

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

Petr Keil^{1, 2, *}, Thorsten Wiegand^{1, 3}, Anikó B. Tóth⁴, Daniel J. McGlinn⁵, 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 Helmholtz Centre for Environmental Research - UFZ, Department of Ecological Modelling, 04318 Leipzig, Germany

⁴ Centre for Ecosystem Sciences, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, 2052 Australia.

⁵ Department of Biology, College of Charleston, Charleston SC 29401, USA

* Correspondence: pkeil@seznam.cz

Article type: Review

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 both spatially implicit and explicit indices of association for both binary and abundance data. We test and compare these approaches on empirical and simulated data, and we provide specific 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 pair-wise 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

33 link to biotic interactions. Specifically, ISA can provide strong tests of biodiversity theories that
34 require multiple patterns to benchmark against, and it can be explored for potentially predictive
35 macroecological patterns.

36 **Keywords:** C-score, co-occurrence, point process, interspecific, intraspecific, conspecific,
37 heterospecific, spatial scale, grain, semivariogram, geostatistics

38

39 “... *there is no absolutely general measure of the degree of dependence*”

40 (Cramér 1924, quoted by Goodman and Kruskal 1979, quoted by Hubálek 1982)

41 INTRODUCTION

42 Organisms are almost never distributed randomly in space. Instead, they clump together or spread
43 apart, and such non-random spatial aggregation and segregation occurs both among individuals within
44 a single species and among species; we call the former the *conspecific spatial aggregation* (CSA), and
45 the latter the *interspecific spatial association* (ISA) (Fig. 1, Appendix S1). Static patterns of CSA and
46 ISA can be generated by various mechanisms. For example, ISA may result from shared habitat
47 requirements among species, dispersal limitation, and/or interspecific interactions (Table 1), and there
48 is a long tradition of inferring these mechanisms from spatial patterns of ISA (Forbes 1907, Cody and
49 Diamond 1979, Gotelli et al. 2010, Calatayud et al. 2020). However, it is not always possible, nor
50 useful, to infer causal mechanisms from emergent static patterns (Peters 1991, McGill and Nekola
51 2010), and this is particularly problematic in the case of ISA for presence-absence data (Blanchet et
52 al. 2020).

53 Fortunately, documenting static patterns of spatial aggregation and association has value regardless of
54 the generative mechanisms, such as in macroecological studies (Currie 2019) and for nature
55 conservation and forecasts (Ladle and Whittaker 2011). For example, patterns of CSA underpin
56 widely applied concepts such as species endemism, range size, and rarity (Table 1), and there are well
57 described connections between CSA and patterns of beta diversity and species-area relationships
58 (Storch et al. 2008, Wiegand and Moloney 2014). Similarly, patterns of ISA have direct applications
59 in classifications of communities to coenoses or biomes (Hoekstra et al. 2004), underpin the concepts
60 of indicator and umbrella species (Roberge and Angelstam 2004), and can improve estimates of site-
61 specific species pools (Karger et al. 2016) and predictions of species distributions (Harris 2016,
62 Norberg et al. 2019). Further, even though it has not been perceived as such, ISA describes a unique
63 facet of biodiversity, particularly when biodiversity is defined as “variation of life at all levels of
64 biological organization” (Gaston and Spicer 2009). Even though biodiversity is typically measured
65 “per site”, it can indeed also be quantified in other ways, such as “per species” (e.g., a number of

66 species associated with a given species). However, the theoretical and empirical connections between
67 ISA and spatial patterns of biodiversity remain poorly understood.

68 Only a handful of studies have attempted to empirically document large-scale patterns of ISA (Gotelli
69 et al. 2010, Lyons et al. 2016, Tóth et al. 2019, Calatayud et al. 2020), and ISA has been missing in
70 recent high-profile studies of the ongoing biodiversity change (Millenium Ecosystem Assessment
71 2005, WWF International 2012, Newbold et al. 2015, Blowes et al. 2019, IPBES 2019). To
72 investigate this potential knowledge gap more systematically, we conducted two literature searches.
73 First, we manually went through 3,856 abstracts of papers published between 1995 and 2019 in three
74 journals with a long history of publishing biodiversity studies—*American Naturalist*, *Ecology*, and
75 *Ecography*. We found that ISA is a marginal topic when compared to patterns of CSA, beta diversity,
76 and community composition, as well as processes such as biotic interactions (Appendix S2). Second,
77 we scanned Clarivate Web of Science for all papers mentioning biodiversity in their title or abstract;
78 among these we looked at frequency of terms associated with ISA and terms associated with other
79 ecological patterns and processes (Appendix S2). We found that this literature has most often focused
80 on quantities at sites (e.g. species richness, functional and phylogenetic diversity), or variation among
81 sites (e.g. beta diversity, Fig. S1). In contrast, biodiversity-related papers paid considerably less
82 attention to ISA, measured by the total number of published papers per year (Appendix S2). Since
83 measurement of ISA requires exactly the same data as analysis of beta diversity (see below), the lack
84 of attention cannot be explained by a lack of data. We propose three reasons for this gap: First, there
85 is a variety of approaches for measuring ISA, with little clear guidance on the advantages and
86 disadvantages of each approach. Second, it is not always clear why ISA may be important or useful
87 when studying biodiversity. Third, little is known about the theoretical connections, or lack thereof,
88 between ISA and other biodiversity patterns.

