Check for updates

REVIEW

*Ecological Monographs*, 91(3), 2021, e01452
© 2021 The Authors. *Ecological Monographs* published by Wiley Periodicals LLC on behalf of Ecological Society of America.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

# Measurement and analysis of interspecific spatial associations as a facet of biodiversity

PETR KEIL,[1,2,3,7] THORSTEN WIEGAND,[1,4] ANIKÓ B. TÓTH,[5]
DANIEL J. MCGLINN,[6] AND JONATHAN M. CHASE[1,2]

*[1]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*
*[2]Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany*
*[3]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Kamýcká 129, Praha – Suchdol,165 00, Czech Republic*
*[4]Department of Ecological Modelling, Helmholtz Centre for Environmental Research - UFZ, 04318 Leipzig, Germany*
*[5]Centre for Ecosystem Sciences, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052 Australia*
*[6]Department of Biology, College of Charleston, Charleston, South Carolina 29401 USA*

*Citation:* Keil, P., T. Wiegand, A. B. Tóth, D. J. McGlinn, and J. M. Chase. 2021. Measurement and analysis of interspecific spatial associations as a facet of biodiversity. Ecological Monographs 91(3):e01452. 10.1002/ecm.1452

*Abstract.* Interspecific spatial associations (ISA), which include co-occurrences, segregations, or attractions among two or more species, can provide important insights into the spatial structuring of communities. However, ISA has primarily been examined in the context of understanding interspecific interactions, while other aspects of ISA, including its relations to other biodiversity facets and how it changes in the face of anthropogenic pressures, have been largely neglected. This is likely because it is unclear what makes ISA useful in a biodiversity context, little is known about the theoretical connections between ISA and other biodiversity facets, and there is a confusing variety of approaches to measuring ISA. Here, we first review the metrics of ISA. These include spatially implicit and explicit indices of association for binary, abundance, and point pattern data. We test and compare these approaches on empirical and simulated data, and we provide recommendations for how to use and interpret them in biodiversity science. We argue that measurements of ISA are more informative when they are spatially explicit (i.e., distance dependent). We then review links of ISA to other classical biodiversity facets, such as alpha, beta, and gamma diversity, and show that they mostly fail to reflect changes/variation in ISA, with the exception of average pairwise beta diversity. This underscores the need for a specific focus on ISA in large-scale biodiversity assessments. Finally, we argue that there are important, and underappreciated, reasons to study ISA that are unrelated to its link to biotic interactions. Specifically, ISA can provide strong tests of biodiversity theories that require multiple patterns to benchmark against, and it can be explored for potentially predictive macroecological patterns.

*Key words: conspecific; co-occurrence;* C-*score; geostatistics; grain; heterospecific; interspecific; intraspecific; point process; semivariogram; spatial scale.*

> . . . there is no absolutely general measure of the degree of dependence. (Cramér 1924, quoted by Goodman and Kruskal 1979, quoted by Hubálek 1982)

## INTRODUCTION

Organisms are rarely distributed independently on each other in space. Instead, individuals within a single species may display non-random *conspecific spatial aggregation* (CSA, Fig. 1, Appendix S1) in which they clump together or spread apart. Additionally, individuals between species may be spatially segregated or display spatial attraction indicative of non-random *interspecific spatial association* (ISA, Fig. 1, Appendix S1). CSA has

Manuscript received 16 July 2020; accepted 24 November 2020. Corresponding Editor: Aimée T. Classen.
[7] E-mail: pkeil@seznam.cz

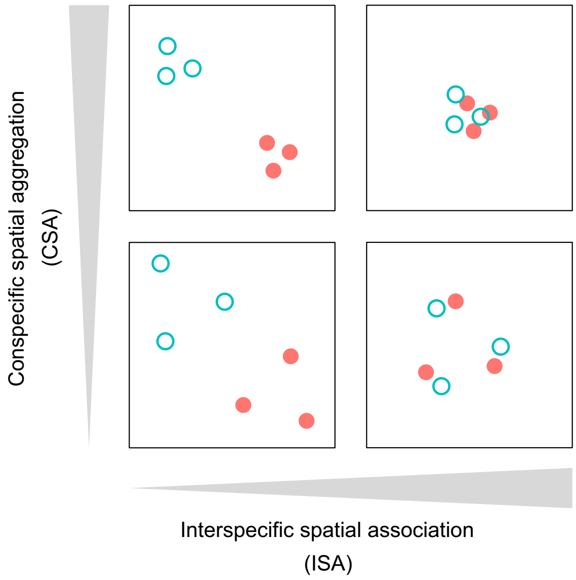


Fig. 1. The difference between conspecific spatial aggregation (CSA, vertical gradient) and interspecific spatial association (ISA, horizontal gradient). While CSA is the dependence (negative or positive) in spatial positions of individual within a species, ISA is the dependence (negative or positive) of spatial position of individuals of one species on spatial position of individuals of the other species. ISA is the subject of this paper. Points are individuals, colors mark species.

known links to fundamental biodiversity concepts such as species accumulation curves, species–area species–area relationships, and beta diversity (Table 1; Storch et al. 2008, McGlinn et al. 2019); in contrast, the connections between ISA and these concepts are less known and scattered across the literature. Further, CSA and ISA are each generated by distinct mechanisms: CSA emerges due to conspecific interactions, the interplay between niche width and environment, and species-specific dispersal limits. In contrast, ISA is a result of interspecific interactions, and the overlap of niches or dispersal limits among species (Table 1). There is a long tradition of inferring these mechanisms from spatial patterns of both CSA (Tilman and Kareiva 1997, Condit et al. 2000) and ISA (Forbes 1907, Cody and Diamond 1979, Gotelli et al. 2010, Calatayud et al. 2020). However, it is not always possible, nor useful, to infer causal mechanisms from emergent static patterns (Peters 1991, McGill and Nekola 2010, Blanchet et al. 2020).

Fortunately, documenting static patterns of spatial aggregation and association has value regardless of the generative mechanisms, such as in macroecological studies (Currie 2019) and for nature conservation and forecasts (Ladle and Whittaker 2011). For example, patterns of CSA underpin widely applied concepts such as species endemism, range size, rarity, and ultimately extinction risk (Hartley and Kunin 2003). Similarly, patterns of ISA have direct applications in classifications of communities to coenoses or biomes (Hoekstra et al. 2004), underpin the concepts of indicator and umbrella species (Roberge and Angelstam 2004), and can improve estimates of site-specific species pools (Carmona and Pärtel 2021) and predictions of species distributions (Harris 2016, Norberg et al. 2019). Further, even though it has not been perceived as such, ISA describes a unique facet of biodiversity, particularly when biodiversity is defined as “variation of life at all levels of biological organization” (Gaston and Spicer 2009). Even though biodiversity is typically measured “per site,” it can also be quantified in other ways, such as “per species” (e.g., a number of species associated with a given species; Dormann et al. 2009). However, the theoretical and empirical connections between “per species” associations and spatial patterns of “per site” biodiversity remain poorly understood.

Although ISA is implicit in spatial gradients of biodiversity (Table 1), only a handful of studies have attempted to document large-scale patterns of ISA empirically and explicitly (Gotelli et al. 2010, Lyons et al. 2016, Tóth et al. 2019, Calatayud et al. 2020), and ISA has been missing in recent high-profile studies of the ongoing biodiversity change (Millenium Ecosystem Assessment 2005, WWF International 2012, Newbold

Table 1. Brief summary of processes that generate CSA and ISA, examples of biodiversity patterns that are linked to CSA and ISA, and the approaches to measure them.

| Question | Conspecific spatial aggregation (CSA) | Interspecific spatial association (ISA) |
|---|---|---|
| What generates it? | species-specific dispersal limits, *conspecific interactions* among individuals, interplay between *niche width*, and spatial aggregation of environmental conditions | dispersal limits common to two or more species, *interspecific interactions* among individuals, interplay between *niche overlap*, and aggregation of environmental conditions |
| Examples of biodiversity facets and patterns sensitive to it | beta diversity and its distance decay, species–accumulation curves, species–area relationships (Storch et al., 2008, McGlinn et al. 2019) | spatial gradients of species richness, assembly patterns (e.g., nestedness, Lewinsohn et al. 2006, Presley et al. 2010) |
| Approaches to measure it | spatial over- or underdispersion using Poisson or Negative Binomial distribution (He and Legendre 2002), spatial Taylor’s law (Taylor 1961), occupancy–area relationship (OAR), and fractal dimension (Kunin 1998), univariate pair correlation function (Wiegand and Moloney 2014) | this paper |

et al. 2015, Blowes et al. 2019, IPBES 2019). To investigate this potential knowledge gap more systematically, we conducted two literature searches (Appendix S2). First, we scanned Clarivate Web of Science for all papers mentioning biodiversity in their title or abstract; among these, we looked at frequency of terms associated with ISA and terms associated with other ecological patterns and processes (Appendix S2). We found that this literature has most often focused on quantities at sites (e.g., species richness, functional and phylogenetic diversity), or variation among sites (e.g., beta diversity). In contrast, biodiversity-related papers paid considerably less attention to ISA, measured by the total number of published papers per year (Appendix S2: Fig. S1). Second, we manually went through 3,856 abstracts of papers published between 1995 and 2019 in three journals with a long history of publishing biodiversity studies: *American Naturalist*, *Ecology*, and *Ecography* (details are in Appendix S2). We found that ISA is a marginal topic when compared to patterns of CSA, beta diversity, and community composition, as well as processes such as biotic interactions (Appendix S2: Fig. S2). Since measurement of ISA requires exactly the same data as analysis of beta diversity, the lack of attention cannot be explained by a lack of data. We propose three reasons for this gap. First, there is a variety of approaches for measuring ISA, with little guidance on the advantages and disadvantages of each approach. Second, it is not always clear why ISA may be important or useful when studying biodiversity. Third, little is known about the theoretical connections, or lack thereof, between ISA and other biodiversity patterns.

In this paper, we first review the current methods to measure ISA. We sort the methods according to the types of data that ecologists are likely to encounter, and we compare the methods on empirical data sets and on simulated pairs of species with known ISA. Next, we address why ISA can be an important facet of biodiversity, and we outline the theoretical links between ISA and other fundamental facets of biodiversity. We offer guidelines to select the appropriate approach to analyzing ISA based on the question of interest and particular type of data at hand. We hope that our overview will stimulate the study of ISA in biodiversity assessments alongside the more traditional spatial measures.

## Schools of Thought

Measurement and analysis of ISA have been approached by several schools of thought, with each of these typically working with a particular data type (Fig. 2) and research agenda. While each of these schools of thought has generated conceptually diverse measures of ISA, there has been little cross-fertilization. The first school uses non-spatial information on presence/absence (or abundance) of species in different sites and has focused on tests of hypotheses related to assembly of ecological communities (Cody and Diamond 1979) by comparing the observed patterns of ISA with simulations of null models (Gotelli and Graves 1996, Ulrich and Gotelli 2010, Sanderson and Pimm 2015). The second school uses similar non-spatial metrics, but has been more descriptive, identifying groups of species that are frequently observed together, a typical goal in phytocoenology (Braun-Blanquet 1964). It relies on methods such as ordinations and cluster analysis (ter Braak 1987, Šmilauer and Lepš 2014). The third school comes from the field of geostatistics (Cressie 2010) and uses the spatial position of the sites to show how covariance between two species changes with spatial distance (Wagner 2003). The fourth school is the analysis of spatial point patterns (Wiegand and Moloney 2014, Baddeley et al. 2015), which requires the most detailed data on the spatial positions of individuals, and offers the richest inference. Below we review the specific measures coming from these diverse schools of thought.

## Spatially Implicit Measures of ISA

We begin our review with the simplest and oldest (Forbes 1907) way to measure ISA using indices that quantify, in a single number, the association between two or more species, irrespective of the spatial distance among sites. This is also the most widely used approach, since it is applicable to any data that can be expressed as a *community matrix* **Y**, which describes the distribution of $S$ species (rows) over $n$ sites (columns; e.g., Gotelli 2000), but others may transpose these. Each element $y_{ij}$ contains either binary incidence or some measure of abundance, and $i \in 1{:}S$ and $j \in 1{:}n$. Since ISA measures that only use **Y** consider no information on the spatial location of the sites, the approach is *spatially implicit*.

To date, nearly 100 indices have been proposed to capture ISA in binary co-occurrence data (Hubálek 1982, Rajagopalan and Robb 2005, Legendre and Legendre 2012, Ulrich and Gotelli 2013, Arita 2017) and for abundance data (Legendre and Legendre 2012, Legendre and De Cáceres 2013), although many fewer are widely used in ecology. Some of the pairwise indices were developed specifically to capture ISA while others were adopted from the literature on beta diversity (Hubálek 1982, Legendre and Legendre 2012, Arita 2017). Most of the indices we describe are available through R packages, such as vegan (Oksanen et al. 2019), EcoSimR (Gotelli et al. 2015), bipartite (Dormann et al. 2008), and betapart (Baselga and Orme 2012), as listed in Tables 2, 3.

