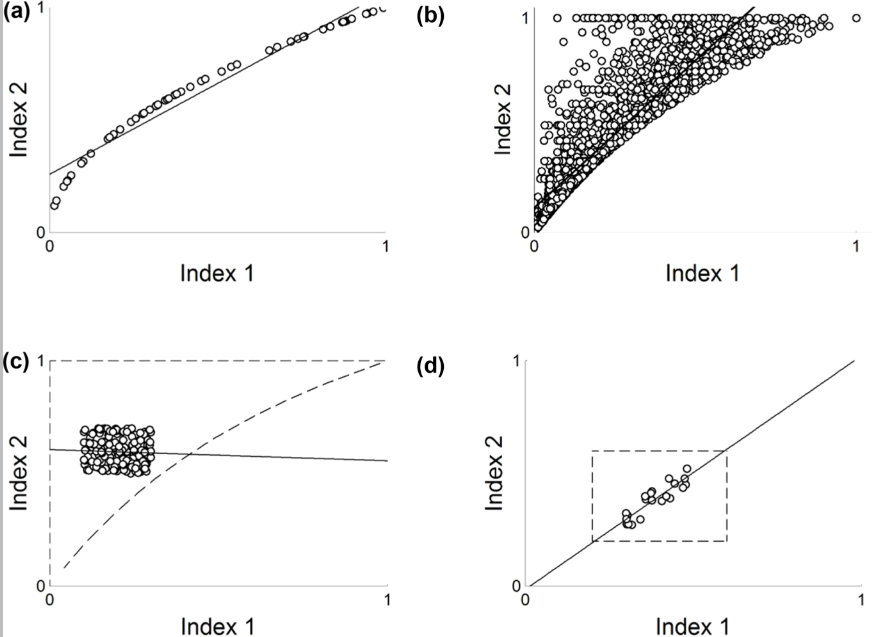


Figure 2. Four examples of different relationships between hypothetical index formulas, and index values (empty rings in a, b, c, d). Full lines stands for regression lines. Empty circles in (a, b) randomly cover the whole space of possible values. Dashed lines in (b, c) show ranges of possible values given by the index formulas, and empty circles show observed index values for particular communities at particular sites. In (a), index formulas are i-dependent and related (sensu Chao et al. 2012), index values are s-dependent; in (b) formulas are i-independent but related, and values are s-dependent; in (c) formulas are i-independent but related, and values are s-independent; and in (d) formulas are i-independent and unrelated, but observed values are s-dependent.

nestendess) have been typically derived using unidirectional implication: a particular state of the phenomenon (e.g. high or low nestedness) affects the value of an index, so it is assumed that the index measures the phenomenon. However, if an index is to be useful, the opposite implication should also hold: a value of the index should always correspond to a unique state of the phenomenon (e.g. only to high or only to low nestedness, not both). One index value should not be therefore assigned to multiple different states of the phenomenon (the problem of 'converse logic', Audi 1999). An example of the ambiguous unidirectional implication is used by Podani and Schmera (2011), who argue that "because species replacement and nestedness reflect contrasting ecological phenomena […] it is meaningful to express nestedness [β_{nps}] with the effect of species replacement [β_{rps}] completely removed from [its] maximum". However, in Fig. 1 we show that cases with 1–13 shared species have equal β_{nps}, although they apparently have different nestedness, conceptually predefined as the extent in which species of smaller community occur also in the larger community (Patterson 1984, Baselga 2010, Podani and Schmera 2011). Similarly, the fact that the absence of nestedness leads to $\beta_{\mathrm{s\O r}}=\beta_{\mathrm{sim}}$ (Baselga 2010, 2012) does not mean that $\beta_{\mathrm{s\O r}}=\beta_{\mathrm{sim}}$ implies no nestedness (Problem 1). This problem of the incorrect implication was mentioned by several authors (Ulrich et al. 2017, 2018, Schmera et al. 2020) but has not been appreciated in practice.

Problem 3. The mathematical operation of subtraction of indices does not remove one effect from another. Baselga (2010, 2012) has suggested that subtraction of an index of turnover component from an index of dissimilarity leads to the nestedness-resultant component, irrespective of the indices used. Baselga (2012) applies the reasoning about nestedness-resultant components equally to β_{t} and β_{G}, so that the nestedness-resultant component of Sørensen dissimilarity (β_{sne}, Table 1) is $\beta_{\mathrm{t}}-\beta_{\mathrm{Sim}}$, and the nestedness-resultant component of Jaccard dissimilarity (β_{jne}, Table 1) is $\beta_{\mathrm{G}}-\beta_{\mathrm{jtu}}$ (β_{Sim}, and β_{jtu} are the 'turnover components' of Sørensen and Jaccard dissimilarities, respectively; Table 1). However, the subtraction can remove the effects only for β_{t} or β_{G} (or none of them), but not for both, and it is unknown which of β_{sne} and β_{jne} would be the 'true' nesteness resultant component. The reasoning is as follows: β_{t} is i-dependent on β_{G} (one can be calculated from the other) and, similarly, β_{jtu} is i-dependent on β_{Sim} (Table 1, Supporting information). If the subtraction produced nestedness-resultant components in both cases, then the subtraction would have to account for the same effect in both cases, and thus β_{sne} and β_{jne} should also be i-dependent. However, it is not the case (Baselga 2012). The problem is that the minus operator in $\beta_{\mathrm{t}}-\beta_{\mathrm{Sim}}\ (\overset{\mathrm{def}}{=}\beta_{\mathrm{sne}})$ has a different meaning than the minus operator in $\beta_{\mathrm{G}}-\beta_{\mathrm{jtu}}\ (\overset{\mathrm{def}}{=}\beta_{\mathrm{jne}})$. Similarly the subtraction in $x-y$ does not have the same meaning as the subtraction in $\log x - \log y$ (since $\log(x) - \log(y) = \log(x/y)$), even though x and $\log x$ are i-dependent. As a result, the logic that subtraction eliminates an effect cannot be applied to both β_{sne} and β_{jne}, and only one of them (if any) can capture the nestedness-resultant component, and we don't know which one.

Problem 4. Although pairwise indices are dimensionless quantities and can take values within the same range (e.g. 0 and 1), they are not universally comparable (Schmera and Podani 2011). For example, the framework introduced in Baselga (2010) subtracts and compares Sørensen dissimilarity, β_{t2}, and Simpson dissimilarity β_{Sim} (Table 1). This has been challenged by Schmera and Podani (2011), who proposed to sum (or subtract) only indices with the same denominator. The problem of commensurability of β_{t2}, and β_{Sim} stems from the fact that they measure two different phenomena, and two measures of different phenomena cannot be subtracted from each other. Simply, two measures with different units cannot be subtracted from each other.

Theory

To overcome the four problems listed above, we explore which indices, or their combinations, provide new and interpretable information about species composition of communities, and which are redundant. The fundamental proposition that justifies this approach is that 'indices that carry identical information cannot be used to quantify different ecological phenomena, while those carrying different information quantify different phenomena'. The new focus on information independence has several key implications. Firstly, we will show that indices can be organized into families of mutually i-dependent measures. We will also demonstrate that a minimal set of two i-independent dimensionless indices combined with the species richness of one community provides all the information necessary to compute any other presence–absence index, including those yet to be invented. We will then show that this information-based framework is key to ecological interpretation of the indices.

Classification of indices based on their information content

Definition of information dependence

In order to evaluate the information provided by an index, we have to treat the definitions of the indices as equations to be solved (Box 2). When we add an equation to a set of n equations, and if this new set of equations provides an identical solution to the previous set, then the new equation is said to be i-dependent on the others and carries no extra information. If the solution of the $n + 1$ equations is a subset of (but not equal to) the solution of the n equations, then the new equation is i-independen*t* of the others and carries extra information. Note that this is a widely accepted mathematical definition of independence in a system of equations (Box 2) and that it follows from the theorem from information theory that only the variables that are uniquely mapped to each other have equal information content (Orlitsky 2003).

We use this mathematical definition of independent equations and define the indices that are mutually i-independent if the

Box 2. How many indices do we need?

For a given system, a multitude of indices can be calculated. What is the minimum number of indices that characterize a given system fully? Here we show that three i-independent indices (including species richness) are sufficient for full characterization of the system, so that any other index can be reconstructed using this information. This allows for two statements: 1) all the focal indices listed in the Supporting information are i-dependent on the pair of Jaccard index, J (Table 1) and Simpson nestedness, β'_{Sim} (Table 1); and 2) all information concerning the difference between two communities is captured by J, β'_{Sim}, and an index that is i-dependent on species richness (including species richness itself). Here we show why these statements hold.

All the focal indices can be defined as a ratio of two linear functions (Supporting information):

$$I \overset{\text{def}}{=} \frac{k_1 S_{X\cap Y} + k_2 S_X + k_3 S_Y}{l_1 S_{X\cap Y} + l_2 S_X + l_3 S_Y}, \left(k_{1-3}, l_{1-3} \in \boldsymbol{R}\right), |l_2| + |l_3| > 0, S_X \ge S_Y > 0. \qquad \text{(B2.1)}$$

For J and β'_{Sim}, $\{k_{1-3}, l_{1-3}\} = \{1,0,0,-1,1,1\}$ and $\{1,0,0,0,0,1\}$, respectively. Following the logic: $I = \frac{a}{b} \rightarrow bI - a = 0$, Eq. B2.1 can be converted into the linear equation

$$S_X\left(Il_2 - k_2\right) + S_Y\left(Il_3 - k_3\right) + S_{X\cap Y}\left(Il_1 - k_1\right) = 0, \qquad \text{(B2.2)}$$

where S_X, S_Y, $S_{X\cap Y}$ are the unknowns and the I is a particular value of an index. Eq. B2.2 turns into

$$S_X J + S_Y J - S_{X\cap Y}\left(J + 1\right) = 0,\ S_Y \beta'_{Sim} - S_{X\cap Y} = 0 \text{ and } S_X R - S_Y = 0 \qquad \text{(B2.3)}$$

for J, β'_{Sim} and R, respectively. The two equations for two indices (Eq. B2.3), and the three unknown variables (S_X, S_Y, $S_{X\cap Y}$), do not provide a unique solution, and have zero at the right side. If the third equation had zero on the right side, the system would either provide multiple solutions or the only solution would be zero (S_X, S_Y, $S_{X\cap Y} = 0$, 0, 0). We therefore need an independent equation with a nonzero right side to get unique information on S_X, S_Y, $S_{X\cap Y}$. This equation is

$$S_X + S_Y = 2\overline{S}, \qquad \text{(B2.4)}$$

where $\overline{S} \neq 0$ is the expected alpha diversity (for the proof that Eq. B2.4 is i-independent of Eq. B2.3, see the Supporting information). The three equations (any two equations from Eq. B2.3, and Eq. B2.4) determine S_X, S_Y, $S_{X\cap Y}$ uniquely, and thus no other index (even if its definition does not follow Eq. B2.1) can further specify the solution. Moreover, if we focus on the indices which can be expressed using Eq. B2.1, we can compute the value of any other index from the values of J and β'_{Sim} or R as

$$I = \frac{J(\beta'_{\text{Sim}}(k_1 + k_2) - k_2 + k_3) + \beta'_{\text{Sim}} k_2}{J(\beta'_{\text{Sim}}(l_1 + l_2) - l_2 + l_3) + \beta'_{\text{Sim}} l_2} \text{ for } J \neq 0, \text{ else } I = \frac{k_2 + k_3 R}{l_2 + l_3 R} \qquad \text{(B2.5)}$$

where k_i, l_i define the new index I (for the proof see the Supporting information).