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

97 SCHOOLS OF THOUGHT

98 Measurement and analysis of ISA has been approached by several schools of thought, with each of
99 these typically working with a particular data type (Fig. 2) and research agenda. While each of these

100 schools of thought has generated conceptually diverse measures of ISA, there has been little cross-
101 fertilization. The first school uses non-spatial information on presence/absence (or abundance) of
102 species in different sites and has focused on tests of hypotheses related to assembly of ecological
103 communities (Cody and Diamond 1979) by comparing the observed patterns of ISA with simulations
104 of null models (Gotelli and Graves 1996, Ulrich and Gotelli 2010, Sanderson and Pimm 2015). The
105 second school uses similar non-spatial metrics, but has been more descriptive, identifying groups of
106 species that are frequently observed together, a typical goal in phytocoenology (Braun-Blanquet
107 1964). It relies on methods such as ordinations and cluster analysis (ter Braak 1987, Šmilauer and
108 Lepš 2014). The third school comes from the field of geostatistics (Cressie 2010) and uses the spatial
109 position of the sites to show how covariance between two species changes with spatial distance
110 (Wagner 2003). The fourth school is the analysis of spatial point patterns (Wiegand and Moloney
111 2014; Baddeley et al. 2015), which requires the most detailed data on the spatial positions of
112 individuals, but offers the richest inference. Below we review the specific measures coming from
113 these diverse schools of thought.

114 SPATIALLY IMPLICIT MEASURES OF ISA

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

123 To date, nearly 100 indices have been proposed to capture ISA in binary co-occurrence data (Hubálek
124 1982, Rajagopalan and Robb 2005, Legendre and Legendre 2012, Ulrich and Gotelli 2013, Arita
125 2017) and for abundance data (Legendre and Legendre 2012, Legendre and De Cáceres 2013),
126 although many fewer are widely used in ecology. Some of the pairwise indices were developed
127 specifically to capture ISA while others were adopted from the literature on beta diversity (Hubálek
128 1982, Legendre and Legendre 2012, Arita 2017). Most of the indices we describe below are available
129 through R packages, such as *vegan* (Oksanen et al. 2019), *EcoSimR* (Gotelli et al. 2015), *bipartite*
130 (Dormann et al. 2008), and *betapart* (Baselga and Orme 2012), as listed in Tables 2 and 3.

131 *Pairwise associations*

132 The conceptually simplest approach is to analyze associations between binary occurrences of two

133 species. Table 2 lists some of the typical pairwise indices of association for binary data that have been
134 widely used, or that represent a unique approach to capturing ISA. They are based on four different
135 quantities: the number of sites occupied uniquely by species 1 (c) and species 2 (b), the number of
136 sites where both species co-occur (a), the number of sites where none occurs (d), and $n = a + b +$
137 $c + d$. These indices vary in their interpretation; the widely used C-score ($C_{seg} = bc$), for example,
138 gives the number of pairs of sites with a “checkerboard” pattern (i.e., one site hosts only species 1, the
139 other only species 2), and it is a measure of segregation between two species. In contrast, togetherness
140 ($C_{tog} = ad$) is a measure of attraction between species, but the C_{seg} and C_{tog} are not complements (i.e.
141 low C_{seg} does not imply high C_{tog}). Another widely used example is the Jaccard index (of association,
142 not beta diversity), a measure of spatial overlap, which describes the proportion of sites occupied by
143 both species among all occupied sites ($C_{jacc} = a/(a + b + c)$). Further, the indices vary in the range
144 of possible values, such as $[-1, 1]$, $[0, 1]$, or $[0, \infty)$. Thus, it is important to be familiar with both the
145 interpretation and interval of these indices, particularly in their raw form (as opposed to their Z-score,
146 see below) (Hubálek 1982, Legendre and Legendre 2012).

147 Table 3 lists some commonly used indices for pair-wise ISA in abundance data. The first family
148 includes both parametric (e.g. CA_{cor}) and non-parametric (e.g. CA_{rho}) indices of covariance and
149 correlation. Another family includes indices that are extensions of the incidence-based similarity
150 measures; an example is Růžička similarity CA_{ruz} , which is one of the abundance-based versions of
151 Jaccard’s index for binary data (Legendre and De Cáceres 2013).

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

153 When the aim is to quantify the overall magnitude of ISA in the entire matrix Y in a single number,
154 one option is to simply use the mean or median of the species-by-species association matrix Z .
155 However, focusing only on pairwise comparisons ignores interactions of higher order [i.e. interactions
156 between 3 or more species (Harris 2016)]. Some methods can capture N-wise species associations
157 when applied to the rows of Y (Baselga 2017, Arita 2017). The problem with these indices is their
158 sensitivity to double zeroes and it is still unclear how common the higher-order associations are [i.e.,
159 motifs (Milo 2002)] in the real world (Marion et al. 2017). Finally, Legendre and De Cáceres (2013)
160 proposed an index of contribution of a single locality to the overall beta diversity in Y . The same
161 index can be calculated for rows instead of columns of Y , thus becoming a measure of species
162 contribution to overall association; however, we are unaware of a study that employs such an
163 approach to species-wise ISA.