### *Pairwise associations*

The conceptually simplest approach is to analyze associations between binary occurrences of two species. Table 2 lists some of the typical pairwise indices of association for *binary data* that have been widely used, or that represent a unique approach to capturing ISA. They are based on four different quantities: the number of sites occupied uniquely by species 1 ($c$) and species 2 ($b$),

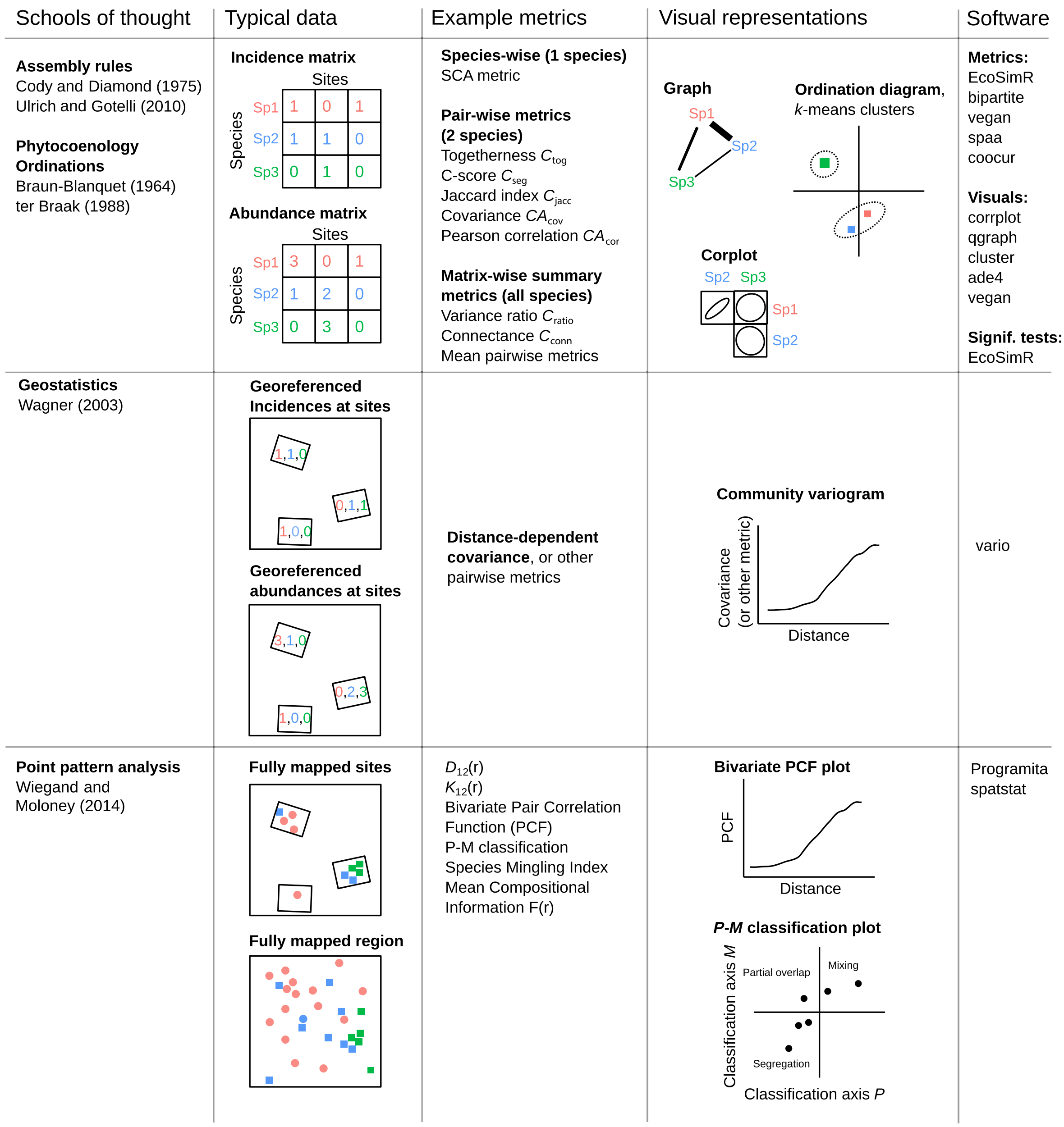


FIG. 2. Approaches to capturing inter-specific spatial association (ISA) among species, classified by the broad ecological schools of thought, together with typical data that are used in the approaches. Note that there is a gradient of the amount of spatial information (detail) in the data. With the exception of Programita, all the listed software are R packages.

the number of sites where both species co-occur ($a$), the number of sites where none occurs ($d$), and $n = a + b + c + d$. These indices vary in their interpretation; the widely used $C$-score ($C_{seg} = bc$), for example, gives the number of pairs of sites with a "checkerboard" pattern (i.e., one site hosts only species 1, the other only species 2), and it is a measure of segregation between two species. In contrast, togetherness ($C_{tog} = ad$) is a measure of attraction between species, but the $C_{seg}$ and $C_{tog}$ are not complements (i.e., low $C_{seg}$ does not imply high $C_{tog}$). Another widely used example is the Jaccard index (of association, not beta diversity), a measure of spatial overlap, which describes the proportion of sites occupied by both species among all occupied sites ($C_{jacc} = a/(a + b + c)$). Further, the indices vary in the range of possible values, such as [−1, 1], [0, 1], or [0, ∞). Thus, it is important to be familiar with both the interpretation and interval of these indices, particularly in their raw form (as opposed to their $Z$-score; Hubálek 1982, Legendre and Legendre 2012).

Table 3 lists some commonly used indices for pairwise ISA in *abundance data*. The first family includes both

Table 2. Measures of ISA for binary community data.

| Measure | Reference | Symbol | Formula or description | R code in |
|---|---|---|---|---|
| (A) Pairwise | | | | |
| *C*-score | Stone and Roberts (1990) | $C_{\text{seg}}$ | $bc$ | bipartite::C.score<br>EcoSimR::c_score<br>vegan::designdist |
| Scaled *C*-score | Ulrich and Gotelli (2013) | $C_{\text{segSc}}$ | $\frac{bc}{n(n-1)/2}$ | vegan::designdist |
| Togetherness | Stone and Roberts (1992) | $C_{\text{tog}}$ | $ad$ | bipartite::togetherness<br>vegan::designdist |
| Scaled togetherness | Ulrich and Gotelli (2013) | $C_{\text{togSc}}$ | $\frac{ad}{n(n-1)/2}$ | vegan::designdist |
| Jaccard similarity | Jaccard (1901) | $C_{\text{jacc}}$ | $\frac{a}{a+b+c}$ | betapart::betapart<br>vegan::betadiver |
| Dice-Sorensen similarity | Dice (1945),<br>Sørensen (1948) | $C_{\text{sor}}$ | $\frac{2a}{2a+b+c}$ | betapart::beta.pair<br>vegan::betadiver |
| Coefficient of association | Forbes (1907) | $C_{\text{forbes}}$ | $\frac{an}{(a+b)(a+c)}$ | vegan::designdist |
| Alroy's coefficient (a new take on $C_{\text{forbes}}$) | Alroy (2015) | $C_{\text{alroy}}$ | $\frac{a(z+\sqrt{z})}{(a+b)(a+c)+a\sqrt{z}+\frac{bc}{2}}$<br>where $z = a + b + c$ | vegan::designdist |
| Pearson tetrachoric correlation | Pearson and Heron (1913) | $C_{\text{pears}}$ | $\frac{ad-bc}{[(a+b)(c+d)(a+c)(b+d)]^{0.5}}$ | vegan::designdist |
| Mid-P variant of Fisher's exact test | Kallio et al. (2011),<br>Tóth et al. (2019) | $C_{\text{FETmP}}$ | $\left(\sum_{x_{\min} \leq x \leq a} \frac{\binom{i}{x}\binom{n-i}{j-x}}{\binom{n}{j}}\right) - \frac{\binom{i}{a}\binom{n-i}{c}}{2\binom{n}{j}}$<br>where $i = a + b, j = a + c$, and $x$ is the possible number of overlaps, and $x_{\min} = \max(a\text{-}d, 0)$ | https://github.com/anikobtoth/FCW |
| Matching coefficient | Sokal and Michener (1958) | $C_{\text{match}}$ | $\frac{a+d}{n}$ | vegan::designdist |
| (B) Matrix-wise | | | | |
| Variance ratio | Schluter (1984) | $C_{\text{ratio}}$ | $S_T^2/\sum\sigma^2$, where $S_T^2$ is the variance of per-site species richness and $\sigma^2$ is the sum of per-species variances of incidences. | EcoSimR::v_ratio<br>https://github.com/mcglinnlab/vario |
| Checker score | Gotelli (2000) | $C_{\text{checker}}$ | Number of species pairs forming perfect checkerboard distributions. | EcoSimR::checker |
| Number of unique species combinations | Gotelli (2000) | $C_{\text{combo}}$ | Number of species that always co-occur. | EcoSimR::species_combo |
| Network connectance | Dormann et al. (2009) | $C_{\text{conn}}$ | $\frac{F}{Sn}$, where $F$ is the number of all co-occurrences. | bipartite::networklevel |

*Notes:* From more than 70 measures (Hubálek 1982, Legendre and Legendre 2012, Keil 2019), we have included those that have either been popular, recommended based on suitable properties, or that represent a distinct approach to ISA. For pairwise indices, $a$ is the number of sites where both species co-occur, $c$ and $b$ are numbers of sites occupied uniquely by each species respectively, $d$ is the number of sites where none of the species occur, $n = a + b + c + d$. For matrix-wise indices, where $\bar{n}$ and $\bar{a}$ are the mean number of occupied sites per species the mean number of species per site, respectively, and $S$ and $n$ are numbers of all species and sites in the matrix, respectively.

parametric (e.g., $CA_{\text{cor}}$) and nonparametric (e.g., $CA_{\text{rho}}$) indices of covariance and correlation. Another family includes indices that are extensions of the incidence-based similarity measures; an example is Růžička similarity $CA_{\text{ruz}}$, which is one of the abundance-based versions of Jaccard's index for binary data (Legendre and De Cáceres 2013).

### *Higher-order, matrix-wise, and per-species indices*

When the aim is to quantify the overall magnitude of ISA in the entire matrix **Y** in a single number, one option is to simply use the mean or median of all pairwise associations (or only negative or only positive associations) in the species-by-species *association matrix* **Z**. However, focusing only on pairwise comparisons ignores interactions of higher order (i.e., interactions between three or more species; Harris 2016). Some methods can capture *N*-wise species associations when applied to the rows of **Y** (Baselga 2017, Arita 2017). The problem with these indices is their sensitivity to double zeroes and it is still unclear how common the higher order associations (i.e., motifs; Milo 2002) are in the real world (Marion et al. 2017). Finally, Legendre and De Cáceres (2013)

TABLE 3. Select measures of ISA for abundance data.

| Measure | Reference | Symbol | Formula or description | Notes | R code in |
|---|---|---|---|---|---|
| (A) Pairwise | | | | | |
| Covariance | Legendre and Legendre (2012) | $CA_{cov}$, $CA_{cov\_hell}$ | $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$ | With optional transformation (Hellinger, log, sqrt) of raw abundances. | stats::cov |
| Pearson correlation (scaled covariance) | Legendre and Legendre (2012) | $CA_{cor}$, $CA_{cor\_hell}$ | $\frac{CA_{cov}(xy)}{\sigma_x\sigma_y}$ | With optional transformation (Hellinger, log, sqrt) of raw abundances. | stats::cor |
| Spearman's Rho | Legendre and Legendre (2012) | $CA_{rho}$ | $CA_{cor}$ between the rank values of $x$ and $y$ | | stats::cor |
| Chi-squared distance | Lebart and Fénelon (1971), Legendre and De Cáceres (2013) | $CA_{chi}$ | $\sqrt{(x_+ + y_+)\sum_{i=1}^{n}\frac{1}{x_i+y_i}\left(\frac{x_i}{x_+}-\frac{y_i}{y_+}\right)^2}$ | | vegan::decostand with stats::dist |
| Hellinger distance | Rao (1995), Legendre and De Cáceres (2013) | $CA_{hell}$ | $\sqrt{\sum_{i=1}^{n}\left(\sqrt{\frac{x_i}{x_+}}-\sqrt{\frac{y_i}{y_+}}\right)^2}$ | | vegan::decostand with stats::dist |
| Percentage difference (former "Bray-Curtis" index) | Odum (1950), Legendre and De Cáceres (2013) | $CA_{bray}$ | $\frac{\sum_{i=1}^{n}\lvert x_i-y_i\rvert}{x_+ + y_+}$ | One of the abundance-based variants of $C_{sor}$ | vegan::vegdist |
| Růžička similarity | Růžička (1958), Oksanen et al. (2019) | $CA_{ruz}$ | $\frac{2CA_{bray}}{1+CA_{bray}}$ | One of the abundance-based variants of $C_{jacc}$ | vegan::vegdist |
| (B) Matrix-wise | | | | | |
| N-wise Růžička | Baselga (2017) | $CA_{ruzN}$ | Long formula, see Baselga (2017) | Multi-species version of $CA_{ruz}$ | betapart::beta.multi.abund |
| Variance ratio | Ulrich and Gotelli (2010) | $CA_{ratio}$ | Same as $C_{ratio}$ (Table 2) | | |

*Notes:* From the plethora of existing measures, we have included those that have been popular, recommended, or that represent a distinct approach to ISA. Specifically, we selected three correlation-based indices, two distance-based indices, two abundance-based variants of the binary indices, and two matrix-wise indices. Here **x** and **y** are the vectors of abundances of two species, $\overline{x}$ and $\overline{y}$ are their means, $x_+$ and $y_+$ are their sums, $\sigma_x$ and $\sigma_y$ are their standard deviations, $x_i$ and $y_i$ are abundances at site $i$, and $n$ is the total number of sites.

proposed an index of contribution of a single locality to the overall beta diversity in **Y**. The same index can be calculated for rows instead of columns of **Y**, thus becoming a measure of species contribution to overall association, but we are unaware of a study that employs such an approach.

### *Statistical significance, null models, Z-scores*

There are several ways to calculate the probability that the observed metric has been produced by chance given the null expectation that there is no ISA, i.e., that species are independent on each other. The simplest test of significance of association in a two-species binary matrix is Fisher's exact test (Arita 2016), which is identical to the more complex procedure of Veech (2013). However, when more than two species are involved, or when the data are abundances rather than incidences, a randomization procedure can produce the null expectation of a given ISA metric (see, e.g., Gotelli 2000, Ulrich and Gotelli 2010, Gotelli et al. 2015). Further, the deviation from the null expectation can also be used as its own ISA metric, such as the $C_{forbes}$ and $C_{FETmP}$ metrics (Table 2), and the so-called $Z$-score (Gotelli and McCabe 2002, Ulrich et al. 2009, Ulrich and Gotelli 2013), defined as $(E_{raw} - E_{exp})/SD_{exp}$, and where $E_{raw}$ is the ISA metric (e.g., from Table 2 or 3) calculated on observed data, $E_{exp}$ is the null expectation of the metric, and $SD_{exp}$ is the standard deviation of the null expectation.