equations of the indices are mutually independent (Box 2). For example, β'_{Sim} (Simpson nestedness) and J (Jaccard similarity) (Table 1) are i-independent because there is no way to convert β'_{Sim} to J. This is easy to see from Fig. 2b where more than one value of β'_{Sim} is attributed to each value of J (except for $J = 0$). Unlike the statistical concept of dependence (hereafter statistical dependence or s-dependence, Box 1), relatedness (Chao et al. 2012) and i-dependence comprise the links between mathematical formulas that constrain the possible values of observed indices. However, there is a vital difference between the two. While relatedness does not imply that one index can be converted into a related index, informational dependence requires a deterministic, one-to-one mapping between indices.

The distinction between relatedness and i-dependence may seem minor, but it is of fundamental importance. Consider, for example, two distinct concepts in physics: the kinetic energy of a moving body ($E_k = 0.5\ mv^2$, where m is mass and v is speed) and its momentum ($p = mv$). Based on their similar mathematical forms and plot of E_k against p values, one might assume that these phenomena are tightly related, highly s-dependent, and therefore interchangeable. But these variables cannot be converted to each other, and they provide vital and complementary information about the physical system. Similarly, in ecology, two indices may be mutually related, or s-dependent (Fig. 2b) yet carry complementary, non-redundant information. For a complete understanding of relationships between ecological communities, it is therefore essential to use indices that, while potentially related, are not informationally dependent.

Families of indices

So far, we have introduced the concept of information interdependence between indices. Now, we will group them to families within which indices share equal information. We will then show how to convert indices within their respective family to each other (Box 3). Mathematical details of these steps are in repository (Supporting information), and here we summarize the results.

To distinguish families of i-dependent indices, we first examined their mutual bivariate relationships, i.e., we simply plotted the indices against each other, using simulated and real-world data (Fig. 3, Supporting information). Where we did not see a one-to-one correspondence (Fig. 2a), i.e., the line joining the points in the plot was neither strictly increasing nor strictly decreasing, the two plotted indices were considered i-independent. Where we found a one-to-one correspondence along a strictly increasing or decreasing curve, we had to prove the i-dependence by mathematical analysis (Supporting information). Following families have emerged:

1. **Jaccard family.** The largest family of indices is grouped around the Jaccard index (*J*, Jaccard 1912). It has nine

Box 3. How to unify indices from across the literature.

A large number of different indices are spread over the literature (Gaston et al. 2007, Baselga 2012, Schmera et al. 2020). However, if one wishes to do a meta-analysis and compare indices from different published sources, one needs to make the indices comparable, and to convert them to only one or two reference indices. Here we show how this can be done, based on the idea that an index is defined by an equation. Firstly, we need a publication that utilizes at least two i-independent indices, I_1, I_2, that are expressed as in Eq. B2.1 in Box 2. Then we write a system of three independent equations: two equations Eq. B2.2 in Box 2, each for one of the indices, and one scaling equation Eq. B2.4. If $\bar{S}$ is missing from the publication, we can put $\bar{S}=1$ without loss of generality. The reason is that the indices to be converted (Eq. B2.1 in Box 2) are i-independent of species richness (Supporting information) and therefore the exact value of $\bar{S}$ does not matter in the case. The third equation then is

$$S_X + S_Y = 2\,. \tag{B3.1}$$

The solution of the three equations (Gauss elimination method) is $S_{\mathrm{X\ comp}}$, $S_{\mathrm{Y\ comp}}$, and $S_{\mathrm{X\cap Y\ comp}}$. This solution is different from the solution based on the original data behind the published source, but we can get the original average values $S_{\mathrm{X\ orig}}$, $S_{\mathrm{Y\ orig}}$, $S_{\mathrm{X\cap Y\ orig}}$ by simple rescaling of the computed values ($S_{X\mathrm{orig}} = \bar{S}S_{X\mathrm{comp}}$ and so on), in case we know $\langle S\rangle$. From $S_{\mathrm{X\ comp}}$, $S_{\mathrm{Y\ comp}}$, $S_{\mathrm{X\cap Y\ comp}}$ we can compute any index that is i-independent of species richness, even for an index that cannot be expressed as Eq. B2.1 in Box 2 (e.g. $I = S_X S_{X\cap Y} / S_Y^2$).

Some special cases are the indices that originated from additive partitioning ($I_1 = I_\mathrm{a} - I_\mathrm{b}$) such as β_{sne}, β_{jne}, or β_{nps} (Table 1). These indices cannot be universally expressed as Eq. B2.1, and their equations are no longer linear. In the case of additive partitioning, we can compute I_b from two i-independent indices I_1 and I_2 using Eq. B3.2

$$I_b^2 D_{\mathrm{LL}} - I_b\left(I_1 D_{\mathrm{LL}} - D_{\mathrm{KL}} + D_{\mathrm{LK}}\right) + I_1 D_{\mathrm{LK}} - D_{\mathrm{KK}} = 0 \tag{B3.2}$$

where D_{XY} are determinants listed in the Supporting information. I_a then follows $I_1 + I_\mathrm{b}$. If I_a, I_2, or I_b, I_2, or I_a, I_b are i-independent then we can follow the above algorithm that uses linear equations and compute any even not yet invented index. However, as Eq. B3.2 may have two realistic solutions (both solutions are within minimum and maximum possible range of I_b), the partitioning often leads to the loss of information. For example, if Šizling et al. (2016) had only reported two partitioned indices (e.g. $\beta_{\mathrm{sne}} = 0.1$ and $\beta_{\mathrm{nps}} = 0.61$ for pre-agricultural landscape, and $\beta_{\mathrm{sne}} = 0.0$ and $\beta_{\mathrm{nps}} = 0.51$ for agricultural landscape), theywould not be sufficient to infer the homogenization of communities. The reason is that these changes of indices may be driven both by increasing or decreasing uniformity (R)(R could have changed from 0.74 to either 1 or 0.32, respectively, see the Supporting information for further details).

indices of (dis)similarity, including Sørensen ($\beta_{\text{SØr}}$, as defined in Gaston et al. 2007), β_{HK} (Harte and Kinzig 1997), β_{G} (Gaston et al. 2001), β_{SR} (Schluter and Ricklefs 1993), β_{t2} (Wilson and Shmida 1984) and β_{DICE} (Raup and Crick 1979). The Bray–Curtis index of dissimilarity (β_{BC}, Bray and Curtis 1957), belongs to the Jaccard family when computed from incidences. A classical index called, N_c (Wright and Reeves 1992) also belongs to the Jaccard family (Fig. 3, Supporting information) when standardized by $S_X + S_Y$ (Gotelli and McCabe 2002). Whittaker beta-diversity index, when applied to pairwise comparison, also belongs to this family (Tuomisto 2010b, Supporting information). An inspection of all indices in this family reveals that they all relate species overlap to the whole set of species, i.e. utilize all three species richness components S_X, S_Y, and $S_{X\cap Y}$.

2. **Simpson family.** The second large family consists of Simpson's beta (β_{Sim}, Simpson 1943), Simpson's nestedness (β'_{Sim}, Lennon et al. 2001) and β_{jtu} which was previously referred to as a turnover component (Baselga 2012). Two classical indices that are supposed to measure nestedness, the discrepancy (D, Brualdi and Sanderson 1999), and the standardized N_C by Wright and Reeves (1992), also belong to the Simpson family under certain circumstances. The pairwise index D does not belong to any family if we compare just two communities. However, if we increase the number of sites being compared, then D standardized as in Greve et al.