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

165 There are several ways to calculate the probability that the observed metric has been produced by
166 chance given the null expectation that there is no ISA. The simplest test of significance of association

167 in a two-species binary matrix is Fisher's exact test (Arita 2016), which is identical to the
168 randomization procedure of Veech (2013). However, when more than two species are involved, or
169 when the data are abundances rather than incidences, a randomization procedure can produce the null
170 expectation of a given ISA metric (see e.g., Gotelli 2000, Gotelli et al. 2015, Ulrich and Gotelli 2010).
171 Further, the deviation from the null expectation can also be used as its own ISA metric, such as the
172 C_{forbes} and C_{FETMP} metrics (Table 2), and the so-called Z-score (Gotelli and McCabe 2002, Ulrich et al.
173 2009, Ulrich and Gotelli 2013), defined as $(E_{raw} - E_{exp})/SD_{exp}$, and where E_{raw} is the ISA metric
174 (e.g. from Table 2 or 3) calculated on observed data, E_{exp} is the null expectation of the index, and
175 SD_{exp} is the standard deviation of the null expectation.

176 SPATIALLY EXPLICIT MEASURES OF ISA

177 [Spatial structure between sites, individuals, and/or geographic ranges can be critical. Consider the](#)
178 [simple example in Figure 3 in which the same community matrix represents either between-species](#)
179 [aggregation or segregation; the aforementioned spatially implicit metrics of ISA conflate these](#)
180 [patterns, and a spatially explicit approach needs to be adopted.](#) When spatial coordinates of sites or
181 individuals are available, one way to make the ISA spatially explicit is to only consider pairs of sites
182 that are within a given distance interval when using any of the metrics from the previous section
183 (Tables 2 and 3), and then to examine a range of distances. This is also the idea behind community
184 variograms (Wagner 2003) described below. A point pattern-based version of this, when the location
185 and identity of each individual is known, is to use bivariate pair correlation functions which describe
186 positive and negative ISA between individuals over continuous distance (Wiegand and Moloney
187 2014). In the following section, we describe principles of these two approaches, and refer to software
188 implementations.

189 *Community variograms*

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

199 To test for non-random patterns of spatial species covariance an appropriate null model is required.

200 The distancedistance-dependent C_{ratio} is typically used with a null expectation of $C_{ratio} = 1$ under
201 species independence, which can be biased in a number of different ways (Palmer and van der Maarel
202 1995). In particular, it can be biased by patterns of within-species clumping. Even if species are
203 arranged independently on a landscape, strong patterns of within-species clumping will create the
204 appearance of spatial segregation at least up to the scale of the within-species clump size. Therefore,
205 the most common type of null model is one in which the within-species spatial clumping is held
206 constant but otherwise species are shuffled randomly (Palmer and van der Maarel 1995, Roxburgh
207 and Chesson 1998).

208 Even though the idea of community variograms has been featured in prominent methodological
209 reviews (Dray et al. 2012), and elements of it occasionally appear in empirical analyses (Wagner et al.
210 2005, Seabloom et al. 2005, Kikvidze et al. 2005, Ovaskainen et al. 2017), we are unaware of its
211 direct use estimating distance-dependent ISA. At the same time the method is the closest thing to the
212 truly spatially-explicit description of ISA for community matrices with additional spatial coordinates,
213 and we thus see potentially important applications. Two R packages that allow calculation of
214 community variograms are spacemakeR (Dray 2019) with the function variomultiv, and package vario
215 (<https://github.com/dmcglinn/vario>) with the function vario.

216 *Point pattern analysis*

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

223 *The bivariate pair correlation function* $g_{12}(r)$ measures the association between species at different
224 spatial scales (Stoyan and Stoyan 1994, Wiegand et al. 2007b); it relates to the density of species 2 at
225 distance r of the nearest species 1. Positive association (attraction) occurs for $g_{12}(r) > 1$ (larger than
226 expected neighborhood density) and negative association (segregation) for $g_{12}(r) < 1$ (smaller than
227 expected neighborhood density). Furthermore, the cumulative version of $g_{12}(r)$ is the K-function
228 $K_{12}(r)$. Popular software to calculate $g_{12}(r)$ and $K_{12}(r)$ are Programita (Wiegand and Moloney 2004,
229 2014), and R package spatstat (pcfcross function) (Baddeley et al. 2015).

230 *The P-M classification scheme* provides a way to further classify a spatial pairwise ISA as either
231 independent, fully segregated, partially overlapping, or “mixing” at a given spatial lag r using a
232 combination of the Z-score transformed statistics of nearest neighbor occurrence (P) and
233 neighborhood density (M) (Wiegand et al. 2007b). This is best used to summarize the overall ISA

234 structure at a given spatial scale (Getzin et al. 2014). It can be calculated using the Programita
235 software (Wiegand and Moloney 2004, 2014), and we also offer an R implementation in package
236 *spasm* that relies heavily on the *spatstat* package (Baddeley et al. 2015).

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

247 COMPARISON OF THE MEASURES

248 [In order to get a comprehensive basis for recommendations, we compare the performance of the](#)
249 [approaches on a common set of empirical and simulated communities.](#) Our aim is to assess
250 redundancy among the approaches, as well as their sensitivity to the common variables such as
251 number of sites, number of species, magnitude of conspecific aggregation, or spatial grain. We are
252 also interested in how well the different measures capture negative associations (i.e. segregation)
253 compared to ositive associations (attraction) of ISA. Code and data for these analyses are at in
254 package *spasm* (see Code and Data section).