## SPATIALLY EXPLICIT MEASURES OF ISA

Spatial structure between sites, individuals, and/or geographic ranges can be critical. Consider the simple example in Figure 3 in which the same community matrix represents either attraction or segregation. The aforementioned spatially implicit metrics of ISA conflate these patterns, and a spatially explicit approach needs to be adopted. When spatial coordinates of sites or individuals are available, one way to make the ISA spatially explicit is to only consider pairs of sites that are within a given distance interval when using any of the metrics from the previous section (Tables 2 and 3), and then to examine a range of distances. This is also the idea behind community variograms (Wagner 2003) described below. A point pattern-based version of this, when the location and identity of each individual are known, is to use bivariate pair correlation functions, which describe positive and negative ISA between individuals over continuous distance (Wiegand and Moloney 2014). In the following section, we describe principles of these two approaches, and refer to software implementations.

### *Community variograms*

Community variograms, borrowed from geostatistics, use the species-by-site community matrix **Y** together with spatial coordinates of each site (Wagner 2003). A community variogram then expresses the species-by-species variance–covariance matrix $\mathbf{Z}(r)$ as a function of spatial distance or lag $r$ between sites and allows the user to analyze within-species (CSA) and between-species (ISA) covariances at a single distance and their change over different distances. The covariance calculated on binary data is closely algebraically tied to the $C_{\mathrm{seg}}$ or $C_{\mathrm{tog}}$ metrics and their scaled versions (Table 2), which summarize only negative or positive covariances respectively. Apart from covariance, many of the metrics from Tables 2 and 3 can be made distance-dependent by using the community variogram framework.

To test for non-random patterns of spatial species covariance an appropriate null model is required. The distance-dependent $C_{\mathrm{ratio}}$ is typically used with a null expectation of $C_{\mathrm{ratio}} = 1$ under species independence, which can be biased in a number of different ways (Palmer and van der Maarel 1995). In particular, it can be biased by patterns of within-species clumping. Even if species are arranged independently on a landscape, strong patterns of CSA will create the appearance of spatial segregation at least up to the scale of the within-species clump size. Therefore, the most common type of null model is one in which the within-species spatial clumping is held constant but otherwise species are shuffled randomly (Palmer and van der Maarel 1995, Roxburgh and Chesson 1998).

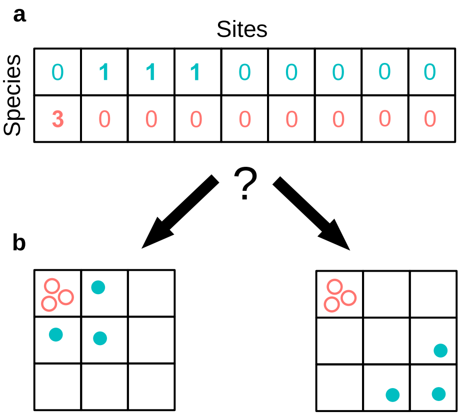


Fig. 3. The problem of disregarding the importance of spatial distance in ISA. (a) A single community matrix can potentially reflect (b) two different spatial arrangements of the community, each with a different magnitude of ISA; the two species in the left spatial matrix are more attracted, while on the right they are more segregated, yet this spatial arrangement is not reflected in the community matrix.

Even though the idea of community variograms has been featured in prominent methodological reviews (Dray et al. 2012), and elements of it occasionally appear in empirical analyses (Wagner et al. 2005, Seabloom et al. 2005, Kikvidze et al. 2005, Ovaskainen et al. 2017), we are unaware of its direct use in estimating distance-dependent ISA. At the same time, the method is close to the truly spatially explicit description of ISA for community matrices with additional spatial coordinates, and we thus see potentially important applications. Two R packages that allow calculation of community variograms are spacemakeR (Dray 2019) with the function variomultiv, and package vario with the function vario (*available online*).[8]

### *Point pattern analysis*

Point patterns capture facets of ISA when the position and identity of every individual within a spatial domain is known, offering the most accurate and spatially explicit quantification of ISA. However, such detailed data are also costly and thus relatively rare, and they have limits when applied to mobile organisms. Consequently, analyses of ISA based on point patterns are less common than analyses based on community matrices. Here, we focus on bivariate pair correlation functions and *P-M* classification scheme (Wiegand et al. 2007*b*, Wiegand and Moloney 2014).

*The bivariate pair correlation function* $g_{12}(r)$ measures the association between two species 1 and 2 at different spatial scales (Stoyan and Stoyan 1994, Wiegand et al. 2007*b*); it relates to the density of species 2 at distance $r$ of the individuals of species 1. Positive association (attraction) occurs for $g_{12}(r) > 1$ (larger than expected neighborhood density) and negative association (segregation) for $g_{12}(r) < 1$ (smaller than expected neighborhood density). Furthermore, the cumulative version of $g_{12}(r)$ is the $K$ function $K_{12}(r)$. Popular software to calculate $g_{12}(r)$ and $K_{12}(r)$ are Programita (Wiegand and Moloney 2004, 2014), and R package spatstat (pcfcross function; Baddeley et al. 2016).

*The P-M classification scheme* provides a way to further classify a spatial pairwise ISA as either independent, fully segregated, partially overlapping, or "mixing" at a given spatial lag $r$ using a combination of the $Z$-score transformed statistics of nearest neighbor occurrence ($P$) and neighborhood density ($M$; Getzin et al. 2014). This is best used to summarize the overall ISA structure at a given spatial scale (Getzin et al. 2014). It

REVIEW

---

[8] https://github.com/dmcglinn/vario

can be calculated using the Programita software (Wiegand and Moloney 2004, 2014).

*Null models for point patterns.* A well-developed toolbox is available to those seeking to compare observed ISA summary statistics [e.g., $g_{12}(r)$, $K_{12}(r)$] with those expected under the null expectations of independence (Wiegand and Moloney 2014). These null models are based on breaking the association between species, while keeping all of the other properties of single-species point patterns (e.g., CSA) intact. Here, we note that the heterogeneous point process null models (also combined with pattern reconstruction) can be useful when estimating biotic interaction from occurrence. Notably, this can be done without environmental data by assuming that the spatial range of biotic interactions is shorter than the scale of environmental autocorrelation (Wiegand et al. 2007*b*). Null models for point patterns are implemented in Programita software (Wiegand and Moloney 2004, 2014) and in the spatstat R package (Baddeley et al. 2016).

## COMPARISON OF THE MEASURES

In order to get a comprehensive basis for recommendations, we compare the performance of the approaches on a common set of empirical and simulated communities. Our aim is to assess redundancy among the approaches, as well as their sensitivity to the common variables such as number of sites, number of species, magnitude of conspecific aggregation, or spatial grain. We are also interested in how well the different measures capture negative associations (i.e., segregation) compared to positive associations (attraction).

### *Spatially implicit measures and empirical data*

Our goal in this exercise was to calculate the common metrics of ISA for a set of spatially implicit species-by-site matrices, and to see how the metrics correlate with each other. We evaluated the metrics in Table 2 using 476 empirical binary community matrices collated by Atmar and Patterson (1995) and Ulrich and Gotelli (2010), and the metrics in Table 3 using 186 empirical abundance matrices collated by Ulrich and Gotelli (2010). We excluded three binary matrices and 52 abundance matrices with too little information to allow meaningful calculation of all of the indices, or with negative or positive infinity values of some of the metrics. For the pairwise metrics, we averaged them over the species-by-species association matrix **Z** to obtain a single number comparable with the matrix-wise measures. We subjected the metrics to principal components analysis (PCA), for which we transformed some of them to have an approximately normal distribution. We plotted the PCA as an ordination biplot, and we also plotted a graph representation of the correlation matrix of the measures (Fig. 4). The correlation matrices are in Appendix S3.

In both the incidence- and abundance-based spatially implicit indices, we found clear differentiation along the PCA axes (Fig. 4a and b) and in the graph (Fig. 4c and d), reflecting the different aspects of ISA. In binary metrics, there was a clear cluster of similarity-based indices ($C_{\mathrm{jacc}}$, $C_{\mathrm{sor}}$), Alroy's index ($C_{\mathrm{alroy}}$), and connectance ($C_{\mathrm{conn}}$; Fig. 4a and c). Two measures that measure departure from Poisson randomness are $C_{\mathrm{FETmP}}$ and $C_{\mathrm{forbes}}$, which were grouped together. The $C$-score ($C_{\mathrm{segSc}}$) and its counterpart, togetherness ($C_{\mathrm{togSc}}$), the matching coefficient ($C_{\mathrm{match}}$), the checkerboard score ($C_{\mathrm{checker}}$), and number of unique combinations ($C_{\mathrm{combo}}$) were largely unrelated to the rest of the metrics. In the abundance-based metrics, we found similar clustering of the similarity-based ISA metrics ($\mathrm{CA}_{\mathrm{ruz}}$, $\mathrm{CA}_{\mathrm{bray}}$, $\mathrm{CA}_{\mathrm{chi}}$, $\mathrm{CA}_{\mathrm{hell}}$), while the correlation-based metrics ($\mathrm{CA}_{\mathrm{cor}}$, $\mathrm{CA}_{\mathrm{cor\_hell}}$, $\mathrm{CA}_{\mathrm{rho}}$) formed another group.

Among the incidence-based metrics, only $C_{\mathrm{combo}}$ was strongly correlated with the total number of sites in a matrix ($n$), while only $C_{\mathrm{checker}}$ correlated strongly with $S$, the total number of species (Fig. 4, Appendix S3). Covariance-based measures ($\mathrm{CA}_{\mathrm{cov}}$, $\mathrm{CA}_{\mathrm{cov\_hell}}$) grouped together with the total number of sites $n$, while most metrics correlated only weakly with $S$ (Fig. 4, Appendix S3).

### *Spatially implicit measures and spatially explicit simulations*

In the second exercise, we devised simulations with known magnitude of ISA between two species (Fig. 5, Appendix S4) to illustrate how the different approaches recover the known ISA. In contrast to spatially implicit simulations performed in other studies (Gotelli 2000, Ulrich and Gotelli 2010), we simulated the position of every individual within a bounded domain. In each simulation, we generated point patterns of two species, with a given magnitude of ISA between them, and with varying conspecific aggregation and number of individuals per species.

We modeled ISA as dependent on spatial distance (Fig. 5b), and it was controlled by a single parameter that we call ISA, with negative values for segregation (ISA $<$ 0), zero for independence (ISA = 0), and positive for attraction (ISA $>$ 0; Fig. 5). To calculate measures based on incidence or abundance, we converted the point patterns to grids of varying resolutions (grains). Across simulations, we measured performance of the metrics from Tables 2 and 3 as their Spearman correlation with the ISA parameter. We also examined how the performance was affected by spatial grain (Appendix S4: Fig. S3) and magnitude of conspecific spatial aggregation (CSA; Appendix S4: Fig. S4). Detailed description of the simulations, and their analysis, is in Appendix S4, and complete code is in R package spasm (see *Code and Data*).

We found that the best correlation with overall ISA, as well as correlation with both positive ISA (attraction) and negative ISA (repulsion), was exhibited by Pearson's