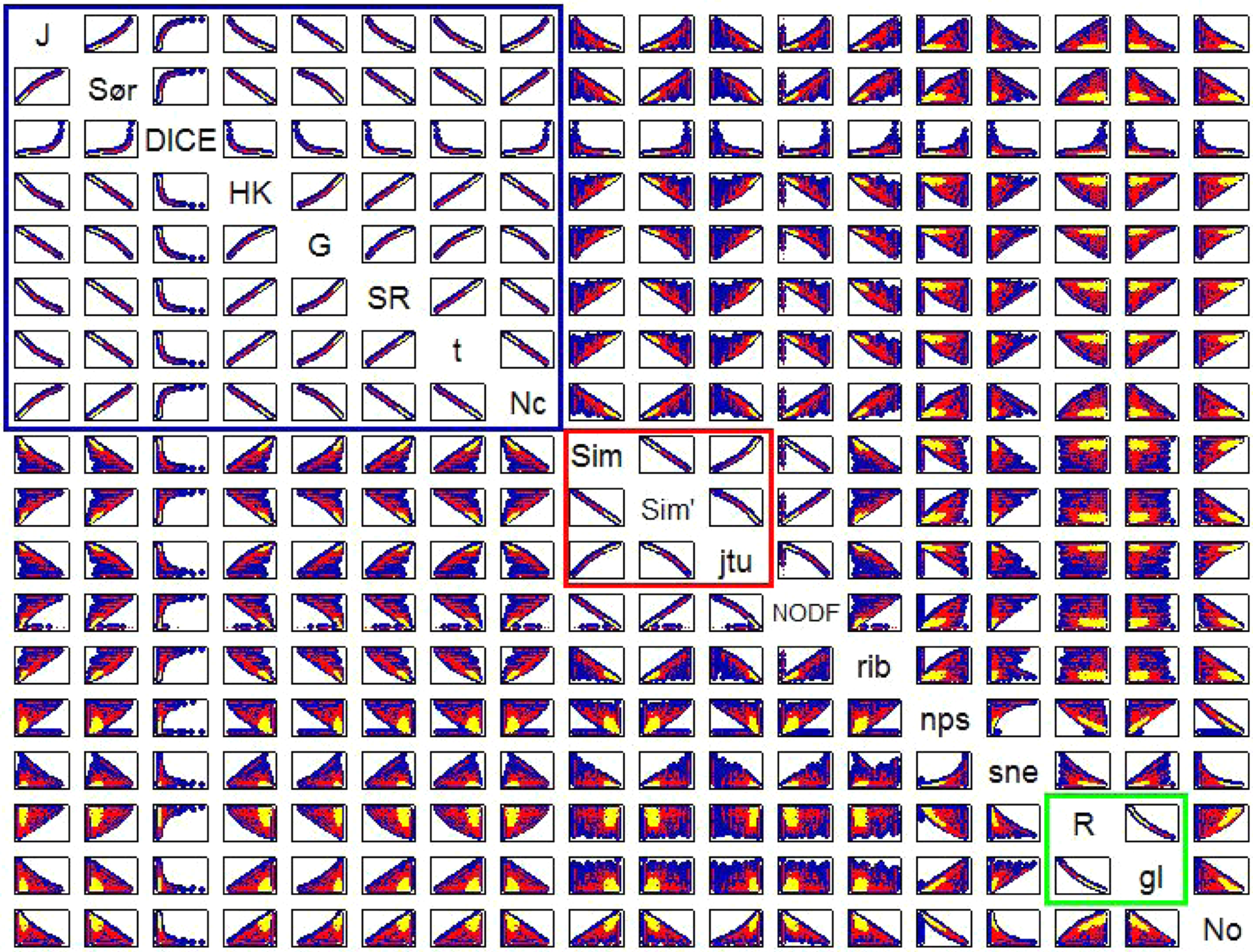


Figure 3. Relationships between the indices for pairs of simulated random (blue dots, n = 1352, see the Supporting information for details) and observed (red – plants, n = 946, yellow – Ice Shield microbes, n = 406) communities (Supporting information). The species richness values of communities in each pair are uncorrelated, i.e., numbers $S_{X\cap Y}$, $b = S_X - S_{X\cap Y}$ and $c = S_Y - S_{X\cap Y}$ vary between 0 and 1, they are drawn from a uniform distribution and are mutually s-independent. In N_0 and N_{NODF} (i.e., where the equality between S_X and S_Y affects the result) species richness is a random integer between 1 and 20 species. Where possible, the notation is adopted from Gaston et al. (2007). The notation in the plot is simplified. From up to down (the indices without a reference are referred in Table 1, the indices with a reference are defined in the Supporting information): J, $\beta_{\text{sØr}}$, β_{DICE} – Raup and Crick (1979), β_{HK} – Harte and Kinzig (1997), β_G, β_{SR} – Schluter and Ricklefs (1993), β_{t2}, N_C – Wright and Reeves (1992) standardized as in Gotelli and McCabe (2002), β_{Sim}, β'_{Sim}, β_{jtu}, N_{NODF}, β_{rib} – Ruggiero et al. (1998), β_{nps}, $\beta_{\text{sne}} = \beta_{\text{NES}}$, R, β_{gl}, N_0 – Patterson and Atmar (1986). Blue, red, yellow, and green rectangles delimit the Jaccard, Simpson and Gaston–Lennon families, respectively.

(2005) quickly converges to the mean value across all β_{Sim} ($D \cong \overline{\beta_{\mathrm{Sim}}}$, Supporting information) belonging to the Simpson family for n of sites > 10 (Supporting information). The pairwise N_{C} also belongs to the Simpson family if standardized as suggested by its authors (Wright and Reeves 1992, Supporting information), and if species richness is high (Supporting information). All indices within this family relate the species overlap to the richness of species-poor community, i.e. they utilize the information about $S_{\mathrm{X \cap Y}}$ and $\min(S_{\mathrm{X}}, S_{\mathrm{Y}})$ and ignore the information about the species-rich community.

3. **Gaston-Lennon family.** This family currently consists of two indices: an index called species richness gradient (β_{gl}) as defined by Lennon et al. (2001) (Table 1), and index R as defined by Newbold et al. (2016) (Table 1). Although β_{gl} is defined as the standardised difference of species richness and R is a simple ratio of minimum to maximum species richness, they can be converted into each other (Supporting information) and thus belong to the same family.
4. **Podani-Schmera family.** This family consists of two indices β_{rps} and β_{nps} (Table 1). Both the indices were introduced by Podani and Schmera (2011) and the authors called them replacement and nestedness, respectively.
5. **Other families.** There are several minor families, each consisting of a single index (Fig. 3): Ruggiero index of beta-diversity (β_{Rib}, Ruggiero et al. 1998), Baselga nestedness-resultant components of both Sørensen dissimilarity (β_{sne}, Baselga 2012, labelled as β_{nes} in Baselga 2010) and Jaccard dissimilarity (β_{jne}, Baselga 2012, not plotted); the classic index of nestedness, N_0 (computed for two sites, Patterson and Atmar 1986), and the N_{NODF} index of nestedness (computed for two sites, Almeida-Neto et al. 2008). The N_{NODF} would have belonged to the Simpson beta family if it did not violate the requirement of continuity (Fig. 3), which makes inferences from the index value problematic.

i-independent combinations of indices

Although we have delineated nine families of indices, it does not mean that there are nine i-independent indices. The reason is that three indices from any three families are i-dependent even if any single pair from this triplet is i-independent. Only two families are then i-independent with certainty, and the indices of the other families can be calculated from these two (Box 2). A clear case of mutual i-dependence of a triplet of indices is the combination of the Jaccard family, the Simpson family and the Gaston–Lennon family. A simple re-arrangement of Eq. B2.5 (where $R = I$ and $k_1, k_2, l_1, l_3 = 0$, $k_3, l_2 = 1$, Box 2), gives

$$\frac{J}{1+J} = \beta'_{\mathrm{Sim}} \frac{R}{1+R}. \tag{1}$$

As all three indices (J, β'_{Sim}, R) and their multiplicative components, $J/(1+J)$, β'_{Sim}, and $R/(1+R)$ increase with increasing 'similarity' of the communities, this result (Eq. 1) indicates that the similarity quantified by the Jaccard family increases with increasing 'similarities' quantified by the Simpson and Gaston–Lennon families.

The results (Eq. 1) can be generalized to all the considered indices that can be defined by Eq. B2.1 (Box 2), which includes all dimensionless indices listed in Table 1, (and the Supporting information), and even indices that have not yet been introduced. This does not include some indices that have been produced using additive partitioning (e.g. β_{sne}, Eq. B3.2). The algorithm in Box 3 provides a tool for converting indices from multiple studies into common terms, assuming that the authors published at least two i-independent presence–absence indices for their data (e.g. as in Xu et al. 2003).

Special cases: families of indices when communities are interrelated

Species communities are often interrelated due to similarity in habitats or due to dispersal. This may lead to two special cases in which the indices that would normally belong to different families appear as i-dependent.

First, let us assume an effect that limits variation of species richness between sites. In extreme, $S_{\mathrm{X}} = S_{\mathrm{Y}}$. In this case, $R = 1$ and Eq. 1 turns into

$$\beta'_{\mathrm{Sim}} = 2\frac{J}{1+J}. \tag{2}$$

In this case, the Jaccard and the Simpson families become a single family of mutually i-dependent indices (Supporting information).

Second, communities can be perfectly nested, i.e., the community poorer in species has no unique species. In this case, $S_{\mathrm{X \cap Y}} = \min(S_{\mathrm{X}}, S_{\mathrm{Y}})$ and thus $\beta'_{\mathrm{Sim}} = 1$. It follows from Eq. 1 that

$$J = R. \tag{3}$$

In this special case the Jaccard and Gaston–Lennon families appear no longer i-independent (Supporting information) and become one index family.

The apparent loss of i-independence in these special cases is trivial, but it illustrates an important point: when species richness of communities become similar, or communities approach complete nestedness, then the indices cluster along a strictly monotonic relationship and thus appear i-dependent (Supporting information).

Statistical non-equivalence of the indices

We have argued that the indices within a given family are equivalent, since they are i-dependent and provide the same information. However, they often scale non-linearly with each other. The values of two indices within one family can be seen as transformed values of each other (Fig. 3, Supporting information). This means that the values of an index and its nonlinear transformation have different frequency distributions, which can affect parametric statistical tests and their sensitivity – it decreases the correlation coefficient and s-dependence between otherwise i-dependent indices, making them statistically non-equivalent (Koleff and Gaston 2002, Lyashevska and Farnsworth 2012). This may crucially affect conclusions based on real-world data with measurement bias (Chao et al. 2005, 2006, Chao and Colwell 2017).

One option to deal with these issues would be to use an index with the most symmetric frequency distribution for the examined dataset. However, the commonly used indices represent a rather poor spectrum of transformations (Fig. 3). We argue that it is better to first pick any index within the family that well describes the phenomenon of interest, and transform it using an appropriate transformation (e.g. logit transformation), rather than pick and choose from the range of existing indices within given family with 'good' statistical properties. Alternatively, it is possible to test the index values against a null model, and then the precise distribution of the values is not an issue (Ulrich and Gotelli 2007, 2013, Chase et al. 2011).

From indices to ecological phenomena

So far, we have examined the formal relationship between the indices, without focusing on the ecological phenomena they quantify. To make a consistent link between the indices and the phenomena it is vital to proceed in two steps. First, we need an intuitive understanding of what are the distinct phenomena we want to measure, i.e. their conceptual predefinitions. Second, we refine the predefinitions in a formal way. To do so, we gradually rank various community arrangements along an axis of increasing 'intensity' of the focal phenomenon. The value of respective index has to strictly increase or decrease along this axis of the ranked arrangements. Gradual ranking ensures continuity (Problem 1), and the strict monotony is crucial for the bidirectional implication (Problem 2). A practical way to rank the arrangements is to define a finite set of community arrangements with inequalities between their intensities that follow from the intuitive understanding of the phenomenon (Fig. 4). These inequalities constrain the indices that quantify the phenomena. Specifically, we typically understand when the intensity of the phenomenon is at its minimum and maximum, and when it is lower and higher – e.g. what arrangement corresponds to maximum and minimum nestedness, and which arrangement represents higher and lower nestedness. Only the indices with the values that rank in the same way as the arrangements can quantify the phenomenon.