255 *Spatially implicit measures and empirical data*

256 [Our goal in this exercise was to calculate the common metrics of ISA for a set of spatially implicit](#)
257 [species-by-site matrices, and to see how the metrics correlate with each other.](#) We evaluated the
258 metrics in Table 2 using 476 empirical binary community matrices collated by Atmar and Patterson
259 (1995) and Ulrich and Gotelli (2010), and the metrics in Table 3 using 186 empirical abundance
260 matrices collated by (Ulrich and Gotelli 2010). We excluded 3 binary matrices and 52 abundance
261 matrices with too little information to allow meaningful calculation of all of the indices, or with
262 negative or positive infinity values of some of the metrics. For the pairwise metrics, we averaged
263 them over the species-by-species association matrix Z to obtain a single number comparable with the
264 matrix-wise measures. We subjected the metrics to principal components analysis (PCA), for which
265 we transformed some of them to have an approximately normal distribution. We plotted the PCA as
266 an ordination biplot, and we also plotted a graph representation of the correlation matrix of the

267 measures (Fig. 4). The raw correlations are provided in Figures S3 and S4.

268 In both the incidence- and abundance-based spatially implicit indices, we found clear differentiation
 269 along the PCA axes (Fig. 4a, b) and in the graph (Fig. 4c, d), reflecting the different aspects of ISA. In
 270 binary metrics, there was a clear cluster of similarity-based indices (C_{jacc} , C_{sor}), Alroy's index (C_{alroy}),
 271 and connectance (C_{conn}) (Fig. 4a, c, Fig. S1). Two measures that explicitly measure departure from
 272 Poisson randomness are C_{FETmP} and C_{forbes} , which were grouped together. The C-score (C_{segSc}) and its
 273 counterpart, togetherness (C_{togSc}), the matching coefficient (C_{match}), the checkerboard score ($C_{checker}$),
 274 and number of unique combinations (C_{combo}) were largely unrelated to the rest of the metrics. In the
 275 abundance-based metrics, we found similar clustering of the similarity-based ISA metrics (CA_{ruz} ,
 276 CA_{bray} , CA_{chi} , CA_{heli}), while the correlation-based metrics (CA_{cor} , CA_{cor_heli} , CA_{rho}) formed another
 277 group.

278 Notably, among the incidence-based metrics, only C_{combo} was strongly correlated with the total
 279 number of sites in a matrix (n), while only $C_{checker}$ correlated strongly with S , the total number of
 280 species (Fig. 4, S3). Covariance-based measures (CA_{cov} , CA_{cov_heli}) grouped together with the total
 281 number of sites n , while most metrics correlated only weakly with S (Fig. 4, S3, S4).

282 *Spatially implicit measures and spatially explicit simulations*

283 In the second exercise, we devised simulations with known magnitude of ISA between two species
 284 (Fig. 5, S5, Appendix S3) [to illustrate how the different approaches recover the known ISA](#). In
 285 contrast to spatially implicit simulations performed in other studies (Gotelli 2000, Ulrich and Gotelli
 286 2010), we simulated the position of every individual within a bounded domain. In each simulation, we
 287 generated point patterns of two species, with a given magnitude of ISA between them, and with
 288 varying conspecific aggregation and number of individuals per species.

289 We modelled ISA as dependent on spatial distance (Fig. 5b), and it was controlled by a single
 290 parameter which we call ISA, with negative values for segregation ($ISA < 0$), zero for independence
 291 ($ISA = 0$), and positive for attraction ($ISA > 0$) (Fig. 5). To calculate measures based on incidence or
 292 abundance, we converted the point patterns to grids of varying resolutions (grains). Across
 293 simulations, we measured performance of the metrics from Tables 2 and 3 as their Spearman
 294 correlation with the ISA parameter. We also examined how the performance was affected by spatial
 295 grain (Fig. S7) and magnitude of conspecific spatial aggregation (CSA) (Fig. S8). Detailed description
 296 of the simulations, and their analysis, is in Appendix S3, and complete code is in package `spasm` (see
 297 Code and Data section below).

298 We found that the best correlation with overall ISA, as well as correlation with both positive ISA
 299 (attraction) and negative ISA (repulsion), was exhibited by Pearson's tetrachoric correlation (C_{pears}),
 300 Forbes index (C_{forbes}), and the mid-P variant of Fisher's Exact Test (C_{FETmP}) for binary data, and

301 Spearman's correlation (CA_{rho}) and Hellinger distance (CA_{hell}) for abundance data (Fig. 6). The best
 302 overall Spearman correlations between the simulated ISA and its estimated measures were around
 303 0.75 (Fig. 6), which we attribute to the inability of the spatially implicit measures to capture the
 304 distance-dependent part of ISA (Fig. 3). We found no clear indication that either abundance or
 305 incidence-based metrics performed best. We also found that measures based on correlation (C_{pears} ,
 306 CA_{rho} , CA_{cor} , CA_{cov}), or on deviation from an expected null association (C_{forbes} , C_{FETmP}) performed
 307 better in capturing ISA, while similarity measures (C_{sor} , C_{jacc} , CA_{bray} , CA_{ruz}) performed worse. This is
 308 also in line with our supplementary analysis of Z-scores (Figs. S6-S8). We found that refining the
 309 spatial grain of the analysis, and increasing the magnitude of CSA, had mostly negative or no effect
 310 on the performance of the metrics (Figs. S7-S8), although there were exceptions.