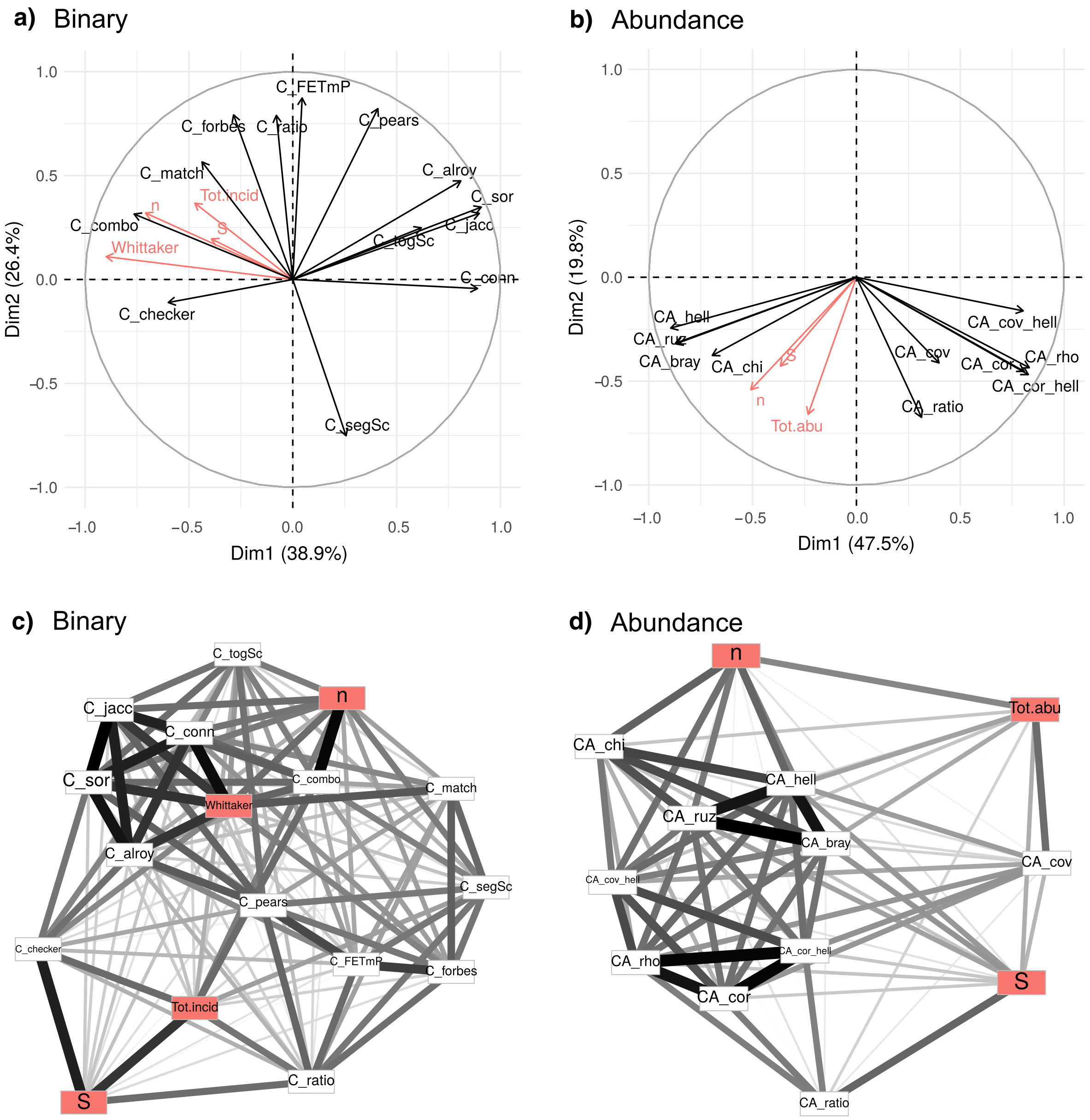


FIG. 4. Comparison of ISA metrics calculated on empirical community matrices of Atmar and Patterson (1995) and Ulrich and Gotelli (2010). Panels a and b use the binary measures from Table 2, and are based on binary (presence/absence) version of all matrices. Panels c and d use the abundance-based measures of Table 3, and use only the abundance matrices of Ulrich and Gotelli (2010). Panels a and b show the first two axes of principal components analysis (PCA), panels c and d show representations of correlation matrices between the metrics. Red indicates variables that are not ISA metrics (number of sites *n*, total number of species *S*, total number of incidences, total abundance of all individuals, Whittaker's index).

tetrachoric correlation ($C_{\text{pears}}$), Forbes index ($C_{\text{forbes}}$), and the mid-*P* variant of Fisher's Exact Test ($C_{\text{FETmP}}$) for binary data, and Spearman's correlation ($\text{CA}_{\text{rho}}$) and Hellinger distance ($\text{CA}_{\text{hell}}$) for abundance data (Fig. 6). The best overall Spearman correlations between the simulated ISA and its estimated measures were around 0.75 (Fig. 6), which we attribute to the inability of the spatially implicit measures to capture the distance-dependent part of ISA (Fig. 3). We found no clear indication that either abundance or incidence-based metrics performed best. We also found that measures based on correlation ($C_{\text{pears}}$, $\text{CA}_{\text{rho}}$, $\text{CA}_{\text{cor}}$, $\text{CA}_{\text{cov}}$), or on deviation from an expected null association ($C_{\text{forbes}}$, $C_{\text{FETmP}}$) performed better in capturing ISA, while similarity measures ($C_{\text{sor}}$, $C_{\text{jacc}}$, $\text{CA}_{\text{bray}}$, $\text{CA}_{\text{ruz}}$) performed worse. Refining the spatial grain of the analysis, and increasing the magnitude of CSA, had

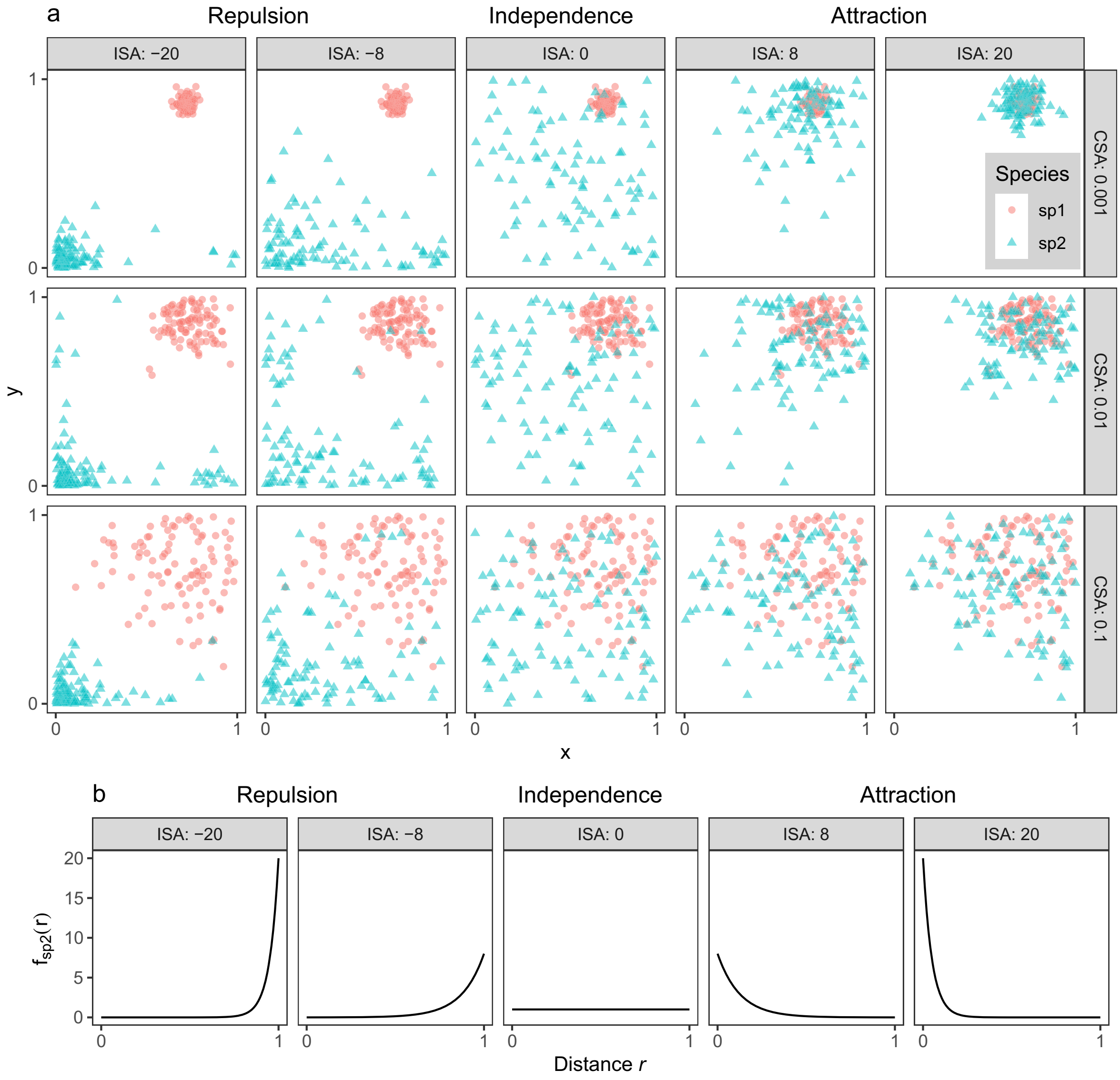


FIG. 5. (a) Simulated spatial distributions of individuals (points) of two species (sp1 and sp2) in a square domain under three levels of conspecific aggregation (CSA) of sp1, and five levels of interspecific association (ISA). (b) Truncated exponential probability density function [$f_{sp2}(r)$ (Keil 2014, 2019)] describes how likely we are to observe an individual of sp2 at a given distance from any individual of sp1. The shape of $f_{sp2}(r)$ depends on a single parameter (here called ISA) that represents various magnitudes of interspecific repulsion (left) and attraction (right), and their relationship with distance.

mostly negative or no effect on the performance of the metrics (Appendix S4: Figs. S3 and S4), although there were exceptions.

Importantly, we found that the spatially implicit metrics tended to saturate at extreme negative (ISA < 10) association (Appendix S4: Fig. S2), most likely because the community matrices are identical above these extreme ISA values (demonstrated in Fig. 3). Thus, most abundance-based metrics showed weak overall correlation with repulsion (Fig. 6b), with the exception of the distance-based Růžička ($CA_{ruz}$) and percentage difference ($CA_{bray}$) indices.

### *Spatially explicit measures and spatially explicit simulations*

Here, we used the two-species simulations to illustrate community variograms and pair correlation functions. Our goal was to show the potential of these methods, rather than to perform the same comprehensive evaluation as above, since these techniques do not measure ISA in a single number, but as a function of spatial distance *r*. Because of that, they have the potential to recover the shape of the entire probability density function $f_{sp2}(r)$ from Fig. 5b. For the spatially explicit

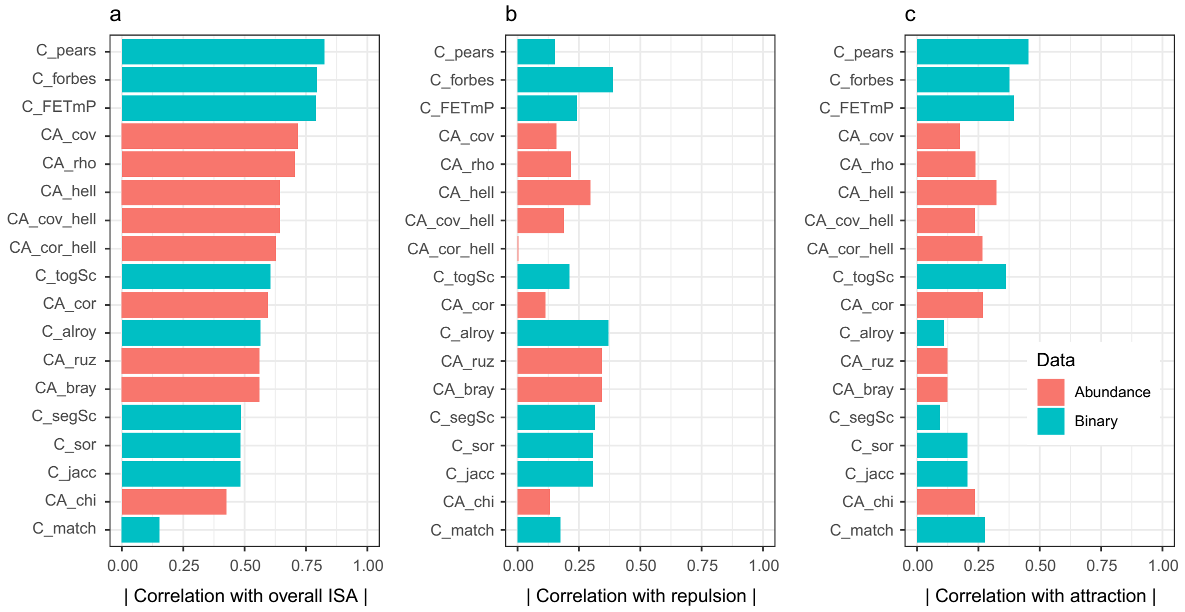


Fig. 6. Ability of pairwise spatially implicit metrics from Tables 2 and 3 to recover the magnitude of ISA from spatially explicit simulations. Panel a shows overall correlation between the true ISA and the metric. Panels b and c show the correlation when only inter-specific repulsion is considered (ISA $< 0$) or when only positive attraction is considered (ISA $> 0$). Absolute values of Spearman correlation coefficients are shown. Note that some metrics separate negative from positive associations very well (i.e., they have good correlation with overall ISA), but within these two categories they have weak correlation with the underlaying ISA parameter, e.g., covariance-based measures ($\mathrm{CA}_{\mathrm{cov}}$, $\mathrm{CA}_{\mathrm{cov\_hell}}$). See Appendix S4: Fig. S2 for details of these relationships.

techniques, we performed nine simulations that correspond to the panels of Fig. 5a (the only difference is that we used 200 individuals per species). For the purpose of fitting the community variograms, we used a single grid resolution with 20 cells along each side of the simulated domain.

Community variograms (Fig. 7a) estimated the relationship between distance $r$ and negative covariance in a way that matched the shape of the original $f_{\mathrm{sp2}}(r)$ from Fig. 5b. However, they were weak at capturing the spatial pattern of attraction when there was a confounding effect of strong conspecific dispersion (CSA $> 0.01$). In contrast to community variograms, pair correlation functions more accurately estimated the distance-dependent ISA (compare Fig. 5b with Fig. 7b), i.e., the estimated relationship between distance and pair correlation function closely matches the shape of $f_{\mathrm{sp2}}(r)$ from Fig. 5b. Similarly to community variograms, there was a somewhat limited performance for attraction and high CSA values, but the problem was less severe. Overall, the point pattern-based bivariate pair correlation functions provided the most complete picture of the "true" simulated ISA pattern.