Step 1. Conceptual predefinitions of the diversity phenomena

We propose that most of the common indices in ecology have originated as measures of four phenomena. Ecologists can intuitively perceive these phenomena in the field even without the knowledge of any mathematically defined indices, similarly as one can compare the speed of moving objects without measuring the speed. Imagine an ecologist roaming the landscape, leaving one site, and approaching another. They can directly observe 1) a change in species composition between communities, regardless of the change in species richness (species turnover); 2) a change

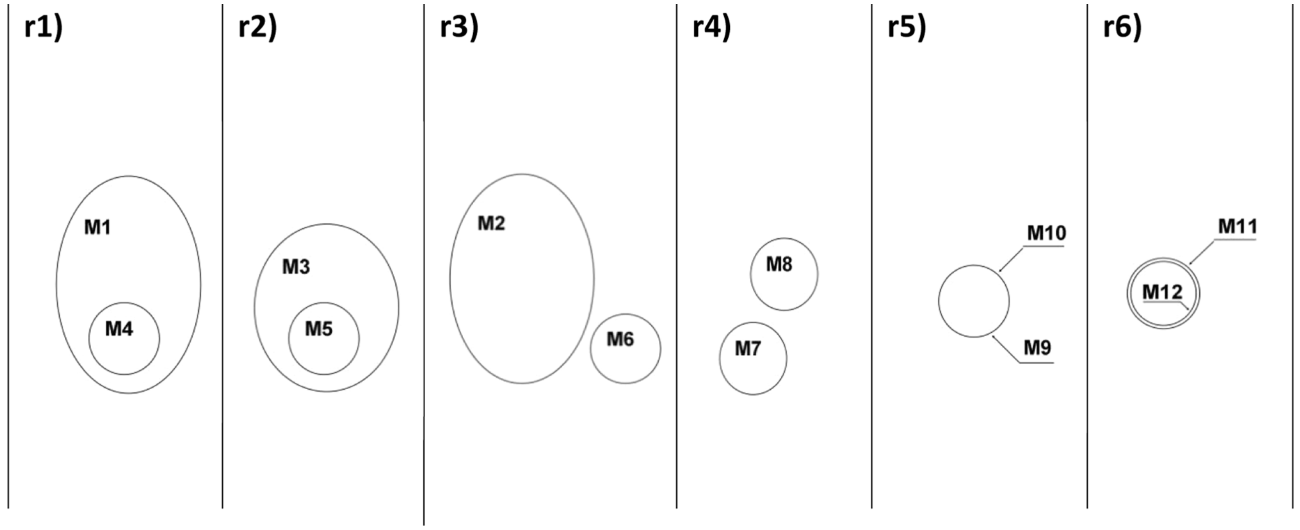


Figure 4. Constraints imposed by the three diversity phenomena (excluding beta-diversity sensu stricto) on their respective indices. Various arrangements of two sets (lists of species) in the Venn diagrams define differences between the phenomena. The sizes of the sets follow M1 = M2 > M3 > M4 = M5 = M6 = M7 = M8 = M9 = M10 = M11 > M12. The sets M9 and M10 are identical. The constraints that define the phenomena are as follows: Nestedness: Min = N[r4] = N[r3] < N[r1] = N[r2] = N[r6] = N[r5] = Max, Turnover: Min = T[r5] < T[r6] < T[r2] < T[r1] < T[r4] = T[r3] = Max, and Richness Uniformity: Min < U[r3] = U[r1] < U[r2] < U[r6] < U[r4] = U[r5] = Max. The 'Min' and 'Max' label the extreme values of the indices (usually Min = 0 and Max = 1) and the letters in brackets refer to the arrangements.

in species richness between communities, regardless of species replacement (species richness uniformity); 3) presence of species of the species-poor community in the species-rich one (nestednes of communities); and 4) a contrast between the average richness of a single community and the total observed richness (Whittaker's (1960) beta-diversity sensu stricto).

The interrelation of the four phenomena is their inevitable property and must be reflected by the respective indices. The reason is that 1) complete absence of turnover implies that all communities are identical and thus have maximum uniformity in species richness; 2) a loss of a species from a species poor community increases turnover and decreases species richness uniformity; and 3) nested communities show lower turnover than non-overlapping communities of the same richness.

The four phenomena were introduced as follows:

Species turnover and replacement. Generally, species turnover addresses the gain and the loss of species in space or time (Cody 1975, Wilson and Shmida 1984, Tuomisto 2010a). Turnover is sometimes narrowed down to replacement, defined as a situation where one species replaces exactly one other species, irrespective of the ecological role or habitat requirements of the species (Podani and Schmera 2011, Schmera et al. 2020). The Baselga's (2010) concept of 'turnover component' is actually this one-for-one replacement as has been pointed by Schmera et al. (2020). However, here we follow the most common approach (Cody 1975, Lennon et al. 2001, Koleff et al. 2003, Gaston et al. 2007) and understand turnover as a more general phenomenon referring to any change of species composition between communities. This does not mean that the concept of turnover predefined by Cody (1975) is unaffected by variation in species richness between sites. It merely states that there are cases where the variation in species richness remains unchanged, yet the species composition (i.e. turnover) changes; and that maximum change in species composition (i.e. turnover) occurs when no species are shared between communities, regardless of their species richness.

Species richness uniformity (the opposite phenomenon to the contrast between species richness). Species turnover, as defined above (Cody 1975, Lennon et al. 2001), and uniformity of species richness between sites are bound by each other. Here we adopt the idea that the uniformity of species richness and species turnover are two distinct phenomena (Lennon et al. 2001), although they interact with each other and their measures are therefore mutually s-dependent.

Nestedness. The idea of nestedness was originally intended to capture the patterns in communities of archipelagos and/or of inland islands (e.g. mountain ridges). Patterson (1984), and Patterson and Atmar (1986) noticed that species on species-poor mountain ridges were almost universally found at species-rich sites. They suggested that this pattern was driven by selective extinction and called it a 'nested pattern'. Their index of nestedness, N_0, was designed to measure the deviation from a perfectly nested communities, in which the species-poorer site has no unique species, and N_0 is thus independent of species richness of species richer community as evident from points 1–4 at pages 69–70 in Patterson (1984). This concept of nestedness fully agrees with the mathematical definition where a subset is nested within the set regardless of the size difference between the sets (Agosti et al. 2013). Although some proposed indices of nestedness imply that nestedness increases with increasing richness of the richer community (Baselga 2010), this is impractical, since it mixes two phenomena together, namely nestedness and species richness uniformity. Such view also deviates from the originally defined nestedness (Patterson 1984, Patterson and Atmar 1986).

Beta-diversity. The original idea behind beta-diversity is that different regions have different relationships between local (alpha) and regional (gamma) diversity (Whittaker 1960). Beta-diversity thus quantifies the contrast between gamma-diversity and average alpha-diversity (Whittaker 1960). Therefore, beta-diversity does not primarily compare two communities with different locations, but instead a set of sub- communities with a merged community of the whole region. The Whittaker (1960) formula that defines beta-diversity sensu stricto was later included in the indices of similarity between two different communities (Koleff et al. 2003) and several mathematical links between similarity indices and Whitaker beta-diversity sensu stricto were highlighted (Koleff et al. 2003, Tuomisto 2010b, Chao et al. 2012). However, in accord with Koleff et al. (2003), Jost (2007) and Tuomisto (2010b) we contend that the two forms of comparison (i.e. between species richness of a community and its sub- communities, and between two separate communities) should not be confused. In the following text, we will not discuss beta-diversity sensu stricto as we focus on pairwise indices.

Step 2. Formal definitions of the phenomena based on their constraints

Here we show the constraints that delimit the phenomena conceptually predefined above. To describe the constraints formally, we use six extreme arrangements of two sets as illustrated in Venn diagram in Fig. 4. Each set represents a list of species. The constraints are as follows:

- Turnover captures the contrast in species composition between two or more communities. Consequently, in Fig. 4, the turnover found in arrangement r3 must be larger than that of r1, which in turn represents a larger turnover than r2. In r5, the species lists are identical, hence there is no turnover at all, which means that the index of turnover is at its minimum. Since turnover is at its maximum where there are no species shared between the communities (Tuomisto 2010a), we state that T[r3] = T[r4] but U[*r*3] < U[*r*4] (Fig. 4). See the Supporting information for relevant mathematical theses and proofs. This phenomenon is quantified by the indices from the Jaccard family.
- Because we conceptually predefine the species richness uniformity as simply the similarity in species richness

between sites, the arrangements r1 and r3 in Fig. 4 represent the same species richness uniformity, as do r4 and r5 in Fig. 4. The latter also represents the maximum species richness uniformity (Supporting information). This phenomenon is quantified by the indices from the Gaston–Lennon family.

- Nestedness has its maximum where all the species from the species-poor community are found also in the species-rich community, and minimum where the two communities have no species in common (Patterson and Atmar 1986, Wright and Reeves 1992, Brualdi and Sanderson 1999, Baselga 2010). Therefore, the two communities are maximally nested if one of them is completely contained within the other, regardless of their size (as in arrangements r1, r2, r5, r6 in Fig. 4). This phenomenon is quantified by the indices from the Simpson family, see the Supporting information.

Note that some indices measure the above phenomena in a reversed way, so that, e.g. Jaccard index of similarity, which measures relative overlap of species composition is a reversed measure of turnover etc. (Table 1). Also, some indices do not quantify any of the phenomena above, see the Supporting information.

Inadequacy of partitioning the indices

The theory above is crucial for understanding the problems stemming from additive partitioning of indices. It makes little sense to develop indices that control one aspect of diversity patterns for the other aspects. An example is the attempt to partition 'total beta-diversity' into the nestedness-resultant and turnover components (Baselga 2010, 2012). The reason is that interdependence is the intrinsic characteristic of the phenomena which bound each other, and removing the interdependence thus necessarily alters the phenomena itself. Other fundamental reasons can be found in Box 4.

We show the irrelevance of s-dependence of indices using the triplet J, β'_{Sim}, R. The ideas demonstrated here are then universally applicable, since any pair of i-independent indices is convertible to a pair of indices J, β'_{Sim}, or R (Box 3). S-dependence (including relatedness, Chao et al. 2012) between these indices is driven by inequalities:

$$0 \le J =_1 R\frac{\beta'_{\mathrm{Sim}}}{R+1-R\beta'_{\mathrm{Sim}}} \le_2 \frac{\beta'_{\mathrm{Sim}}}{2-\beta'_{\mathrm{Sim}}}\text{, and } 0 \le J \le_3 R. \qquad (4)$$

The first equality ($=_1$) results from Eq. 1, the following inequality ($\le_2$) involves β'_{Sim} and is a consequence of

Box 4. Different phenomena are incomparable.