311 Importantly, we found that the spatially implicit metrics tended to saturate at extreme negative (ISA <
 312 10) association (Fig. S6), most likely because the community matrices are identical above these
 313 extreme ISA values (demonstrated in Fig. 3). Thus, most abundance-based metrics showed weak
 314 overall correlation with repulsion (Fig. 6b), with the exception of the distance-based Ruzicka (CA_{ruz})
 315 and percentage difference (CA_{bray}) indices. Lastly, we found that when the metrics were averaged at a
 316 given spatial grain, they varied in their relationship with grain (Fig. S7).

317 *Spatially explicit measures and spatially explicit simulations*

318 Here, we used the 2-species simulations to illustrate community variograms and pair correlation
 319 functions. Our goal was to show the potential of these methods, rather than to perform the same
 320 comprehensive evaluation as above, since these techniques do not measure ISA in a single number,
 321 but as a function of spatial distance r . Because of that, they have the potential to actually recover the
 322 shape of the entire probability density function $f_{sp2}(r)$ from Figure 5b. For the spatially explicit
 323 techniques, we performed 9 simulations that correspond to the panels of Figure 5a (the only difference
 324 is that we used 200 individuals per species). For the purpose of fitting the community variograms, we
 325 used a single grid resolution with 20 cells along each side of the simulated domain.

326 Community variograms (Fig. 7a) estimated the relationship between distance r and negative
 327 covariance in a way that matched the shape of the original $f_{sp2}(r)$ from Figure 5b. However, they
 328 were weak at capturing the spatial pattern of attraction when there was a confounding effect of strong
 329 conspecific dispersion (CSA > 0.01). In contrast to community variograms, pair correlation functions
 330 more accurately estimated the distance-dependent ISA (compare Fig. 5b with Fig. 7b), i.e. the
 331 estimated relationship between distance and pair correlation function closely matches the shape of
 332 $f_{sp2}(r)$ from Figure 5b. Similarly to community variograms, there was a somewhat limited
 333 performance for attraction and high CSA values, but the problem was less severe. Overall, the point
 334 pattern-based bivariate pair correlation functions provided the most complete picture of the "true"

335 simulated ISA pattern.

336 RECOMMENDATIONS

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

339 *Best indices*

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

354 From the other indices for pairwise binary data, among the all-purpose metrics that capture both
355 repulsion and attraction, we recommend the mid-P index (C_{FETmP}), Forbes's coefficient of association
356 (C_{forbes}), or Pearson's tetrachoric correlation (C_{pears}); the latter is also recommended by Hubálek
357 (1982). Jaccard's index (C_{jacc}) is a good alternative, since it captures both negative and positive ISA,
358 and it can easily interpreted as proportional overlap, although it may not reflect the overall ISA as
359 accurately as the correlation-based indices. For pairwise abundance data, apart from covariance- and
360 correlation-based parametric indices, we also advocate for the Spearman rank-rank correlation (C_{rho}),
361 since it requires no prior transformation, and it captures both negative and positive ISA reliably.

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

368 *ISA is better spatially explicit*

369 Our results highlight a serious limitation of spatially implicit indices of ISA; they fail to consider
370 spatial proximity of two species when they are already 100% spatially segregated (Fig. 3). The
371 problem becomes more severe towards finer spatial resolutions. This partly explains why none of the
372 examined spatially implicit metrics gave a perfect correlation with the simulated distance-dependent
373 ISA, and why we observed the saturation of performance towards extreme values of ISA in our
374 simulations (Fig. S6). To address this systematically, we recommend that whenever there is
375 information on spatial position of the sites, indices of ISA are better considered as a function of
376 spatial distance. Both the community variograms and point pattern analysis offer straightforward ways
377 to do that.

378 Going spatially explicit (when data permit) also makes ISA more biologically interpretable, since it
379 can identify spatial distances over which biotic interactions really matter. This can be done, for
380 example, by integrating community variograms into joint species distribution models (JSDM, Warton
381 et al. 2015, Ovaskainen et al. 2017, Zurell et al. 2018). Specifically, in a JSDM, the spatially implicit
382 species-by-species covariance (or correlation) matrix Z can be replaced by distance-dependent $Z(r)$,
383 as also mentioned by Ovaskainen et al. (2017). Interestingly however, the spatially explicit approach
384 can also separate ISA caused by short-distance interactions from ISA caused by the environment
385 *without* the need of modelling the effect of the environment, as demonstrated by Wiegand et al.
386 (2012). The trick is to use null models that only randomize individuals locally, i.e. only within
387 neighborhoods with radius R , which removes potential signals of small-scale interactions at scales $r <$
388 R , if we can reasonably assume that environmental conditions within the neighborhood are constant
389 (Wiegand et al. 2012).