## RECOMMENDATIONS

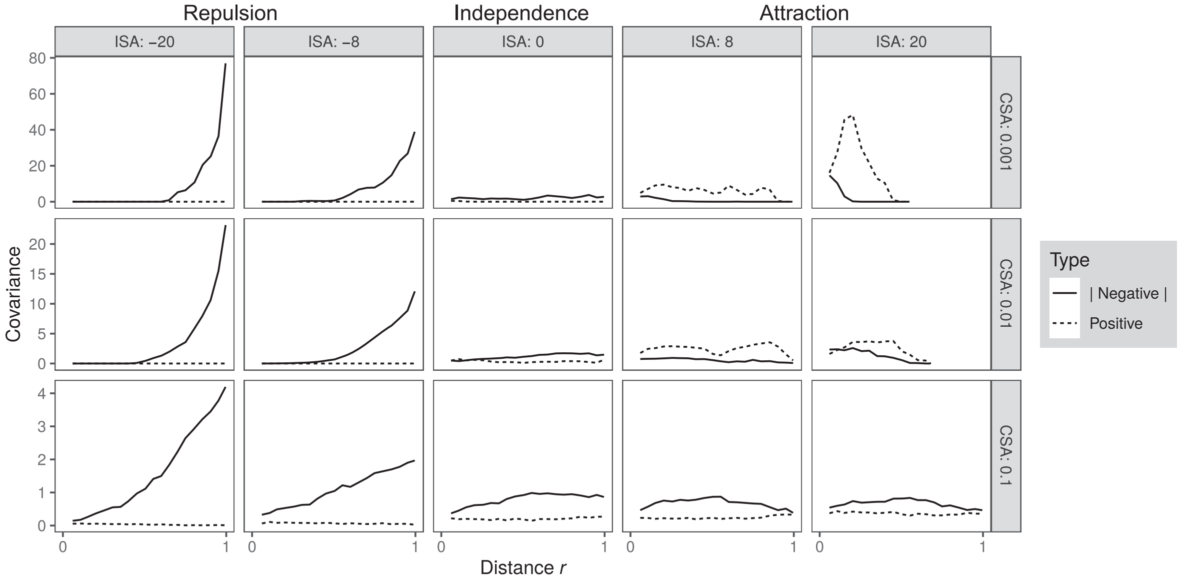
Based on our comparisons of the methods and on the existing literature, we make several recommendations for capturing ISA from observational data.

### *Best indices*

There is a variety of indices that were designed to capture different aspects of ISA and some of them are uncorrelated with the others (see our results, but also Hubálek [1982]). Thus, in empirical assessments, we recommend not relying on a single metric (see also Ulrich and Gotelli [2013]). Overall, we suggest that the most promising indices for spatially implicit analyses are those that can capture both negative and positive ISA. They should also offer the option to be eventually integrated into, or compared with, spatially explicit analyses such as community variograms, it should be possible to contrast them with indices of conspecific aggregation, and they should be applicable in association matrices of parametric joint species distribution models (Ovaskainen et al. 2017). For abundance data, these are covariance ($\mathrm{CA}_{\mathrm{cov}}$) and Pearson correlation ($\mathrm{CA}_{\mathrm{cor}}$), coupled with a data transformation (e.g., Hellinger or log) if appropriate (Legendre and Legendre 2012). For incidence data, these are the *C*-score ($C_{\mathrm{seg}}$, $C_{\mathrm{segSc}}$) and togetherness ($C_{\mathrm{tog}}$, $C_{\mathrm{togSc}}$), which capture the positive and negative ISA, respectively, and are mathematically linked to covariance. We caution that that low *C*-score values do not indicate high togetherness, and vice versa. Thus, both measures should typically be employed.

From the other indices for pairwise binary data, among the all-purpose metrics that capture both repulsion and attraction, we recommend the mid-*P* index

**a)** Community variograms



**b)** Bivariate pair correlation functions

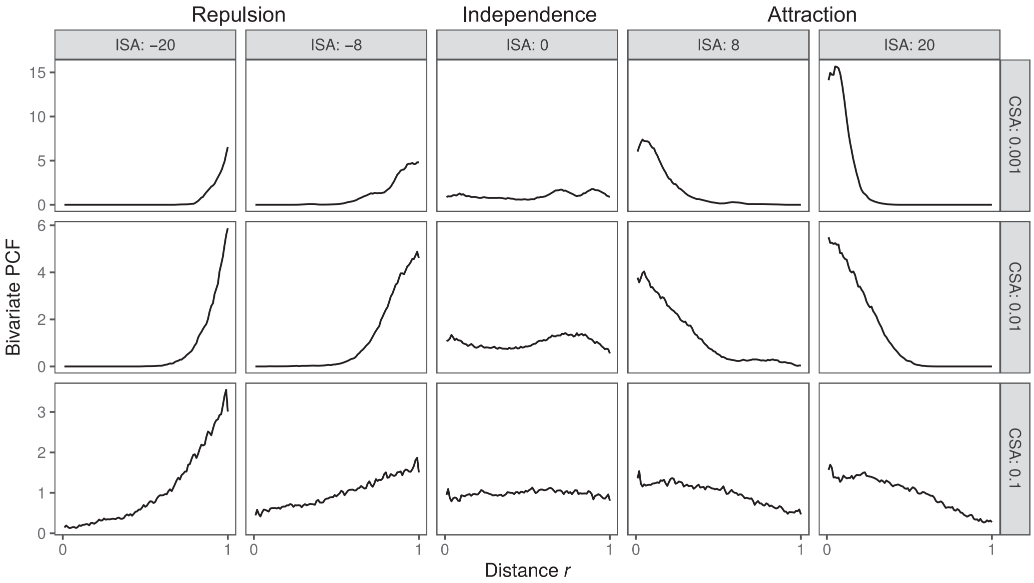


FIG. 7. Ability of two spatially explicit approaches to recover ISA patterns of the two species from Fig. 4b. (a) Community variograms calculated on abundance data obtained by aggregation of the point pattern in a 20 × 20 pixel grid. Also, 200 points were simulated for each species, instead of the 100 points in Fig. 5. (b) Bivariate pair correlation function calculated directly from the point patterns.

($C_{\mathrm{FETmP}}$) and Forbes's coefficient of association ($C_{\mathrm{forbes}}$), both of which have the advantage of explicitly quantifying the deviation of the observed ISA from the ISA expected under independent distribution of incidences. Based on its performance, as well as ability to capture repulsion as well as attraction, we also recommend Pearson's tetrachoric correlation ($C_{\mathrm{pears}}$), which is also recommended by Hubálek (1982). Jaccard's index ($C_{\mathrm{jacc}}$) is a good alternative, since it captures both negative and positive ISA, and it can easily be interpreted as proportional overlap, although it may not reflect the overall ISA as accurately as the correlation-based indices. For pairwise abundance data, apart from covariance and correlation-based parametric indices, we advocate for the Spearman rank-rank correlation ($\mathrm{CA}_{\mathrm{rho}}$), since it requires no prior transformation, and it captures both negative and positive ISA reliably.

Concerning the community-wide measures that quantify the magnitude of ISA for all species in a single number, an obvious choice is to use summarized pairwise metrics, although one needs to beware that some summarizations (e.g., averaging) blend repulsions and attractions. One way to avoid this problem is to decompose the pairwise association matrix to positive vs negative ISA, by separately summarizing positive vs negative covariances in abundance data (this is also the approach used in community variograms), or by considering both togetherness and *C*-score in analyses of incidence data.

### *ISA is better spatially explicit*

Our results highlight a serious limitation of spatially implicit indices of ISA; they fail to consider spatial proximity of two species when they are already 100% spatially segregated (Fig. 3). The problem becomes more severe toward finer spatial resolutions. This partly explains why none of the examined spatially implicit metrics gave a perfect correlation with the simulated distance-dependent ISA, and why we observed the saturation of performance toward extreme values of ISA in our simulations (Appendix S4: Fig. S2). To address this systematically, we recommend that whenever there is information on spatial position of the sites, indices of ISA are better considered as a function of spatial distance. Both the community variograms and point pattern analysis offer straightforward ways to do that.

Going spatially explicit (when data permit) also makes ISA more biologically interpretable, since it can identify spatial distances over which biotic interactions really matter. This can be done, for example, by integrating community variograms into joint species distribution models (JSDM; Warton et al. 2015, Ovaskainen et al. 2017, Zurell et al. 2018). Specifically, in a JSDM, the spatially implicit species-by-species covariance (or correlation) matrix $\mathbf{Z}$ can be replaced by distance-dependent $\mathbf{Z}(r)$, as also mentioned by Ovaskainen et al. (2017). Interestingly, however, the spatially explicit approach can also separate ISA caused by short-distance interactions from ISA caused by the environment *without* the need of modeling the effect of the environment, as demonstrated by Wiegand et al. (2012). The trick is to use null models that only randomize individuals locally, i.e., only within neighborhoods with radius $R$, which removes potential signals of small-scale interactions at scales $r < R$, if we can reasonably assume that environmental conditions within the neighborhood are constant (Wiegand et al. 2012).

### *Spatial scale*

Most of the commonly studied facets of biodiversity depend on spatial scale. In the case of ISA, scale has been approached from two angles. The first focuses on ISA as a function of the average area of a site in a community matrix, which is equivalent to spatial resolution (grain) of a grid (Äkland 1994, Hui 2009, Segurado et al. 2012, Araújo and Rozenfeld 2013, McNickle et al. 2018). These studies show that ISA is grain-dependent, but neither theory nor empirical observations predict a systematic direction of the grain dependence; ISA-area relationships can be increasing, decreasing, or hump-shaped. One issue that complicates consensus in this context is the use of various metrics of ISA across studies. Our simulations (Appendix S4: Fig. S3) show that selection of the ISA metric affects the grain dependency; for example, some metrics may have slightly hump-shaped relationship with grain (e.g., *C*-score, Appendix S4: Fig. S3, see also McNickle et al. [2018]), while other metrics may monotonically increase or decrease with grain (Appendix S4: Fig. S3). This is something that the field needs to sort out before any empirical scaling of ISA is explored and interpreted.

The second approach to spatial scaling of ISA uses distance instead of grain in the form of community variograms or bivariate pair correlation functions (Wagner 2003, Wiegand and Moloney 2014). One advantage of this approach is its straightforward biological interpretation, particularly in the context of biotic interactions. Both the grain-based and distance-based approaches to the scaling of ISA can be combined in a single analysis (see Wiegand et al. 2012). Finally, point-pattern analysis has theory that links the area-based with the distance-based approaches to ISA (Wiegand and Moloney 2014). We thus recommend that future investigations of ISA should explicitly embrace the issue of scale (both the grain and perhaps also extent) by focusing on the ISA–-area relationships (Araújo and Rozenfeld 2013, McNickle et al. 2018), or by focusing on the spatially explicit approaches to ISA, which handle scale more naturally than the spatially implicit ones.

### *Null models and Z-scores*

Most of the literature on ISA emphasizes the importance of null models, be it spatially implicit (Ulrich and Gotelli 2013) or explicit (Wiegand and Moloney 2014). In Appendix S4: Figs. S3 and S4, we show that transforming ISA indices to *Z*-scores makes them more robust against the confounding effects of conspecific aggregation or varying resolution, while (Ulrich et al. 2018) show that null models also account for the confounding effect of total species richness ($S$). The same sentiment underlies Wiegand and Moloney (2014): to analyze spatially explicit patterns of ISA, one should contrast them against a null expectation of no ISA. However, null models can be computationally demanding, posing a problem for large biodiversity data; in such cases analytical shortcuts may be handy. For example, in spatially implicit methods, measures such as $C_{\mathrm{FETmP}}$ or $C_{\mathrm{forbes}}$ already have the deviation from the expected randomness implicit in their definition.

### *Measuring macroecological patterns of ISA*

To empirically document patterns of ISA, and their generality or variation over broad spatial and temporal scales, we need analytical approaches designed specifically to measure ISA over broad scales. From all of the reviewed approaches, we see community variograms as the most promising, since (1) community variograms are spatially explicit, offering rich biological interpretation; (2) community variograms are applicable to a wide range of data types, including structured site-by-species data sets, such as the U.S. North American Breeding Bird Survey (Sauer et al. 2017), data sets of metacommunity-level pattern and process (e.g., the CESTES database; Jeliazkov 2019), as well as classical large-scale data such as the IUCN Red List biodiversity data; and (3) calculation of community variograms is computationally feasible.

Given that there is an ever increasing availability of biodiversity data from large spatial extents (Jetz et al. 2012), we see an exciting opportunity here for exploration of empirical macroecological patterns of ISA, particularly through community variograms. Some of the possible macroecological patterns to explore with community variograms are latitudinal or altitudinal gradients of ISA, and their relationship with patterns such as distance decay of compositional similarity, with distance-dependent CSA, or with broad geographic patterns of species richness and rarity. This has not yet been done, yet bringing ISA to macroecology may provide new explanations for old patterns, for example through the hypothetical (albeit controversial) link between ISA and biotic interactions, or through a completely unexplored link between ISA and diversification.

## ISA IN THE CONTEXT OF BIODIVERSITY

Now that we have demonstrated the ways that ISA can be measured, we return to our earlier argument that ISA deserves more attention in biodiversity research. To study ISA in the context of biodiversity, one should be aware of how it does, or does not, relate to other biodiversity metrics. In the sections below, we will show that although there are biodiversity facets that are affected by ISA, such as pairwise measures of beta diversity, others are insensitive to ISA, including local and regional diversity, and their ratio (i.e., Whittaker's index [Whittaker 1960], and $\alpha$, $\beta$, and $\gamma$ diversity).

### *ISA and alpha diversity*

First, we show that local species richness at any given site can be sensitive to changes in ISA, while mean local richness ($\overline{\alpha}$ diversity) is insensitive to changes in ISA. This argument was also made by Plotkin et al. (2000) and Storch (2016), and is inherent in connections between occupancy and species richness (e.g., Šizling and Storch 2004, Arita 2017).

Let $\alpha_j$ be local species richness at a site $j$, where $j \in 1{:}n$, and where $n$ is the total number of sites within a given spatial domain. Let $O_i$ be the number of occupied sites (i.e., occupancy) by species $i$, where $i \in 1{:}S$, and where $S$ is the total number of species (i.e., gamma diversity) present among sites. When we manipulate ISA in the system but keep CSA constant, some values of $\alpha_j$ change as a result. For example, in Fig. 8a, a change from interspecific segregation to attraction results in corresponding changes in each site's $\alpha_j$ (one site gains species and two sites lose species).