There are several reasons why the presence–absence indices phenomena are incomparable. The most fundamental one stems from the theory of measurement (ISO 2009, Mari et al. 2023), which states that: 1) different concepts of measurement (kinds of quantities) are incomparable even though they may have the same units (e.g. torque and work are incomparable although they have the same unit, the Newton meter), and 2) dimensionless quantities are not commensurable if they capture different concepts. The phenomena addressed here are defined using different sets of constraints (Fig. 4) and thus represent different concepts of community similarity. We demonstrate this using the phenomena of turnover and nestedness. Turnover is conceptually defined as species overlap ($S_{X\cap Y}$) relative to the richness of both communities (S_X, S_Y) (e.g. relative to total richness in the case of Jaccard, and mean richness in the case of Sørensen index). In contrast, nestedness is conceptually defined using only species richness of the poorer community (min(S_X, S_Y), $S_{X\cap Y}$), which makes it i-independent of the species richness of the richer community (max(S_X, S_Y)). This means that the index of turnover must include both S_X and S_Y in the denominator (Supporting information), whereas the denominator of the index of nestedness can include either S_X or S_Y, but not both. Different denominators make the indices of turnover and nestedness incomparable (Podani and Schmera 2012). This incomparability of different phenomena implies that their quantities cannot be summed nor subtracted (see Šizling et al. 2025 for further details).

More formally, an index of total beta-diversity (I_β) cannot be additively partitioned into its turnover and nestedness components (I_{TC}, I_{NC}, respectively, Baselga 2010, 2012). An additive partitioning means that

$$I_\beta = I_{TC} + I_{NC}\,. \qquad \text{(B4.1)}$$

There are two possibilities. First, each pair of these indices is i-dependent. Such situation is trivial, since then all the indices are equivalent. Second, at least one pair is i-independent. Our Eq. 1 shows that the third index then can be calculated from these two i-independent indices - for instance I_β, and I_{TC} are i-independent, and I_{NC} is a non-partioned function of these two ($I_{NC} = \mathrm{F}(I_\beta, I_{TC})$, Box 2). Therefore

$$I_\beta = I_{\mathrm{TC}} + F\left(I_\beta, I_{\mathrm{TC}}\right). \qquad \text{(B4.2)}$$

Consequently, the 'component' I_{NC} comprises all the information from I_{TC} (see Šizling et al. 2025 for further details). This is a direct consequence of the fact that the phenomena bound each other at the conceptual level.

$0 \le R \le 1$, and the last inequality ($\le_3$) is a consequence of both $0 \le \beta'_{\text{Sim}} \le 1$ and the first equality ($=_1$). The inequality $\le_3$ expresses the well known relatedness between J and species richness uniformity, which has led to the invention of β_{Sim} (Table 1). Both bounds ($\le_{2,3}$) that lead to s-dependence of indices have been taken as an argument that J and β_{Sim} must be corrected for the effect of the other indices (Simpson 1943, Baselga 2010). But 1) the J quantifies turnover regardless of its relatedness with R (Supporting information), and 2) the bounds ($\le_{2,3}$) only show the limits of turnover given the level of nestedness or uniformity in species richness (R) across sites. One could argue that then we could use the value of J relative to its maximum ($\le_1$), but this value reflects just the level of uniformity R (Fig. 5), so using this relative J is equivalent to direct measurement of R.

A general intuition illuminating the inadequacy of the partitioning of indices (defined by Eq. B2.1) stems from the realization that any arrangement of communities is fully described by two i-independent indices (plus species richness as a scaling parameter), so that there are no degrees of freedom for residual information about the community arrangement which would follow from the partitioning. A partitioned index that combines two i-independent indices thus does not measure a phenomenon controlled for the effect of the other phenomenon, but a combination of the phenomena, which is a different phenomenon. A good example is the attempt of Simpson (1943) to control turnover for species richness uniformity, which has led to β'_{Sim}, i.e. the index which directly measures nestedness (Almeida-Neto et al. 2008). Any partitioning thus leads either 1) to an index that measures one of the phenomena we already measure by another established index (as in the above case) or 2) an index that is not strictly monotonically related to any of the phenomena that we want to measure (as in Baselga´s approach in Fig. 1).

Real-world data examples: how to make inference from the indices?

To provide real-world examples of an inference based on the pairwise indices, we use real-world data from Šizling et al. (2016) who studied temporal changes of central European plants during Holocene, and Xu et al. (2023) who reported global patterns of angiosperm trees. The main purpose of these examples is to show that the popular partitioning of

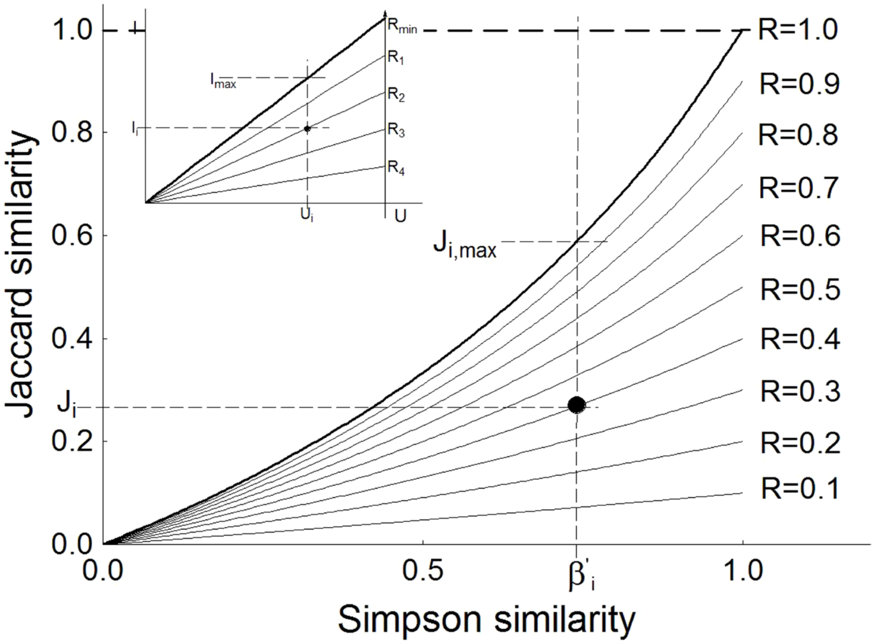


Figure 5. The reason why mutual dependence of indices cannot be eliminated by controlling for effects of other phenomena. The relationship between Jaccard similarity (J) and Simpson similarity (β'_{Sim}) is determined by species richness uniformity R (Eq. 4), each line shows points with equal R (the dot has R=0.4). J is limited by 1 (dashed bold line) and each particular J_i is limited by $J_{i,\max}$, which is a function of β'_{Sim} (bold line, Eq. 4). The value of J_i relative to the limit of one should be read as a reversed measure of turnover without a need for any additive partitioning of J or β'_{Sim} into components, i.e. it is a reversed measure of turnover per se. The value J_i rescaled by $J_{i,\max}$ shows R_i, and thus represents richness uniformity per se. The fact that R affects J is a matter of i-dependence between triplet J, β'_{Sim}, R, and thus the interplay between J and R should not invalidate their unique meaning determined by the constraints. This is like the Ohm's law in physics where the triplet of voltage (U), current (I), and resistance (R) are mutually dependent ($U = RI$, see inset). The understanding in physics is that I_i is a measure of current regardless of its dependence on R and U. No one would relativize I to R, nor say that in the absence of voltage U equals I, and thus U minus I measures the 'voltage resultant component'. In the same way, J_i is a measure of turnover even though it also scales with richness uniformity. For the same reason β'_{Sim} is a measure of nestedness, without a possibility of additive partitioning into controlled components.

these indices (Baselga 2010, 2012) may lead to interpretations that do not correspond to any of the four phenomena intuitively distinguished by ecologists. Baselga's (2010, 2012) frameworks cannot be directly compared with our approach because Baselga (2010, 2012) uses different concepts of turnover and nestedness. However, we show that our framework leads to interpretations which better correspond to the intuitive conceptual predefinitions of respective phenomena.

Example 1 (Temporal change in Holocene plants). Šizling et al. (2016) reported a temporal increase in mean species richness $\langle S\rangle$ from 27 to 32, almost no change of β'_{Sim} ($\beta'_{\mathrm{Sim}} \cong 0.67$), and a steep increase of average J from ca 0.4 to ca 0.5 during the short period (5900–5700 BP) when agriculture was introduced in the region. J and β'_{Sim} uniquely determine R (Eq. 1), which increased from 0.74 to 0.99 (Fig. 6). In addition, R uniquely determines the contribution of species-poor ($S_{\mathrm{poor}} = \langle S\rangle\, 2R/(R+1)$) and species-rich ($S_{\mathrm{rich}} = R^{-1} \bullet S_{\mathrm{poor}}$) sites to the average alpha diversity (Supporting information). S_{poor} increased from ca 23 to 32, and S_{rich} increased from ca 31 to 32. We conclude that recent plant communities in central Europe show lower spatial turnover, and higher richness uniformity (Fig. 6c) than those before the introduction of agriculture. Since the average nestedness has not changed from the period before agriculture until today (Fig. 6c), the species poor sites have also kept a constant ratio of unique to common species. The introduction of agriculture was therefore accompanied by homogenization of plant communities, with proportionally higher increase of richness at species-poor sites, and increasing average species richness of all sites. This homogenization of communities as a consequence of