390 *Spatial scale*

391 Most of the commonly studied facets of biodiversity depend on spatial scale. In the case of ISA, scale
392 has been approached from two angles. The first focuses on ISA as a function of the average area of a
393 site in a community matrix, which is equivalent to spatial resolution (grain) of a grid (Økland 1994,
394 Hui 2009, Segurado et al. 2012, Araújo and Rozenfeld 2013, McNickle et al. 2018). These studies
395 show that ISA is grain-dependent, but neither theory nor empirical observations predict a systematic
396 direction of the grain dependence; ISA-area relationships can be increasing, decreasing, or hump-
397 shaped. One thing that complicates consensus is the use of various metrics of ISA across studies. Our
398 simulations (Fig. S7) show that selection of the ISA metric affects the grain dependency; for example,
399 some metrics may have slightly hump-shaped relationship with grain [e.g. C-score, Fig. S7, see also
400 McNickle et al. (2018)], while other metrics may monotonically increase or decrease with grain (Fig.
401 S7). This is something that the field needs to sort out before any empirical scaling of ISA is explored
402 and interpreted.

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

413 *Null models and Z-scores*

414 Most of the literature on ISA emphasizes the importance of null models, be it spatially implicit
415 (Ulrich and Gotelli 2013) or explicit (Wiegand and Moloney 2014). In Figures S7 and S8, we show
416 that transforming ISA indices to Z-scores makes them more robust against the confounding effects of
417 conspecific aggregation or varying resolution, while (Ulrich et al. 2018) show that null models also
418 account for the confounding effect of total species richness (γ). The same sentiment underlies
419 Wiegand and Moloney (2014): to analyze spatially explicit patterns of ISA, one should contrast them
420 against a null expectation of no ISA. However, null can be computationally demanding, which can be
421 a problem for large biodiversity data; in such cases analytical shortcuts may be handy. For example,
422 in spatially implicit methods, measures such as C_{FETmP} or C_{forbes} already have the deviation from the
423 expected randomness implicit in their definition.

424 *Measuring macroecological patterns of ISA*

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

434 Given that there is an ever increasing availability of biodiversity data from large spatial extents (Jetz
435 et al. 2012), we see an exciting opportunity here for exploration of empirical macroecological patterns
436 of ISA, particularly through community variograms. Some of the possible macroecological patterns to

437 explore with community variograms are latitudinal or altitudinal gradients of ISA, and their
438 relationship with patterns such as distance decay of compositional similarity, with distance-dependent
439 con-specific aggregation, or with broad geographic patterns of species richness and rarity. This has
440 not yet been done, yet bringing ISA to macroecology may provide new explanations for old patterns,
441 for example through the hypothetical (albeit controversial) link between ISA and biotic interactions,
442 or through a completely unexplored link between ISA and diversification.

443 ISA IN THE CONTEXT OF BIODIVERSITY

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

450 *ISA and alpha diversity*

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

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

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

$$463 \quad \bar{\alpha} = \frac{\sum_{j=1}^n \alpha_j}{n} = \sum_{i=1}^S \frac{O_i}{n} \quad (1)$$

464 Importantly, O_i and O_i/n are spatially implicit, meaning that it does not matter which sites are
465 occupied, or whether species are spatially attracted or segregated; as long as O_i is constant, no re-
466 arrangement of occupied cells in space has an effect on $\bar{\alpha}$. That is, $\bar{\alpha}$ is insensitive to ISA. In contrast,
467 $\bar{\alpha}$ is sensitive to changes in O_i/n , which is a measure of conspecific aggregation (CSA). For example,

468 consider the scenario in Figure 8a in which there was a shift in the ISA from segregated to attracted.
469 Because O_i , n and S remain constant, $\bar{\alpha}$ will not be affected by the change in ISA. Hence, an empirical
470 assessment of biodiversity which summarizes net change of $\bar{\alpha}$ over many locations may be blind to
471 changes of ISA. Further, this insensitivity of $\bar{\alpha}$ propagates to species-area relationships and to species
472 accumulation curves diversity, as we mention below.

473 *ISA and gamma diversity*

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

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

490 The second proposition operates on relative abundances. If the total area and density of individuals
491 are constant, but there is an increase of the total S , then the average per-species abundance must
492 decrease, which will reduce the likelihood of detecting inter-specific co-occurrences in communities
493 (Hubbell and Foster 1986, Lieberman and Lieberman 2007, Wiegand et al. 2007b, 2012, Volkov et al.
494 2009, Rajala et al. 2019).

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

502 to the metacommunity, the weaker their observed average negative association gets.

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

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

512 *ISA and beta diversity*

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

518 The connection between ISA and beta-diversity is best illustrated on a spatially implicit site-by-
519 species community matrix Y . In short, ISA is the association among *species* (Simberloff and Connor
520 1979, Hubálek 1982, Bell 2005, Legendre and Legendre 2012), whereas beta diversity is the
521 similarity among *sites*. In this simplified case, both ISA and beta diversity are calculated using exactly
522 the same data and indices (Tables 2, 3); the only difference between them is whether they are applied
523 to the rows ("R-mode" of Legendre & Legendre 2012) or columns of Y ("Q-mode"). In other words,
524 any index of beta diversity that is normally applied to sites can be applied to species and can be
525 meaningfully interpreted as an index of ISA, and vice versa (Legendre and Legendre 2012, Arita
526 2017). For example, Araújo and Rozenfeld (2013) define a "co-occurrence score" as the "ratio of the
527 number of geographical cells where species A and B co-occur to the total number of occupied cells".
528 Although not stated, this is equivalent to the classic Jaccard index for comparing site similarity. Thus,
529 the list of measures that have been typically used to measure ISA (Tables 2 and 3) can be expanded
530 with Jaccard or Sørensen-type indices (Arita 2017). Inversely, the list of commonly used measures of
531 beta diversity (Koleff et al. 2003) can be expanded by some typical ISA indices. Here we see a
532 potential for C_{forbes} (Table 2) which explicitly quantifies the deviation of the observed ISA (or beta
533 diversity) from the ISA (or beta diversity) expected under random distribution of incidences (Forbes
534 1907). One exception is the Whittaker index $(\frac{S}{\alpha} = \frac{n}{\bar{n}}$, where \bar{n} is the average number of occupied sites