Let us now consider the average local number of species, $\overline{\alpha}$, across all sites, which can also be calculated as the sum of prevalences $O_i/n$ (Šizling and Storch 2004) as

$$\overline{\alpha} = \frac{\sum_{j=1}^{n} \alpha_j}{n} = \sum_{i=1}^{S} \frac{O_i}{n}. \quad (1)$$

Importantly, $O_i$ and $O_i/n$ are spatially implicit, meaning that it does not matter which sites are occupied, or whether species are spatially attracted or segregated; as long as $O_i$ is constant, no rearrangement of occupied cells in space has an effect on $\overline{\alpha}$. That is, $\overline{\alpha}$ is insensitive to ISA. For example, consider the scenario in Fig. 8a in which there was a shift in the ISA from segregated to attracted. Because $O_i$, $n$, and $S$ remain constant, $\overline{\alpha}$ will not be affected by the change in ISA. Hence, an empirical assessment of biodiversity that summarizes net change of $\overline{\alpha}$ over many locations is blind to changes of ISA. Further, this insensitivity of $\overline{\alpha}$ propagates to species–area relationships and to species accumulation curves, as we mention below. Note, however, that $\overline{\alpha}$ is sensitive to changes in $O_i/n$, which is a measure of conspecific aggregation (CSA).

### *ISA and gamma diversity*

From a purely geometrical perspective, simply rearranging mutual positions (ISA) of species within a given spatial domain has no effect on total number of species $S$ (gamma diversity) in the domain (given that $n$ is constant). A more interesting question is what happens when the effect is reversed, such as how ISA changes when $S$ increases or decreases. This has biological implications, since $S$ limits the magnitude of ISA within a given domain, which can also limit biotic interactions. It also has methodological implications, since variation in $S$ across spatial domains can hinder direct comparisons of ISA within these domains, which may require statistical treatment (Ulrich et al. 2018). Several propositions about the relationship between ISA and $S$ have been made and some of them have been empirically tested.

The first was summarized by McGill (2010): as $S$ in the metacommunity increases, the weaker the associations will appear even in the presence of strong associations. This can be seen by considering that the total number of possible pairwise associations in the domain is $S(S - 1)/2$. If every species is significantly spatially associated with $k$ species in a symmetrical fashion, then the total number of significant spatial associations in the domain is $(Sk)/2$. Thus, for any given $k < (S - 1)$, the

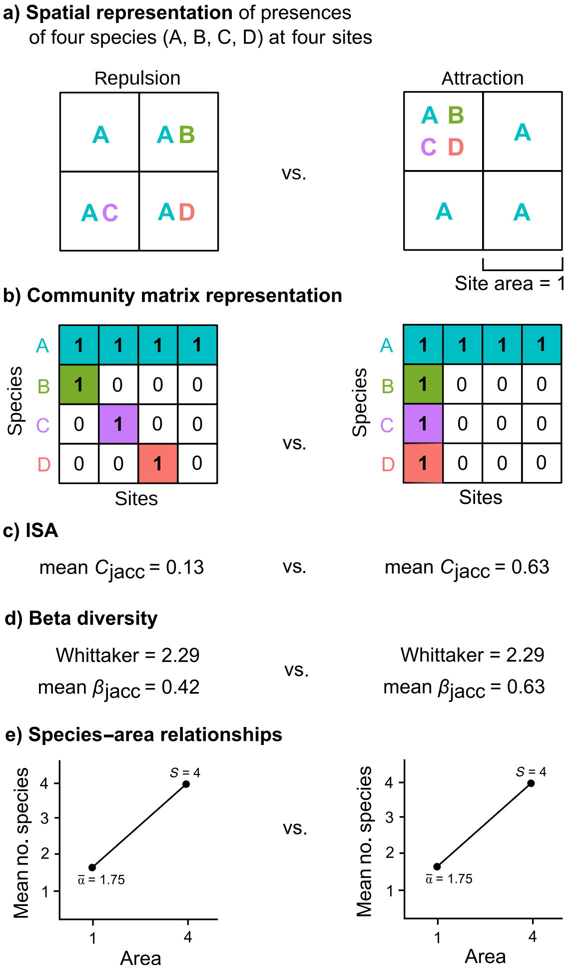


FIG. 8. Insensitivity of nested species–area relationships (SAR) and Whittaker's index to ISA, as also mentioned by Plotkin et al. (2000) and Storch (2016). (a) Two communities consist of four square sites each, with four species (A, B, C, D) either present or absent. These communities have constant CSA, but differ in the magnitude ISA; species in the left community are segregated, those in the right one are attracted to each other. (b) These communities can be described by spatially implicit community matrices **Y**. From **Y**, we can calculate (c) mean between-species $C_{\text{jacc}}$, (d) mean between-site Jaccard similarity $\beta_{\text{jacc}}$ and Whittaker index of beta diversity $s/\bar{\alpha}$, and (e) SAR. Both mean pairwise metrics ($C_{\text{jacc}}$ and $\beta_{\text{jacc}}$) are sensitive to varying ISA (c, d), while Whittaker index and the SAR slope remain constant even though ISA changes (e).

total number of all possible pairwise associations increases faster with S than the number of actually significant associations.

The second proposition operates on relative abundances. If the total area and density of individuals are constant, but there is an increase of the total $S$, then the average per-species abundance must decrease, which will reduce the likelihood of detecting inter-specific co-occurrences in communities (Hubbell and Foster 1986, Lieberman and Lieberman 2007, Wiegand et al. 2007*b*, 2012, Volkov et al. 2009, Rajala et al. 2019).

The third proposition exposes the mathematical constraint on the possible values of *negative* associations in a species-by-species association matrix (Brown et al. 2004). Simply put, if species A and B have strong negative association, then it is mathematically impossible for a third species C to have strong negative association with both A and B (Brown et al. 2004). Fox (2012) also showed that, under the assumption that all species have the same negative correlation with each other (e.g., due to competition-induced compensatory dynamics), the minimum possible value of the correlation approaches 0 as species richness increases. Thus, on average, the more competing species are added to the metacommunity, the weaker their observed average negative association gets.

The fourth proposition suggests that communities with more species may only be stable if the interactions get weaker on average, as found by May (1972) in an analysis of the stability of Lotka-Voltera type multispecies models. Stone (2016) generalized this proposition by showing that stability and feasibility under increasing species richness requires a reduction in the mean and standard deviation of the value of the interspecific competition coefficients.

Based on these arguments, we should expect the overall ISA to weaken as $S$ increases. Indeed, Wiegand et al. (2012) found that species associations were significantly weaker in rich forest communities compared to species poor ones, even after the null-expected associations were taken into account.

### *ISA and beta diversity*

Here we show how ISA is conceptually related to aspects of beta diversity (i.e., the differentiation of species composition in space, across sites). We also demonstrate how one particularly popular matrix-wise measure of beta diversity, Whittaker's index, is insensitive to ISA, while pairwise indices of beta diversity can be sensitive to ISA (Fig. 8). Although we still lack the exact mathematical theory for the latter, we show that point pattern analysis may offer the right toolbox to build such a theory.

The connection between ISA and beta-diversity is best illustrated on pairwise metrics applied to a spatially implicit site-by-species community matrix **Y**. In short, ISA is the association among *species* (Simberloff and Connor 1979, Hubálek 1982, Bell 2005, Legendre and Legendre 2012), whereas beta diversity is the similarity among *sites*. In this simplified case, both ISA and beta diversity are calculated using exactly the same data and indices (Tables 2, 3); the only difference between them is whether they are applied to the rows ("R-mode" of Legendre and Legendre 2012) or columns of **Y** ("Q-mode"). In other words, any index of beta diversity that is normally applied to sites can be applied to species and can be meaningfully interpreted as an index of ISA, and vice versa (Legendre and Legendre 2012, Arita 2017). For example, Araújo and Rozenfeld (2013) define a "co-occurrence score" as the "ratio of the number of geographical cells where species A and B co-occur to the total number of occupied cells." Although not stated, this is equivalent to the classic Jaccard index for comparing site similarity. Thus, the list of measures that have been typically used to measure ISA (Tables 2, 3) can be expanded with Jaccard or Sørensen-type indices (Arita 2017). Inversely, the list of commonly used measures of beta diversity (Koleff et al. 2003) can be expanded by some typical ISA indices. One exception is the Whittaker index ($s/\overline{\alpha} = n/\overline{n}$, where $\overline{n}$ is the average number of occupied sites per species), which is identical for both the analysis by sites ($s/\overline{\alpha}$) and by species ($n/\overline{n}$); however, this index is different from the pairwise indices of both beta diversity and ISA since it does not capture the within-matrix similarity between sites or species; instead, it only reflects the proportional matrix fill, i.e., the fraction of cells in $\mathbf{Y}$ filled with 1s (Arita 2017).

Little has been written about the actual relationship between beta diversity and ISA within a given spatial domain, i.e., given constant $n$, $S$, and CSA. In other words, what happens with beta diversity if we vary ISA and keep everything else intact? We know that Whittaker's index ($s/\overline{\alpha} = n/\overline{n}$) must be insensitive to ISA, which is demonstrated in Fig. 8, and which follows from the aforementioned insensitivity of $\overline{\alpha}$ to ISA (Eq. 1). However, we are unaware of any study directly focusing on the link between ISA and pairwise beta diversity. In Fig. 8, we show a case of average pairwise between-site Jaccard beta diversity being *sensitive* to ISA, and this is new. Further, in Appendix S5: Fig. S1 we show that average pairwise between-site Jaccard beta is correlated with average between-species $C_{\text{jacc}}$, albeit imperfectly. Šizling et al. (2011) and McGlinn and Hurlbert (2012) hint on a potential explanation by showing the relationship between average pairwise Jaccard beta diversity and the Whittaker's index, which is modulated by the occupancy frequency distribution (see also McGeoch and Gaston [2002]). This reasoning could perhaps be extended to provide a link between pairwise beta diversity and ISA.

In contrast to the spatially implicit indices for binary and abundance data, the connection between ISA and beta diversity is well known in analyses of point patterns (Wiegand and Moloney 2014). The ISA-beta connection can be demonstrated in the spatially-explicit version of Simpson's evenness index $\beta(r)$ (Shimatani 2001, Wiegand and Moloney 2014: section 3.1.5.1). Unlike the traditional spatially implicit version of the Simpson's index (Simpson 1949; i.e., the probability that two randomly selected individuals are heterospecifics), which is a measure of evenness, $\beta(r)$ is a measure of beta diversity, since it captures dissimilarity over a given distance (Shimatani 2001; i.e., the probability that two randomly selected individuals distance $r$ apart are heterospecifics). The index is defined as

$$\beta(r) = \sum_{i=1}^{S} \sum_{\substack{j=1\\ j\neq i}}^{S} f_i f_j \frac{g_{ij}(r)}{g(r)} = 1 - \sum_{m=1}^{S} \frac{f_m^2 g_{mm}(r)}{g(r)}. \tag{2}$$

Note the two alternative but equivalent definitions. In the first definition in Eq. 2, $f_i$ and $f_j$ are the relative abundances of species $i$ and $j$, $g_{ij}(r)$ is the bivariate pair correlation function describing the spatially explicit ISA of species $i$ and $j$ and $g(r)$ is the pair correlation function of all individuals in the community, regardless of species identity. As expected, if there are no spatial patterns of ISA (i.e., when $g_{ij}(r) = 1$), we obtain the non-spatial Simpson index, and depending on the abundances and ISA of the different species, beta diversity will be larger or smaller than this point of reference. The second definition in Eq. 2 operates purely with con-specific aggregation (CSA), measured by within-species pair correlation function $g_{mm}(r)$. We can see that the spatially explicit $\beta(r)$ depends on the balance between the ISA and CSA, whose overall effect sums up to 1.

We thus conclude that point pattern analysis, through $\beta(r)$, offers a comprehensive framework that can link abundances, CSA, ISA, gamma diversity, and alpha diversity, each with an exactly defined and mathematically tractable metrics. Not only does it stress the importance of making all of the diversity facets spatially explicit, but it also potentially offers a roadmap for future unification macroecology that deals with spatially implicit data on abundances or incidences.

### *ISA, species–area relationships, and species–accumulation curves*

Here, we demonstrate that species–area relationships and rarefaction curves are insensitive to ISA. We have stated that $\overline{\alpha}$ and $S$ in a given domain are insensitive to ISA, given that spatial extent is constant. These are the two components of Whittaker's index of beta diversity (Whittaker 1960), which is $S/\overline{\alpha}$, and so Whittaker's index is insensitive to ISA. It means that nested species–area relationships (SAR), which are exactly related the Whittaker index over a continuous range of $\overline{\alpha}$ and $S$ (Crist and Veech 2006) must also be insensitive to ISA (Fig. 2).

When every individual's spatial position and identity is known, point pattern analysis also makes it clear that there is no direct link between ISA and SAR. The relevant equation is (Shimatani and Kubota 2004)

$$S(r) = \sum_{i=1}^{\gamma} H_i(r) \tag{3}$$

where $S(r)$ is number of species present within $r$ from an arbitrarily chosen "test" location, $H_i$ is the spherical contact distribution function for species $i$, which is the probability that the first neighbor of species $i$ is distance $r$ away from the test location. $S(r)$ becomes a species–area curve when $r$ is converted to $\pi r^2$. Importantly, the $H_i$ is insensitive to ISA, since it is only based on the locations of species $i$. We note that point pattern analysis also has a scaling curve that is sensitive to ISA: the individual species–area relationship (ISAR; Wiegand et al. 2007), which quantifies the species richness in neighborhoods within radius $r$ of a focal species $f$

$$\mathrm{ISAR}_f(r) = \sum_{i=1}^{\gamma} D_{fi}(r) \qquad (4)$$

where $D_{fi}(r)$ is the bivariate nearest neighbor distribution function (i.e., the probability that the nearest point of species $i$ is distance $r$ away from an average point of focal species $f$). Again, $r$ can be converted to area as $\pi r^2$.