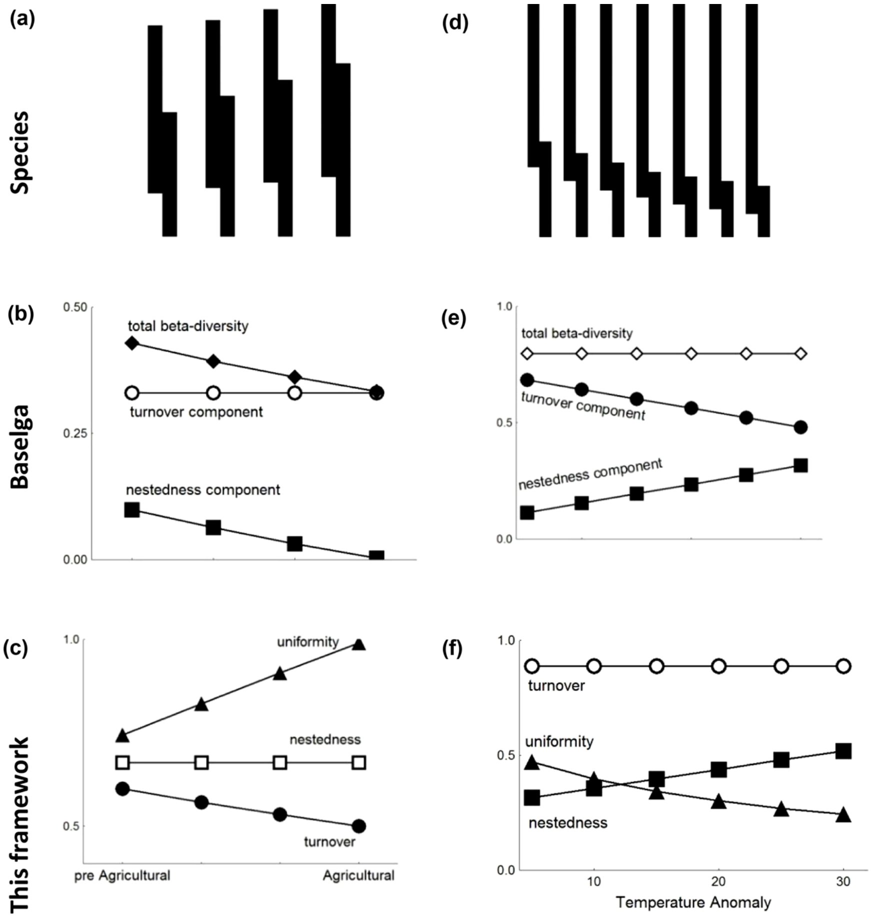


Figure 6. The illustration of the difference between the approach that uses Baselga's (2010) additive partitioning of pairwise indices (b, e) and the framework presented here (c, f), using data from Šizling et al. (2016, left) and Xu et al. (2023, right). (a) and (d) show the species communities and respective overlaps in the lists of species. We used the algorithm in Box 2–3 to compute (a) and (b) from (c), and (f) and (d) from (e), as only data in (c) and (e) was published (the computed S_X ranges from 75.62 to 89.54, $S_Y \cong$ 35.66–21.74 and $S_{X\cap Y} \cong$ 11.28 in Xu. et al. 2023, S_X ranges from 30.97 to 32.16, $S_Y \cong$ 23.03–31.84 and $S_{X\cap Y} \cong$ 15.43–21.33 in Šizling et al. 2016). The plots (b) and (e) display trends for β_{Sim} (circles), β_{sne} (squares), and β_{t2} (diamonds), labeled according to Xu et al. (2023) as turnover component, nestedness component, and total beta-diversity, respectively ($\beta_{\mathrm{Sim}} + \beta_{\mathrm{sne}} = \beta_{t2}$). The plots (c) and (f) show trends for $1 - J$ (circles), β'_{Sim} (squares), and R (triangles), labeled according to this report as turnover, nestedness, and uniformity, respectively. Differences between (b) and (c), as well as (e) and (f), lead to divergent interpretations when applying our framework compared to Baselga's (2010, 2012). The lines show least square regressions; where the regression was not significant ($\alpha > 0.05$, open symbols) we plotted a constant line.

agriculture is in agreement with other analyses (Kolář et. al. 2022) and corresponds to the visual inspection of Fig. 6a.

When we apply Baselga's (2012) partitioning method to the data of Šizling et al. (2016), we get $\beta_{\mathrm{sne}} = 0.1$, $\beta_{\mathrm{Sim}} = 0.33$ for the pre-agricultural landscape and $\beta_{\mathrm{sne}} = 0.00$, $\beta_{\mathrm{Sim}} = 0.33$ for early agricultural landscape. This reads in the Baselga's (2010, 2012) framework as no temporal change of 'turnover component', and decreases in both the 'nestedness-resultant component' and total beta-diversity (Fig. 6b). This interpretation based on Baselga's (2010, 2012) partitioning thus directly contradicts our interpretation (Fig. 6c). Baselga's framework misses the increasing uniformity of species richness (due to the increase of species richness of species-poor communities) which seems to be driving the other observed patterns.

Example 2 (Global patterns of tree diversity). Xu et al. (2023) used the Baselga (2010) framework and reported a general increase of the nestedness-resultant component (β_{sne}) with an increasing temperature anomaly from ca. 0.11 to 0.32, and a decrease of the turnover component (β_{Sim}) from ca 0.68 to 0.48 (Fig. 6e). The total beta-diversity ($\beta_{\mathrm{sne}} + \beta_{\mathrm{Sim}}$) did not show a significant trend along the temperature anomaly gradient.

To reanalyze Xu et al.'s (2023) data, we used our framework (Box 2–3, and Calculator RI4 in Šizling et al. (2025)) and computed the average S_{X}, S_{Y} and $S_{\mathrm{X} \cap \mathrm{Y}}$ (Fig. 6d) that correspond to the indices reported by Xu et al. (2023) (Fig. 6e). Since Xu et al. (2023) did not report species richness, we standardized it to $S_X + S_Y - S_{X \cap Y} \cong 100$ without loss of generality. Our reanalysis of the data from Xu et al. (2023) suggests no significant trend in turnover, increasing nestedness, and decreasing species richness uniformity with increasing temperature anomaly (Fig. 6f). We thus detect a heterogenization of communities as the temperature anomaly increases, while Baselga's (2010, 2012) framework indicates community homogenization (i.e. increasing similarity as quantified by an increasing 'nestedness component', and a decreasing 'turnover component', Fig. 6e). Visual inspection of the arrangement of species communities (Fig. 6d) shows that the main pattern is the decreasing uniformity in species richness of both communities, which is consequently associated with increasing nestedness, and no pattern of species turnover (as the overlap remains constant). This pattern of community heterogenization aligns with our interpretation of the indices, but is in striking contrast to the interpretation within Baselga's framework.

Discussion

We have shown that the information content of indices is their key feature for understanding their interrelationships, classification, and a proper link to biologically relevant phenomena. The phenomena measured by the indices constrain each other, and respective indices are thus necessary related sensu Chao et al. (2012), leading to their statistical dependence. We demonstrated that it is pointless to control one index for an influence of another index, because any effort to eliminate s-dependence results in a measure of a different phenomenon. In other words, there are no s-independent indices that capture mutually dependent phenomena.

On the other hand, i-independent indices are desirable, as they provide (by definition) different information about the system. However, a sufficient number of i-independent indices have already been invented, and even those not yet invented are i-dependent on J and β_{Sim}, or R (Box 2–3). There is thus no need for new indices. We have mathematically proven that two i-independent indices, in combination with a species richness value (which acts as a scaling parameter), provide complete information about the communities. This does not hold for the additively partitioned indices because partitioning often leads to information loss (Eq. B3.2 in Box 3).

Moreover, when variation of species richness across sites is negligible, or when there are no species that are unique to the species-poorer community (i.e. perfect nestedness), i-independent indices appear as if they were i-dependent, and can be derived from each other. The interrelation between communities which leads to these effects thus leads to higher correlations between the observed index values than what would follow from the theory (Ulrich et al. 2017, Supporting information), which makes some families of indices statistically indistinguishable (Supporting information). This is apparently the reason why it has been so difficult to achieve agreement regarding which indices characterize different phenomena (Lennon et al. 2001, Gaston et al. 2007, Ulrich and Gotelli 2007, Tuomisto 2010a, 2010b, Baselga 2012, Podani and Schmera 2011, Ulrich et al. 2017, Schmera et al. 2020).

In terms of practical utility of the indices (Box 5, Table 2), the finding that the phenomena distinguished by ecologists are inevitably mutually dependent and that two i-independent indices fully characterize the system implies that it is reasonable to calculate just two indices belonging to different families (Box 2) to make a proper inference. These indices (for example J and R) can be, if necessary, converted to any other index within their respective family (and vice versa). Different indices within a family of i-dependent indices are different transformations of each other and their particular values thus reflect just different rescaling. On the other hand, indices from different families cannot be directly compared to each other (and thus also subtracted from each other), as their values characterize different phenomena.

We have focused only on pairwise indices that compare spatially or temporarily non-overlapping communities, avoiding the topic of beta-diversity sensu stricto (Whittaker 1960), which comprises the relationship between local and regional species richness, or, more specifically, between alpha and gamma diversity. This relationship is to some extent related to turnover, since when alpha is considerably lower than gamma, there must be high turnover among communities. However, the precise mathematical links between pairwise community turnover and Whittaker's beta is a separate issue (Koleff et al. 2003, Tuomisto 2010b, Chao et al. 2012) and there is no straightforward way to derive turnover from beta-diversity sensu stricto for more than two communities (Šizling et al. 2011).

Box 5. Practical guidelines.

Practical inference from pairwise indices can have different purposes (Anderson et al. 2010), for instance: 1) exploration of diversity (Qian et al. 2009) or (dis)similarity (Simpson 1943) of biotas, 2) revealing non-random origin of a spatial or temporal biodiversity patterns (Patterson and Atmar 1986, Ulrich and Gotelli 2013), 3) meta-analysis based on indices extracted from the literature, and 4) exploration of the behavior of the indices along temporal, spatial or environmental gradients. Here are practical recommendations for using the presence–absence indices:

1. Avoid partitioning of indices that supposedly removes an effect of one phenomenon from an effect of another phenomenon (e.g. controlling turnover for the effect of richness uniformity or controlling nestedness for the effect of turnover). Partitioned indices are either mathematically flawed (e.g. they violate the requirement of continuity), and their meaning is thus unclear, and/or they measure a different phenomenon from what they were originally claimed to measure. An example of the latter is the standardization of an index relative to its maximum value. For instance, standardization of Jaccard index by its maximum determined by the nestedness (bold line in Fig. 5) is in fact a measure of species richness uniformity (Fig. 5). The reason is that nestedness, together with richness uniformity R, uniquely determine J, so that measuring J relative to $J_{\max}$ (bounded by nestedness) is equivalent to measuring R.
2. When choosing an index, first consider which phenomenon it should capture (Table 1–2), then select a corresponding family of indices (Fig. 3, Supporting information). It does not matter which index within the family is selected, as all indices within any family can be converted to each other and are thus practically equivalent. If there is a need for an appropriate statistical distribution of the index, a proper transformation is preferable over an invention of, or search for, a new index.
3. When comparing already published indices, use the equations in the Supporting information or the Calculator RI4 in Šizling et al. (2025) to convert them to a common index that is the most desirable for a given purpose. If a desirable common index is not available in the publication, it can be calculated from any i-dependent index, or any two i-independent indices (Box 2–3). The indices of two different phenomena (two i-independent indices) are never commensurable.
4. Use null models that randomize species incidences to assess the statistical significance of the phenomenon in focus (Gotelli and Ulrich 2012). Null models are useful for distinguishing inherent correlations (Chao et al. 2012) from ecologically meaningful correlations and relationships. Note, however, that relevant null models are still under development (Chase et al. 2011, Legendre 2019), and it may not always be clear what exactly to randomize in the site-by-species matrix.
5. When publishing indices from your research, ensure that these indices conserve complete information, i.e., publish at least two i-independent indices (J, and β_{Sim} or R are recommended) plus regional or mean local species richness, if available. Use our Calculator RI4 in Šizling et al. (2025) when in doubt about whether the employed indices carry complete information.