535 per species), which is identical for both the analysis by sites ($\frac{S}{\alpha}$) and by species ($\frac{n}{\bar{\alpha}}$); however, this
 536 index is different from the pairwise indices of both beta diversity and ISA since it does not capture the
 537 within-matrix similarity between sites or species; instead, it only reflects the proportional matrix fill,
 538 i.e. the fraction of cells in Y filled with 1s (Arita 2017).

539 However, little has been written about the actual relationship between beta diversity and ISA within a
 540 given spatial domain, i.e. given constant n , S , and CSA. In other words, what happens with beta
 541 diversity if we vary ISA and keep everything else intact? We know that Whittaker's index ($\frac{S}{\alpha} = \frac{n}{\bar{\alpha}}$)
 542 must be insensitive to ISA, which is demonstrated in Fig. 8, and which follows from the
 543 aforementioned insensitivity of $\bar{\alpha}$ to ISA (eq. 1). However, we are unaware of any study directly
 544 focusing on the link between ISA and pair-wise beta diversity. In Figure 8 we show a case of average
 545 pairwise between-site Jaccard beta diversity being *sensitive* to ISA, and this is new. Šizling et al.
 546 (2011) and McGlinn and Hurlbert (2012) hint on a potential explanation by showing the relationship
 547 between average pairwise Jaccard beta diversity and the Whittaker's index, which is modulated by the
 548 occupancy frequency distribution [see also McGeoch and Gaston (2002)]. This reasoning could
 549 perhaps be extended to provide a link between pairwise beta diversity and ISA.

550 In contrast to the spatially implicit indices for binary and abundance data, the connection between ISA
 551 and beta diversity is well known in analyses of point patterns (Wiegand and Moloney 2014). The ISA-
 552 beta connection can be demonstrated in the spatially-explicit version of Simpson's evenness index
 553 $\beta(r)$ (Shimatani 2001, Wiegand and Moloney 2014 section 3.1.5.1). Unlike the traditional spatially
 554 implicit version of the Simpson's index (Simpson 1949) (i.e., the probability that two randomly
 555 selected individuals are heterospecifics), which is a measure of evenness, $\beta(r)$ is a measure of beta
 556 diversity, since it captures dissimilarity over a given distance (Shimatani 2001) (i.e., the probability
 557 that two randomly selected individuals distance r apart are heterospecifics). The index is defined as:

$$558 \quad \beta(r) = \frac{\sum_{i=1}^S \sum_{\substack{j=1 \\ j \neq i}}^S f_i f_j \frac{g_{ij}(r)}{g(r)}}{\sum_{i=1}^S f_i^2} = 1 - \frac{\sum_{m=1}^S f_m^2 g_{mm}(r)}{g(r)} \quad (2)$$

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

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

573 *ISA, species-area relationships and species-accumulation curves*

574 Here, we demonstrate that species-area relationships and rarefaction curves are insensitive to ISA. We
575 have stated that $\bar{\alpha}$ and S in a given domain are insensitive to ISA, given that spatial extent or a study
576 remains constant. These are the two components of Whittaker's index of beta diversity (Whittaker
577 1960), which is $S/\bar{\alpha}$, and so Whittaker's index is insensitive to ISA. It means that nested species-area
578 relationships (SAR), which are exactly related the Whittaker index over a continuous range of $\bar{\alpha}$ and S
579 (Crist and Veech 2006) must also be insensitive to ISA (Fig. 2).

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

583
$$S(r) = \sum_{i=1}^Y H_i(r), \quad (3)$$

584 where $S(r)$ is number of species present within r from an arbitrarily chosen "test" location, H_i is the
585 spherical contact distribution function for species i , which is the probability that the first neighbor of
586 species i is distance r away from the test location. $S(r)$ becomes a species-area curve when r is
587 converted to πr^2 . Importantly, the H_i is insensitive to ISA, since it is only based on the locations of
588 species i . We note that point pattern analysis also has a scaling curve that is sensitive to ISA: the
589 individual species-area relationship (ISAR; Wiegand et al. 2007) which always focuses on a focal
590 species f :

591
$$IS(r) = \sum_{i=1}^Y D_{fi}(r), \quad (4)$$

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

595 Finally, we turn to species-accumulation curves, from which the classical examples are the spatially
596 implicit individual-based and sample-based rarefaction curves (Gotelli and Colwell 2001), and their
597 spatially-explicit versions (McGlenn et al. 2019). In the former, individuals or samples are
598 accumulated randomly, irrespectively to their spatial position, which effectively breaks any pattern of
599 both CSA or ISA, making the spatially implicit curves indeed insensitive to ISA. In the latter spatially

600 explicit case, samples (plots) are accumulated by nearest neighbors, which makes these curves closely
601 related to SARs, which we have shown to be sensitive to CSA, but insensitive to ISA. This is in line
602 with the core idea of partitioning of rarefaction curves to their components (McGlenn et al. 2019),
603 which are the regional species-abundance distribution, density of individuals, and con-specific
604 aggregation, but not ISA.