Finally, we turn to species–accumulation curves, from which the classical examples are the spatially implicit individual-based and sample-based rarefaction curves (Gotelli and Colwell 2001), and their spatially explicit versions (McGlinn et al. 2019). In the former, individuals or samples are accumulated randomly, irrespectively to their spatial position, which effectively breaks any pattern of both CSA or ISA, making the spatially implicit curves indeed insensitive to ISA. In the latter spatially explicit case, samples (plots) are accumulated by nearest neighbors, which makes these curves closely related to SARs, which we have shown to be sensitive to CSA, but insensitive to ISA. This is in line with the core idea of partitioning of rarefaction curves to their components (McGlinn et al. 2019), which are the regional species–-abundance distribution, density of individuals, and conspecific aggregation, but not ISA.

## WHY SHOULD BIODIVERSITY SCIENTISTS CARE ABOUT ISA?

Why, in the context of biodiversity, should we consider ISA patterns in space and time? And why should we care about ISA, when we have just demonstrated that many of the key biodiversity metrics are not affected by it? We argue that if we aim to describe a more complete picture of the multi-faceted nature of biodiversity, we need to consider approaches designed specifically to capture ISA, precisely because it is not captured by the traditional measures. Any biodiversity assessment that relies only on simple *per-site* measures of diversity and composition runs into a risk of missing variation, or temporal change, in ISA. Below we give more specific reasons for why capturing ISA as a facet of biodiversity may be useful.

### *ISA as evidence for interactions*

Perhaps the best known, albeit perennially controversial, reason for analyzing patterns of ISA is a notion that they give hints about biotic interactions among species (Gotelli et al. 2010, Blois et al. 2014, Harris 2016, Morueta-Holme et al. 2016, Thurman et al. 2019, Calatayud et al. 2020). The effort that gained traction in the 1980s and 1990s (Cody and Diamond 1979, Connor and Simberloff 1979), and has recently been revived with the promise of joint species distribution models as a tool to disentangle interactions from shared environmental requirements among species (Warton et al. 2015, Ovaskainen et al. 2017, Zurell et al. 2018). The various approaches of revealing interactions from presence-absence co-occurrence data, as well as a suite of cautionary arguments against the endeavor, has been summarized recently by Blanchet et al. (2020). While some of the arguments presented in Blanchet et al. (2020) on presence-absence data can be remedied by analysis of abundance or point pattern data, we agree that the utility of ISA as a direct evidence for interactions is limited. Nevertheless, we argue that there are other reasons for measuring ISA, which we give below.

### *ISA when interactions are given*

Sometimes biotic interactions are not what we want to infer from the data, because we already know how the species interact. Examples are well documented trophic interactions or mutualistic interactions. These known interspecific relationships can generate hypotheses concerning geographic patterns of ISA. For example, a large-scale assessment of biodiversity change may specifically look at patterns of ISA among pairs of pollinators and plants. If there is a significant trend of spatial segregation over time, it may indicate a potential disruption of pollination services, a finding that could be further investigated with additional data or experiments. Here we see a particularly exciting prospect in bringing together network ecology with biogeography.

### *ISA improves forecasts and predictions*

ISA and its patterns are useful even when assuming no mechanistic underpinning, because patterns can improve predictions and forecasts based on inductive logic, rather than causality. Examples of classical predictive biodiversity patterns are the endemic-area relationship predicting extinctions due to habitat loss (Keil et al. 2015), or richness–environment correlations that can be used for spatial interpolations and predictions of diversity patterns (Algar et al. 2009). Similarly, measurements of ISA can potentially improve estimates of other metrics, for example estimates of site-specific species pools (Bruelheide et al. 2020, Carmona and Pärtel 2021) or predictions of species distributions in joint species distribution models (Harris 2016, Norberg et al. 2019). In both cases, the reason for why ISA can improve the predictions is not only as a proxy for species interactions, but also because co-occurring species can act as proxies for suitable environments that may be difficult to measure in the field.

### *ISA as a summary statistic*

ISA captures a unique spatial pattern of communities, on a par with other popular summary statistics such as species richness or beta diversity. Simply reporting how richness or beta diversity vary geographically repeatedly proved to be among the most powerful starting stimuli in the field, generating countless explanatory and testable hypotheses (Brown 1995, Gaston 2000, Lomolino et al. 2010), and such patterns can be as useful for our understanding as the processes that generated them (Currie 2019). An example is the very existence of latitudinal and altitudinal diversity patterns, which have fascinated ecologists for centuries. Similar approach has recently gained traction in summarizing empirical patterns of co-occurrence. For example, Lyons et al. (2016), Tóth et al. (2019), and Calatayud et al. (2020) have documented broad-scale patterns of ISA, and although they do offer interpretations involving biotic interactions, these are part of the post hoc interpretation of the documented patterns, rather than the main goal of the analyses.

### *ISA as a benchmark for theories or mechanistic models*

Rosenzweig and Abramsky (1997) describe the idea of "dipswitch theory," i.e., a theory that makes a bundle of unique predictions, which are then compared with different empirical patterns. The ability of the theory to fit not just one, but multiple patterns, is then a step toward a "stronger" test of the theory (McGill 2003). Here we argue that ISA can allow for stronger tests of theories by providing a unique biodiversity pattern that a theory needs to fit. For instance, it has been demonstrated that neutral theory (Hubbell 2001) or the maximum entropy theory (Harte 2011) both reproduce realistic species–-area relationships, but patterns of ISA, along with other features, might help to distinguish among theories (see, e.g., Ulrich 2004 and Bell 2005). Here, ISA seems particularly promising since it is largely independent on some of the classical patterns such as species–area relationship, or patterns of beta diversity, as we have demonstrated. ISA can also be used together with inverse approaches in individual-based models where known (or hypothesized) individual-level interactions are explicitly modeled and ISA patterns emerge at the community level (Grimm and Railsback 2005). The underlying processes structuring the community can then be inferred by testing how closely the emerging patterns of the model match the observed data. For example, May et al. (2015) used a neutral individual-based model to quantitatively predict patterns observed in a 50-ha tropical forest plot, including beta-diversity (Eq. 2). Surprisingly, the model was able to match five emerging patterns simultaneously, but was unable to match the species–-area relationship and beta-diversity simultaneously, pointing to missing processes. The ability to explain patterns of ISA can thus provide useful information for validating theories and mechanistic models.

## Conclusion

We have argued that interspecific spatial associations (ISA) are an underrepresented topic in biodiversity science and macroecology, and that this is an important knowledge gap worth exploring. Apart from its connection to biotic interactions, ISA can also provide a benchmark for judging different types of ecological theories, and it can serve as a summary statistic capturing unique properties of nature. This is underscored by the fact that many of the biodiversity statistics that have been monitored are insensitive to ISA. We provide an overview of the main approaches to measure ISA, which we sorted into three main schools of thought, based on the data that they use: spatially implicit indices, community variograms, and bivariate pair correlation functions. One of our main conclusions is that considering space, and particularly spatial distance, is vital for the progress of the field, and for any broad-scale assessment of patterns of ISA in geographic space and in time. In all, we hope that our overview of ISA, its measures, and its utility provides a starting point for researchers interested in broadening the scope of biodiversity facets that they study.

## Acknowledgments

The contributions of P. Keil, J. M. Chase, and D. McGlinn were facilitated by support from the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the German Research Foundation (FZT 118). P. Keil was supported by REES (Research Excellence in Environmental Sciences) grant from Faculty of Environmental Sciences, Czech University of Life Sciences in Prague. We thank to Nathan Kraft and two anonymous referees for helpful comments.

## Literature Cited

Algar, A. C., H. M. Kharouba, E. R. Young, and J. T. Kerr. 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. Ecography 32:22–33.

Alroy, J. 2015. A new twist on a very old binary similarity coefficient. Ecology 96:575–586.

Araújo, M. B., and A. Rozenfeld. 2013. The geographic scaling of biotic interactions. Ecography 37:406–415.

Arita, H. T. 2016. Species co-occurrence analysis: pairwise versus matrix-level approaches: Correspondence. Global Ecology and Biogeography 25:1397–1400.

Arita, H. T. 2017. Multisite and multispecies measures of overlap, co-occurrence, and co-diversity. Ecography 40:709–718.

Atmar, W., and B. D. Patterson. 1995. Nestedness temperature calculator. AICS Research Inc, University Park, New Mexico, and The Field Museum, Chicago, Illinois, USA.

Baddeley, A., E. Rubak, and R. Turner. 2016. Spatial point patterns: methodology and applications with R. CRC Press, Taylor & Francis Group, Boca Raton, Florida, USA.

Baselga, A. 2017. Partitioning abundance-based multiple-site dissimilarity into components: balanced variation in abundance and abundance gradients. Methods in Ecology and Evolution 8:799–808.

Baselga, A., and C. D. L. Orme. 2012. betapart: an R package for the study of beta diversity: *Betapart package.* Methods in Ecology and Evolution 3:808–812.

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86:1757–1770.

Blanchet, F. G., K. Cazelles, and D. Gravel. 2020. Co-occurrence is not evidence of ecological interactions. Ecology Letters 23:1050–1063.

Blois, J. L., et al. 2014. A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. Ecography 37:1095–1108.

Blowes, S. A., et al. 2019. The geography of biodiversity change in marine and terrestrial assemblages. Science 366:339–345.

Braun-Blanquet, J. 1964. Pflanzensoziologie: Grundzüge der Vegetationskunde. Springer, Wien, Austria.

Brown, J. H. 1995. Macroecology. University of Chicago Press, Chicago, Illinois, USA.

Brown, J. H., E. J. Bedrick, S. K. M. Ernest, J.-L.-E. Cartron, and J. F. Kelly. 2004. Constraints on negative relationships: mathematical causes and ecological consequences. Pages 298–323 *in* M. L. Taper, and S. R. Lele, editors. The nature of scientific evidence: statistical, philosophical, and empirical considerations. University of Chicago Press, Chicago, Illinois, USA.

Bruelheide, H., B. Jiménez-Alfaro, U. Jandt, and F. M. Sabatini. 2020. Deriving site-specific species pools from large databases. Ecography 43:1215–1228.

Calatayud, J., et al. 2020. Positive associations among rare species and their persistence in ecological assemblages. Nature Ecology & Evolution 4:40–45.

Carmona, C. P., and M. Pärtel. 2021. Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. Global Ecology and Biogeography 30:316–326.

Cody, M. L., and J. M. Diamond, editors. 1979. Ecology and evolution of communities. Belknap Press, Cambridge, Massachusetts, USA.

Condit, R., et al. 2000. Spatial patterns in the distribution of tropical tree species. Science 288:1414.

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

Cramér, H. 1924. Remarks on correlation. Skandinavisk Aktuarietidskrift 7:220–240.

Cressie, N. A. C. 2010. Statistics for spatial data. Wiley, New York, New York, USA.

Crist, T. O., and J. A. Veech. 2006. Additive partitioning of rarefaction curves and species-area relationships: unifying alpha-, beta- and gamma-diversity with sample size and habitat area. Ecology Letters 9:923–932.

Currie, D. J. 2019. Where Newton might have taken ecology. Global Ecology and Biogeography 28:18–27.

Dice, L. R. 1945. Measures of the amount of ecologic association between species. Ecology 26:297–302.

Dormann, C. F., J. Frund, N. Bluthgen, and B. Gruber. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. Open Ecology Journal 2:7–24.

Dormann, C. F., B. Gruber, and J. Frund. 2008. Introducing the bipartite package: Analysing ecological networks. R News 8:8–11.

Dray, S., et al. 2012. Community ecology in the age of multivariate multiscale spatial analysis. Ecological Monographs 82:257–275.

Dray, S. 2019. spacemakeR: Spatial modelling. R-Forge. https://rdrr.io/rforge/spacemakeR

Forbes, S. A. 1907. On the local distribution of certain Illinois fishes: an essay in statistical ecology. Bulletin of the Illinois State Laboratory of Natural History 7:273–303.

Fox, J. W. 2012. An important but little known fact about compensatory dynamics. https://dynamicecology.wordpress.com

Gaston, K. J. 2000. Global patterns in biodiversity. Nature 405:220–227.

Gaston, K. J., and J. I. Spicer. 2009. Biodiversity: an introduction. Blackwell, Malden, Massachusetts, USA.

Getzin, S., T. Wiegand, and S. P. Hubbell. 2014. Stochastically driven adult-recruit associations of tree species on Barro Colorado Island. Proceedings of the Royal Society B: Biological Sciences 281:20140922.

Goodman, L. A., and W. H. Kruskal. 1979. Measures of association for cross classifications. Springer, New York, New York, USA.

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

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379–391.

Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. Smithsonian Press, Washington, D.C., USA.

Gotelli, N. J., G. R. Graves, and C. Rahbek. 2010. Macroecological signals of species interactions in the Danish avifauna. Proceedings of the National Academy of Sciences USA 107:5030–5035.

Gotelli, N. J., E. M. Hart, and A. M. Ellison. 2015. EcoSimR: Null model analysis for ecological data. CRAN. https://rdrr.io/cran/EcoSimR

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

Grimm, V., and S. F. Railsback. 2005. Individual-based modelling in ecology. Princeton University Press, Princeton, New Jersey, USA.

Harris, D. J. 2016. Inferring species interactions from co-occurrence data with Markov networks. Ecology 97:3308–3314.

Harte, J. 2011. Maximum entropy and ecology: A theory of abundance, distribution, and energetics. Oxford University Press, Oxford, UK.

Hartley, S., and W. E. Kunin. 2003. Scale dependency of rarity, extinction risk, and conservation priority. Conservation Biology 17:1559–1570.

He, F., and P. Legendre. 2002. Species diversity patterns derived from species-area models. Ecology 83:1185–1198.

Hoekstra, J. M., T. M. Boucher, T. H. Ricketts, and C. Roberts. 2004. Confronting a biome crisis: global disparities of habitat loss and protection: confronting a biome crisis. Ecology Letters 8:23–29.

Hubálek, Z. 1982. Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. Biological Reviews 57:669–689.

Hubbell, S. P. 2001. The unified theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey, USA.