Our findings regarding the pairwise comparison of two communities can also be applied to pairwise comparisons of two species ranges, and generally to all pairwise indices of interspecific spatial association derived from binary data (Keil et al. 2021). In such a case, S_X, S_Y and $S_{X \cap Y}$ can represent the number of sites occupied by the first species, second species, and both species together, respectively. For instance, the co-occurrence of two species over multiple sites (Diamond 1975, Connor and Simberloff 1979, Diamond and Gilpin 1982, Gotelli and McCabe 2002) can be measured by the Jaccard index (Supporting information). Analogically, indices from the Simpson family and Gaston–Lennon family can be used for characterizing nestedness of species' geographic ranges (Šizling et al. 2009) and the uniformity in species extent of occurrence (range size).

The relationships between indices and between respective families have important implications for understanding the forces shaping distance decay (an increase or decrease of an index with increasing distance) in community similarity (Nekola and White 1999). It is meaningless to assume a universal functional form of the distance decay of community similarity, e.g. exponential (Nekola and McGill 2014). This follows from the mutual non-linear scaling between different indices, even between the i-dependent ones. Consequently, if one index reveals, say, an exponential distance decay, another index can reveal non-exponential distance decay.

Our results are relevant for the interpretation of hundreds of empirical assessments that used the additive partitioning of pairwise indices. To date, the original publication describing the partitioning to nestedness and turnover components (Baselga 2010) has 4160 citations and the R package 'betapart' which does the partitioning (Baselga and Orme 2012) has 2435 citations (Google Scholar accessed on 23 September 2025). Most of these citations are empirical studies, including some high-profile ones in top journals (Molinos et al. 2016, Gotelli et al. 2017, Rocha et al. 2018, Blowes et al. 2019, Chase et al. 2020, Xu et al. 2023). The approach has also been gaining momentum in young fields, for example in microbial ecology

Table 2. Table of possible inference from the variation of pairwise indices along temporal, geographic, or environmental gradients. Some inference can be based on variation of a single index, regardless of other indices (first three rows). However, more specific inference follows from situations when one index changes along the gradient, while the other remains constant or shows only negligible variation (all remaining rows).

Variation along a gradient		
Yes	No	Inference
J		Change in turnover between regions/periods; change in proportion of shared species relative to species richness of both communities.
β'_{Sim}		Change in nestedness compared between regions/periods (i.e. nestedness for region 1 versus nestedness for region 2); change in proportion of common species relative to richness of species poorer sites.
R		Change in species richness uniformity of communities; change in variation of alpha diversity.
β'_{Sim}	S_{rich}	Change in nestedess between regions/periods is driven by exchange of species at species poor sites.
$\langle S \rangle$	S_{rich}	Difference of alpha diversity between regions/periods is driven by sites with fewer species.
$\langle S \rangle$	S_{poor}	Difference of alpha diversity between regions/periods is driven by species-richer sites.
R	$\langle S \rangle$	Regions/periods differ in spatial variation of alpha diversity without change of its average value.
β'_{Sim}	J	Change in nestedness, but not turnover; simultaneous change in lists of species that are unique to species-richer and species-poorer sites; variation in β'_{Sim}; no change in J; Eq. 1 induce necessarily a change in R.
J	β'_{Sim}, S_{poor}	Change in turnover, but not nestedness; change in list of species that are unique to species-richer sites; Eq. 1 induce necessarily a change in R.

(Shade et al. 2013). Based on the conceptual problems that we have described, we call for a critical re-evaluation of the practice. A need may also arise for re-analyzing and re-interpreting some of the studies; in this effort, our equations from Box 2 and 3 (or Calculator RI4 in Šizling et al. (2025)) can be helpful.

In conclusion, we have shown that the combination of two indices from two informationally independent families, namely Jaccard and Simpson families, or a combination of one index from these families and species richness uniformity, is sufficient to full characterization of all biologically relevant phenomena. All the other indices (including those not yet invented) are equivalent to one or two classical indices and represent their different transformations, or (in the case of partitioned indices) do not provide reliable measure of the phenomena ecologists are interested in. Our framework allows recalculation of any index using two i-independent indices, enabling standardization across published studies, and provides a universal tool for the comprehensive study of patterns of species occupancy and their variation across space and time.

Acknowledgements – Special thanks to A. Bryson and G. Ridder for help with English, F. He and W. E. Kunin for their comments on the first versions, Z. Sokolíčková, E. Šizlingová for their help in the field, J. Klimešová and A. Bernard for lessons in Arctic plants, and R. K. Colwell and H. Tuomisto for insightful comments. The project was supported by the Czech Science Foundation (EXPRO 20-29554X). We used the Czech Scientific Infrastructure of the University of South Bohemia, J. Svoboda house on Svalbard – CzechPolar2 LM2015078 supported by MŠMT. P. Keil was funded by the European Union (ERC, BEAST, 101044740). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Funding – The work was supported by the Czech Science Foundation (grant number 20-29554X). All data were collected by the authors by non-destructive methods and in accord with law.

Conflict of interest – The authors declare no conflict of interest.

Author contributions

Arnost L. Šizling: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Supervision (equal); Writing – original draft (lead). **Petr Keil**: Methodology (equal); Writing – original draft (equal). **Even Tjørve**: Data curation (equal); Methodology (equal); Writing – original draft (equal). **Kathleen M. C. Tjørve**: Writing – original draft (equal). **Jakub D. Žárský**: Data curation (lead); Resources (equal); Writing – original draft (supporting). **David Storch**: Conceptualization (equal); Investigation (equal); Methodology (equal); Project administration (lead); Resources (equal); Writing – original draft (equal).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ecog.07996.

Data availability statement

Data, index definitions, theses and proofs, calculator and its code are available from the ZenodoDigital Repository: https://doi.org/10.5281/zenodo.8084396 (Šizling et al. 2025).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Agosti, M., Esposito, F., Ferilli, S. and Ferro, N. 2013. Digital libraries and archives, IRCDL 2012, Bari, Italy, February 9–10, 2012, revised selected papers. – Springer.

Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D. and Ulrich, W. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. – Oikos 117: 1227–1239.

Almeida-Neto, M., Frensel, D. M. B. and Ulrich, W. 2012. Rethinking the relationship between nestedness and beta diversity: a comment on Baselga (2010). – Global Ecol. Biogeogr. 21: 772–777.

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L. and Harrison, S. P. 2010. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. – Ecol. Lett. 14: 19–28.

Audi, R. 1999. The Cambridge dictionary of philosophy, 2nd edn. – Cambridge Univ. Press.

Barwell, L. J., Isaac, N. J. B. and Kunin, W. E. 2015. Measuring ß-diversity with species abundance data. – J. Anim. Ecol. 84: 1112–1122.

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. – Global Ecol. Biogeogr. 19: 134–143.

Baselga, A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. – Global Ecol. Biogeogr. 21: 1223–1232.

Baselga, A. and Orme, C. D. L. 2012. betapart: an R package for the study of beta diversity. – Methods Ecol. Evol. 3: 808–812.

Blowes, S. A. et al. 2019. The geography of biodiversity change in marine and terrestrial assemblages. – Science 336: 339–345.

Bray, J. R. and Curtis, J. T. 1957. An ordination of the upland forest communities of southern Wisconsin. – Ecol. Monogr. 27: 325–349.

Brualdi, R. A. and Sanderson, J. G. 1999. Nested species subsets, gaps, and discrepancy. – Oecologia 119: 256–264.

Carvalho, J. C., Cardoso, P. and Gomes, P. 2012. Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. – Global Ecol. Biogeogr. 21: 760–771.

Chao, A. and Colwell, R. K. 2017. Thirty years of progeny from Chao's inequality: estimating and comparing richness with incidence data and incomplete sampling. – Stat. Oper. Res. Trans. 41: 1–52.

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. 2005. A new statistical approach for assessing compositional similarity based on incidence and abundance data. – Ecol. Lett. 8: 148–159.

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. – Biometrics 62: 361–371.

Chao, A., Chiu, C. H. and Hsieh, T. C. 2012. Proposing a resolution to debates on diversity partitioning. – Ecology 93: 2037–2051.

Chao, A., Chiu, C. H. and Jost, L. 2014a. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. – Annu. Rev. Ecol. Evol. Syst. 45: 297–324.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K. and Ellison, A. M. 2014b. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. – Ecol. Monogr. 84: 45–67.

Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M. and Inouye, B. D. 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. – Ecosphere 2: 24.

Chase, J. M., Blowes, S. A., Knight, T. M., Gerstner, K. and May, F. 2020. Ecosystem decay exacerbates biodiversity loss with habitat loss. – Nature 584: 238–243.

Chiu, C. H., Jost, L. and Chao, A. 2014. Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. – Ecol. Monogr. 84: 21–44.

Cody, M. L. 1975. Towards a theory of continental species diversities. – In: Cody, M. L. and Diamond, J. M. (eds), Ecology and evolution of communities. Harvard Univ. Press, pp. 214–257.

Connor, E. F. and Simberloff, D. 1979. The assembly of species communities: chance or competition? – Ecology 60: 1132–1140.

Diamond, J. M. 1975. Assembly of species communities. – In: Cody, M. L. and Diamond, J. M. (eds), Ecology and evolution of communities. Harvard Univ. Press, pp. 342–444.

Diamond, J. M. and Gilpin, M. E. 1982. Examination of the 'null' model of Connor and Simberloff for species cooccurrences on islands. – Oecologia 52: 64–74.

Dornelas, M., Gotelli, N. J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C. and Magurran, A. E. 2014. Assemblage time series reveal biodiversity change but not systematic loss. – Science 344: 296–299.

Gaston, K. J., Rodrigues, A. S. L., Van Rensburg, B. J., Koleff, P. and Chown, S. L. 2001. Complementary representation and zones of ecological transition. – Ecol. Lett. 4: 4–9.

Gotelli, N. J. and Ulrich, W. 2012. Statistical challenges in null model analysis. – Oikos 121: 171–180.

Gaston, K. J., Evans, K. L. and Lennon, J. J. 2007. The scaling of spatial turnover: pruning the thicket. – In: Storch, D., Marquet, P. A. and Brown, J. H. (eds), Scaling biodiversity. Cambridge Univ. Press, pp. 181–222.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. – Ecology 81: 2606–2621.

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

Gotelli, N. J., Shimadzu, H., Dornelas, M., McGill, B., Moyes, F. and Magurran, A. E. 2017. Community-level regulation of temporal trends in biodiversity. – Sci. Adv. 3: e1700315.

Greve, M., Gremmen, N. J., Gaston, K. J. and Chown, S. L. 2005. Nestedness of southern Ocean Island biotas: ecological perspectives on a biogeographical conundrum. – J. Biogeogr. 32: 155–168.

Harte, J. and Kinzig, A. P. 1997. On the implications of species–area relationships for endemism, spatial turnover, and food web patterns. – Oikos 80: 417–427.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. – Ecology 54: 427–432.

ISO 2009. ISO 80000-1:2009(en). Quantities and units — Part 1: general. – International Organization for Standardization.

Jaccard, P. 1912. The distribution of the flora in the alpine zone. – New Phytol. 11: 37–50.

Jost, L. 2007. Partitioning diversity into independent alpha and beta components. – Ecology 88: 2427–2439.

Jurasinski, G., Retzer, V. and Beierkuhnlein, C. 2009. Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. – Oecologia 159: 15–26.

Keil, P., Wiegand, T., Tóth, A. B., McGlinn, D. J. and Chase, J. M. 2021. Measurement and analysis of interspecific spatial associations as a facet of biodiversity. – Ecol. Monogr. 91: e01452.

Kolář, J., Macek, M., Tkáč, P., Novák, D. and Abraham, V. 2022. Long-term demographic trends and spatio-temporal distribution of past human activity in central Europe: comparison of archaeological and palaeoecological proxies. – Quat. Sci. Rev. 297: 107834.

Koleff, P. and Gaston, K. J. 2002. The relationships between local and regional species richness and spatial turnover. – Global Ecol. Biogeogr. 11: 363–375.

Koleff, P., Gaston, K. J. and Lennon, J. J. 2003. Measuring beta diversity for presence–absence data. – J. Anim. Ecol. 72: 367–382.

Legendre, P. 2019. A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. – Ecol. Evol. 9: 3500–3514.

Legendre, P. and De Cáceres, M. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. – Ecol. Lett. 16: 951–963.

Lennon, J. J., Koleff, P., Greenwood, J. J. D. and Gaston, K. J. 2001. The geographical structure of British bird distributions: diversity, spatial turnover and scale. – J. Anim. Ecol. 70: 966–979.

Lyashevska, O. and Farnsworth, K. D. 2012. How many dimensions of biodiversity do we need? – Ecol. Indic. 18: 485–492.

Mari, L., Wilson, M. and Maul, A. 2023. Measurement across the sciences, 2nd edn. – Springer Nature.

Matthews, T., Aspin, T., Ulrich, W., Baselga, A., Kubota, Y., Proios, K., Triantis, K. A., Whittaker, R. J. and Strona, G. 2019. Can additive beta diversity be reliably partitioned into nestedness and turnover components? – Global Ecol. Biogeogr. 28: 1146–1154.

Molinos, J. G., Halpern, B. S., Schoeman, D. S., Brown, C. J., Kiessling, W., Moore, P. J., Pandolfi, J. M., Poloczanska, E. S., Richardson, A. J. and Burrows, M. T. 2016. Climate velocity and the future global redistribution of marine biodiversity. – Nat. Clim. Change 6: 83–88.

Nekola, J. C. and White, P. S. 1999. The distance decay of similarity in biogeography and ecology. – J. Biogeogr. 26: 867–878.

Nekola, J. C. and McGill, B. J. 2014. Scale dependency in the functional form of the distance decay relationship. – Ecography 37: 309–320.

Neumann, J. von. and Morgenstern, O. 1953. Theory of games and economic behavior. – Princeton Univ. Press.

Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Gray, C. L., Scharlemann, J. P. W., Börger, L., Phillips, H. R. P., Sheil, D., Lysenko, I. and Purvis, A. 2016. Global patterns of terrestrial assemblage turnover within and among land uses. – Ecography 39: 1151–1163.

Orlitsky, A. 2003. Information theory. – In: Meyers, R. A. (ed.), Encyclopedia of physical science and technology. Elsevier Science, Academic Press, pp. 751–769.

Patterson, B. D. 1984. Mammalian extinction and biogeography in the Southern Rocky Mountains. – In: Nitecki, M. H. (ed.), Extinctions. Univ. of Chicago Press, pp. 247–293.

Patterson, B. D. and Atmar, W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. – Biol. J. Linn. Soc. 28: 65–82.

Podani, J. and Schmera, D. 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. – Oikos 120: 1625–1638.

Podani, J. and Schmera, D. 2012. A comparative evaluation of pairwise nestedness measures. – Ecography 35: 889–900.

Qian, H., Badgley, C. and Fox, D. L. 2009. The latitudinal gradient of beta diversity in relation to climate and topography for mammals in North America. – Global Ecol. Biogeogr. 18: 111–122.

Raup, D. M. and Crick, R. E. 1979. Measurement of faunal similarity in paleontology. – J. Paleontol. 53: 1213–1227.

Rocha, L. A., Pinheiro, H. T., Shepherd, B., Papastamatiou, Y. P., Luiz, O. J., Pyle, R. L. and Bongaerts, P. 2018. Mesophotic coral ecosystems are threatened and ecologically distinct from shallow water reefs. – Science 361: 281–284.

Ruggiero, A., Lawton, J. H. and Blackburn, T. M. 1998. The geographic ranges of mammalian species in South America: spatial patterns in environmental resistance and anisotropy. – J. Biogeogr. 25: 1093–1103.

Schluter, D. and Ricklefs, R. E. 1993. Species diversity: an introduction to the problem. – In: Ricklefs, R. E. and Schluter, D. (eds), Species diversity in ecological communities: historical and geographical perspectives. Chicago Univ. Press, pp. 1–12.

Schmera, D. and Podani, J. 2011. Comments on separating components of beta-diversity. – Comm. Ecol. 12: 153–160.

Schmera, D., Podani, J. and Legendre, P. 2020. What do beta diversity components reveal from presence–absence community data? Let us connect every indicator to an indicandum! – Ecol. Indic. 117: 106540.

Shade, A., McManus, P. S. and Handelsman, J. 2013. Unexpected diversity during community succession in the apple flower microbiome. – mBio 4: e00602-12.

Simpson, G. G. 1943. Mammals and the nature of continents. – Am. J. Sci. 241: 1–31.

Šizling, A. L., Šizlingová, E., Storch, D., Reif, J. and Gaston, K. J. 2009. Rarity, commonness, and the contribution of individual species to species richness patterns. – Am. Nat. 174: 82–93.

Šizling, A. L., Kunin, W. E., Šizlingová, E., Reif, J. and Storch, D. 2011. Between geometry and biology: the problem of universality of the species–area relationship. – Am. Nat. 178: 602–611.

Šizling, A. L., Pokorný, P., Juřičková, L., Horáčková, J., Abraham, V., Šizlingová, E., Ložek, V., Tjørve, E., Tjørve, K. M. C. and Kunin, W. 2016. Can people change the ecological rules that appear general across space? – Global Ecol. Biogeogr. 25: 1072–1084.

Šizling, A. L., Keil, P., Tjørve, E., Tjørve, K. M. C., Žárský, J. D. and Storch, D. 2025. Data, SW, manual, SW code, theses and proofs from: proofs from: Mathematically and biologically consistent framework for presence–absence pairwise indices. – Zenodo Digital Repository, https://doi.org/10.5281/zenodo.8084396.

Sørensen, T. A. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content, and its application to analyses of the vegetation on Danish commons. – Kongelige Dan. Videnskabernes Selsk. Biol. Skrifter 5: 1–34.

Tjørve, E., Tjørve, K. M. C., Šizlingová, E. and Šizling, A. L. 2018. Great theories of species diversity in space and why they were forgotten: the beginnings of a spatial ecology and the Nordic early 20th-century botanists. – J. Biogeogr. 45: 530–540.

Tuomisto, H. 2010a. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Quantifying beta diversity and related phenomena. – 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.

Tuomisto, H. 2012. An updated consumer's guide to evenness and related indices. – Oikos 121: 1203–1218.

Ulrich, W. and Gotelli, N. J. 2007. Disentangling community patterns of nestedness and species co-occurrence. – Oikos 116: 2053–2061.

Ulrich, W. and Gotelli, N. J. 2013. Pattern detection in null model analysis. – Oikos 122: 2–18.

Ulrich, W., Almeida-Neto, M. and Gotelli, N. J. 2009. A consumer's guide to nestedness analysis. – Oikos 118: 3–17.

Ulrich, W., Baselga, A., Kusumoto, B., Shiono, T., Tuomisto, H. and Kubota, Y. 2017. The tangled link between β- and ϒ-diversity: a Narcissus effect weakens statistical inferences in null model analyses of diversity patterns. – Global Ecol. Biogeogr. 26: 1–5.

Ulrich, W., Kubota, Y., Kusumoto, B., Baselga, A., Tuomisto, H. and Gotelli, N. J. 2018. Species richness correlates of raw and standardized co-occurrence metrics. – Global Ecol. Biogeogr. 27: 395–399.

Vellend, M. 2001. Do commonly used indices of β-diversity measure species turnover? – J. Veg. Sci. 12: 545–552.

Whittaker, R. H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. – Ecol. Monogr. 30: 279–338.

Wilson, M. V. and Shmida, A. 1984. Measuring beta diversity with presence–absence data. – J. Ecol. 72: 1055–1064.

Wright, D. H. and Reeves, J. H. 1992. On the meaning and measurement of nestedness of species assemblages. – Oecologia 92: 416–428.

Xu, W.-B. et al. 2023. Global beta-diversity of angiosperm trees is shaped by quaternary climate change. – Sci. Adv. 9: eadd8553.