Hubbell, S. P., and R. B. Foster. 1986. Biology, chance, and history and the structure of tropical rain forest tree communities. Pages 314–329 *in* J. M. Diamond and T. J. Case, editors. Community ecology. Harper and Row Publishers, New York, New York, USA.

Hui, C. 2009. On the scaling patterns of species spatial distribution and association. Journal of Theoretical Biology 261:481–487.

IPBES. 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services. IPBES secretariat, Bonn, Germany.

Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37:547–579.

Jeliazkov, A. 2019. A global database for metacommunity ecology, integrating species, traits, environment and space. Scientifc Data 7:1–6.

Jetz, W., J. M. McPherson, and R. P. Guralnick. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. Trends in Ecology & Evolution 27:151–159.

Kallio, A., K. Puolamäki, M. Fortelius, and H. Mannila. 2011. Correlations and co-occurrences of taxa: The role of temporal, geographic, and taxonomic restrictions. Palaeontologia Electronica 14:1–14.

Keil, P. 2014. Limits of uncertainty about estimates of probability of ecological events. PeerJ PrePrints 2:e446v1.

Keil, P. 2019. Z-scores unite pairwise indices of ecological similarity and association for binary data. Ecosphere 10:e02933.

Keil, P., D. Storch, and W. Jetz. 2015. On the decline of biodiversity due to area loss. Nature Communications 6:8837.

Kikvidze, Z., F. I. Pugnaire, R. W. Brooker, P. Choler, C. J. Lortie, R. Michalet, and R. M. Callaway. 2005. Linking patterns and processes in Alpine plant communities: a global study. Ecology 86:1395–1400.

Koleff, P., K. J. Gaston, and J. J. Lennon. 2003. Measuring beta diversity for presence–absence data. Journal of Animal Ecology 72:367–382.

Kunin, W. E. 1998. Extrapolating species abundance across spatial scales. Science 281:1513–1515.

Ladle, R. J., and R. J. Whittaker, editors. 2011. Conservation biogeography. Wiley-Blackwell, Chichester, UK.

Lebart, L., and J. P. Fénelon. 1971. Statistique et informatique appliqées. Dunod, Paris, France.

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

Legendre, P., and L. Legendre. 2012. Numerical ecology. Elsevier, Amsterdam, the Netherlands.

Lewinsohn, T. M., P. Inácio Prado, P. Jordano, J. Bascompte, and J. M. Olesen. 2006. Structure in plant-animal interaction assemblages. Oikos 113:174–184.

Lieberman, M., and D. Lieberman. 2007. Nearest-neighbor tree species combinations in tropical forest: the role of chance, and some consequences of high diversity. Oikos 116:377–386.

Lomolino, M. V., B. R. Riddle, R. J. Whittaker, and J. H. Brown. 2010. Biogeography. Fourth edition. Sinauer Associates, Sunderland, Massachusetts, USA.

Lyons, S. K., et al. 2016. Holocene shifts in the assembly of plant and animal communities implicate human impacts. Nature 529:80–83.

Marion, Z. H., J. A. Fordyce, and B. M. Fitzpatrick. 2017. Pairwise beta diversity resolves an underappreciated source of confusion in calculating species turnover. Ecology 98:933–939.

May, F., A. Huth, and T. Wiegand. 2015. Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. Proceedings of the Royal Society B 282:20141657.

May, R. M. 1972. Will a large complex system be stable? Nature 238:413–414.

McGeoch, M. A., and K. J. Gaston. 2002. Occupancy frequency distributions: patterns, artefacts and mechanisms. Biological Reviews 77:311–331.

McGill, B. 2003. Strong and weak tests of macroecological theory. Oikos 102:679–685.

McGill, B. J. 2010. Towards a unification of unified theories of biodiversity. Ecology Letters 13:627–642.

McGill, B. J., and J. C. Nekola. 2010. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. Oikos 119:591–603.

McGlinn, D. J., and A. H. Hurlbert. 2012. Scale dependence in species turnover reflects variance in species occupancy. Ecology 93:294–302.

McGlinn, D. J., X. Xiao, F. May, N. J. Gotelli, T. Engel, S. A. Blowes, T. M. Knight, O. Purschke, J. M. Chase, and B. J. McGill. 2019. Measurement of Biodiversity (MoB): A method to separate the scale-dependent effects of species abundance distribution, density, and aggregation on diversity change. Methods in Ecology and Evolution 10:258–269.

McNickle, G. G., E. G. Lamb, M. Lavender, J. F. Cahill, B. S. Schamp, S. D. Siciliano, R. Condit, S. P. Hubbell, and J. L. Baltzer. 2018. Checkerboard score-area relationships reveal spatial scales of plant community structure. Oikos 127:415–426.

Millenium Ecosystem Assessment. 2005. Ecosystems and human well-being: synthesis. Island Press, Washington, D.C., USA.

Milo, R. 2002. Network motifs: simple building blocks of complex networks. Science 298:824–827.

Morueta-Holme, N., B. Blonder, B. Sandel, B. J. McGill, R. K. Peet, J. E. Ott, C. Violle, B. J. Enquist, P. M. Jørgensen, and J.-C. Svenning. 2016. A network approach for inferring species associations from co-occurrence data. Ecography 39:1139–1150.

Newbold, T., et al. 2015. Global effects of land use on local terrestrial biodiversity. Nature 520:45–50.

Norberg, A., et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecological Monographs 89.

Odum, E. P. 1950. Bird populations of the highlands (North Carolina) plateau in relation to plant succession and avian invasion. Ecology 31:587–605.

Äkland, R. H. 1994. Patterns of bryophyte associations at different scales in a Norwegian boreal spruce forest. Journal of Vegetation Science 5:127–138.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M. Suggests. 2019. vegan: Community Ecology Package. https://cran.r-project.org/web/packages/vegan/index.html

Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecology Letters 20:561–576.

Palmer, M. W., and E. van der Maarel. 1995. Variance in species richness, species association, and niche limitation. Oikos 73:203.

Pearson, K., and D. Heron. 1913. On theories of association. Biometrika 9:159–315.

Peters, R. H. 1991. A critique for ecology. Cambridge University Press, Cambridge, UK.

Plotkin, J. B., M. D. Potts, N. Leslie, N. Manokaran, J. Lafrankie, and P. S. Ashton. 2000. Species-area curves, spatial aggregation, and habitat specialization in tropical forests. Journal of Theoretical Biology 207:81–99.

Presley, S. J., C. L. Higgins, and M. R. Willig. 2010. A comprehensive framework for the evaluation of metacommunity structure. Oikos 119:908–917.

Rajagopalan, S., and R. Robb. 2005. Assessment of similarity indices to quantify segmentation accuracy of scaffold images for tissue engineering. Page 1636 *in* J. M. Fitzpatrick and J. M. Reinhardt, editors. Proceedings of Society of Photo-Optical Instrumentation Engineers. SPIE, San Diego, California, USA.

Rajala, T., S. C. Olhede, and D. J. Murrell. 2019. When do we have the power to detect biological interactions in spatial point patterns? Journal of Ecology 107:711–721.

Rao, C. R. 1995. A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. Qüestiió 19:23–63.

Roberge, J.-M., and P. Angelstam. 2004. Usefulness of the umbrella species concept as a conservation tool. Conservation Biology 18:76–85.

Rosenzweig, M. L., and Z. Abramsky. 1997. Two gerbils of the Negev: A long-term investigation of optimal habitat selection and its consequences. Evolutionary Ecology 11:733–756.

Roxburgh, S. H., and P. Chesson. 1998. A new method for detecting species associations with spatially autocorrelated data. Ecology 79:2180–2192.

Růžička, M. 1958. Anwendung mathematisch-statistischer Methoden in der Geobotanik (synthetische Bearbeitung von Aufnahmen). Biologia, Bratislava 13:647–661.

Sanderson, J. G., and S. L. Pimm 2015. Patterns in nature: the analysis of species co-occurrences. The University of Chicago Press, Chicago, Illinois, USA.

Sauer, J. R., K. L. Pardieck, D. J. Ziolkowski, A. C. Smith, M.-A.-R. Hudson, V. Rodriguez, H. Berlanga, D. K. Niven, and W. A. Link. 2017. The first 50 years of the North American Breeding Bird Survey. Condor 119:576–593.

Schluter, D. 1984. A variance test for detecting species associations, with some example applications. Ecology 65:998–1005.

Seabloom, E. W., O. N. Bjørnstad, B. M. Bolker, and O. J. Reichman. 2005. Spatial signature of environmental heterogeneity, dispersal, and compatition in successional grasslands. Ecological Monographs 75:199–214.

Segurado, P., W. E. Kunin, A. F. Filipe, and M. B. Araújo. 2012. Patterns of coexistence of two species of freshwater turtles are affected by spatial scale. Basic and Applied Ecology 13:371–379.

Shimatani, K. 2001. Multivariate point processes and spatial variation of species diversity. Forest Ecology and Management 142:215–229.

Shimatani, K., and Y. Kubota. 2004. Quantitative assessment of multispecies spatial pattern with high species diversity: Multispecies spatial pattern. Ecological Research 19:149–163.

Simberloff, D., and E. F. Connor 1979. Q-mode and R-mode analyses of biogeographic distributions: null hypotheses based on random colonization. Pages 123–138 *in* P. P. Ganapati and M. L. Rosenzweig, editors. Contemporary quantitative ecology and related ecometrics. International Cooperative Publishing House, Fairland, Maryland, USA.

Simpson, E. H. 1949. Measurement of diversity. Nature 163:688.

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

Šizling, A. L., and D. Storch. 2004. Power-law species-area relationships and self-similar species distributions within finite areas. Ecology Letters 7:60–68.

Šmilauer, P., and J. Lepš. 2014. Multivariate analysis of ecological data using Canoco 5. Second edition. Cambridge University Press, Cambridge, UK.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. University of Kansas Scientific Bulletin 38:1409–1438.

Sørensen, T. 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. Biologiske Skrifter/Kongelige Danske Videnskabernes Selskab 5:1–34.

Stone, L. 2016. The Google matrix controls the stability of structured ecological and biological networks. Nature Communications 7:12857.

Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. Oecologia 85:74–79.

Stone, L., and A. Roberts. 1992. Competitive exclusion, or species aggregation?: An aid in deciding. Oecologia 91:419–424.

Storch, D. 2016. The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. Journal of Vegetation Science 27:880–891.

Storch, D., A. L. Šizling, J. Reif, J. Polechová, E. Šizlingová, and K. J. Gaston. 2008. The quest for a null model for macroecological patterns: geometry of species distributions at multiple spatial scales. Ecology Letters 11:771–784.

Stoyan, D., and H. Stoyan. 1994. Fractals, random shapes and point fields. Methods of geometrical statistics. John Wiley & Sons, Chichester, UK.

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature 189:732–735.

ter Braak, C. 1987. CANOCO—a FORTRAN program for canonical community ordination by partial detrended canonical correspondence analysis, principal components analysis and redundancy analysis (version 2.1). Agricultural Mathematics Group, Wageningen University, Wageningen, the Netherlands.

Thurman, L. L., A. K. Barner, T. S. Garcia, and T. Chestnut. 2019. Testing the link between species interactions and species co-occurrence in a trophic network. Ecography 42:1658–1670.

Tilman, D., and P. M. Kareiva, editors. 1997. Spatial ecology: the role of space in population dynamics and interspecific interactions. Princeton University Press, Princeton, New Jersey, USA.

Tóth, A. B., et al. 2019. Reorganization of surviving mammal communities after the end-Pleistocene megafaunal extinction. Science 365:1305–1308.

Ulrich, W. 2004. Species co-occurrences and neutral models: reassessing. J. M. Diamond's assembly rules. Oikos 107:603–609.

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

Ulrich, W., and N. J. Gotelli. 2010. Null model analysis of species associations using abundance data. Ecology 91:3384–3397.

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

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

Veech, J. A. 2013. A probabilistic model for analysing species co-occurrence. Global Ecology and Biogeography 22:252–260.

Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan. 2009. Inferring species interactions in tropical forests. Proceedings of the National Academy of Sciences USA 106:13854–13859.

Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology 84:1045–1057.

Wagner, H. H., R. Holderegger, S. Werth, F. Gugerli, S. E. Hoebee, and C. Scheidegger. 2005. Variogram analysis of the spatial genetic structure of continuous populations using multilocus microsatellite data. Genetics 169:1739–1752.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2015. So many variables: Joint modeling in community ecology. Trends in Ecology & Evolution 30:766–779.

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

Wiegand, T., C. V. S. Gunatilleke, I. A. U. N. Gunatilleke, and A. Huth. 2007*a*. How individual species structure diversity in tropical forests. Proceedings of the National Academy of Sciences USA 104:19029–19033.

Wiegand, T., S. Gunatilleke, and N. Gunatilleke. 2007*b*. Species associations in a heterogeneous Sri Lankan dipterocarp forest. American Naturalist, 170:E77–E95.

Wiegand, T., A. Huth, S. Getzin, X. Wang, Z. Hao, C. V. S. Gunatilleke, and I. A. U. N. Gunatilleke. 2012. Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. Proceedings of the Royal Society B 279:3312–3320.

Wiegand, T., and K. A. Moloney. 2004. Rings, circles, and null-models for point pattern analysis in ecology. Oikos 104:209–229.

Wiegand, T., and K. A. Moloney. 2014. Handbook of spatial point-pattern analysis in ecology. CRC Press, Boca Raton, Florida, USA.

WWF International. 2012. Living planet report 2012: Biodiversity, biocapacity and better choices. WWF International, Gland, Switzerland.

Zurell, D., L. J. Pollock, and W. Thuiller. 2018. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? Ecography 41:1812–1819.

## SUPPORTING INFORMATION

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecm.1452/full

## OPEN RESEARCH

Code and data used for this study are archived in Zenodo (https://doi.org/10.5281/zenodo.4431673).