605 WHY SHOULD BIODIVERSITY SCIENTISTS CARE ABOUT ISA?

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

614 *ISA as evidence for interactions.* Perhaps the best-known, albeit perennially controversial, reason for
615 analyzing patterns of ISA is a notion that they give hints about biotic interactions among species
616 (Gotelli et al. 2010, Blois et al. 2014, Harris 2016, Morueta-Holme et al. 2016, Thurman et al. 2019,
617 Calatayud et al. 2020). The effort that gained traction in the 1980's and 1990's (Cody and Diamond
618 1979, Connor and Simberloff 1979), and has recently been revived with the promise of joint species
619 distribution models as a tool to disentangle interactions from shared environmental requirements
620 among species (Warton et al. 2015, Ovaskainen et al. 2017, Zurell et al. 2018). The various
621 approaches of revealing interactions from presence-absence co-occurrence data, as well as a suite of
622 cautionary arguments against the endeavor, has been summarized recently by Blanchet et al. (2020).
623 While some of the arguments presented in Blanchet et al. (2020) on presence-absence data can be
624 remedied by analysis of abundance or point pattern data, we agree that the utility of ISA as a direct
625 evidence for interactions is limited. Nevertheless, we argue that there are other reasons for measuring
626 ISA, which we give below.

627 *ISA when interactions are given.* Sometimes biotic interactions are not what we want to infer from
628 the data, because we already know how the species interact. Examples are well documented trophic
629 interactions or mutualistic interactions. These known interspecific relationships can generate
630 hypotheses concerning geographic patterns of ISA. For example, a large-scale assessment of
631 biodiversity change may specifically look at patterns of ISA among pairs of pollinators and plants; if
632 there is a significant trend of spatial segregation over time, it may indicate a potential disruption of
633 pollination services, a finding that could be further investigated with additional data or experiments.

634 Here we see a particularly exciting prospect in bringing together network ecology with biogeography.

635 *ISA improves forecasts and predictions.* ISA and its patterns are useful even when assuming no
636 mechanistic underpinning, because patterns can improve predictions and forecasts based on inductive
637 logic, rather than causality. Examples of classical predictive biodiversity patterns are the endemic-
638 area relationship predicting extinctions due to habitat loss (Keil et al. 2015), or richness-environment
639 correlations that can be used for spatial interpolations and predictions of diversity patterns (Algar et
640 al. 2009). Similarly, measurements of ISA can potentially improve estimates of other metrics, for
641 example estimates of site-specific species pools (Karger et al. 2016, Bruelheide et al. 2020) or
642 predictions of species distributions in joint species distribution models (Harris 2016, Norberg et al.
643 2019). In both cases, the reason for why ISA can improve the predictions is not only as a proxy for
644 species interactions, but also because co-occurring species can act as proxies for suitable
645 environments which we may be difficult to measure in the field.

646 *ISA as a summary statistic.* ISA captures a unique spatial pattern of communities, on a par with other
647 popular summary statistics such as species richness or beta diversity. Simply reporting how richness
648 or beta diversity vary geographically repeatedly proved to be among the most powerful starting
649 stimuli in the field, generating countless explanatory and testable hypotheses (Brown 1995, Gaston
650 2000, Lomolino et al. 2010), and such patterns can be as useful for our understanding as the processes
651 that generated them (Currie 2019). An example is the very existence of latitudinal and altitudinal
652 diversity patterns, which have fascinated ecologists for centuries. Similar approach has recently
653 gained traction in summarizing empirical patterns of co-occurrence. For example, Lyons et al. (2016),
654 Tóth et al. (2019) and Calatayud et al. (2020) have documented broad-scale patterns of ISA, and
655 although they do offer interpretations involving biotic interactions, these are part of the post hoc
656 interpretation of the documented patterns, rather than the main goal of the analyses.

657 *ISA as a benchmark for theories or mechanistic models.* Rosenzweig and Abramsky (1997) describe
658 the idea of “dipswitch theory”, i.e. a theory that makes a bundle of unique predictions, which are then
659 compared with different empirical patterns. The ability of the theory to fit not just one, but multiple
660 patterns, is then a step towards a “stronger” test of the theory (McGill 2003). Here we argue that ISA
661 can allow for stronger tests of theories by providing a unique biodiversity pattern that a theory needs
662 to fit. For instance, it has been demonstrated that neutral theory (Hubbell 2001) or the maximum
663 entropy theory (Harte 2011) both reproduce realistic species-area relationships, but patterns of ISA,
664 along with other features, might help to distinguish among theories [see e.g. Ulrich (2004) and Bell
665 (2005)]. Here, ISA seems particularly promising since it is largely independent on some of the
666 classical patterns such as species-area relationship, or patterns of beta diversity, as we have
667 demonstrated. ISA can also be used together with inverse approaches in individual-based models
668 where known (or hypothesized) individual-level interactions are explicitly modeled and ISA patterns

669 emerge at the community level (Grimm et al. 2005). The underlying processes structuring the
670 community can then be inferred by testing how closely the emerging patterns of the model match the
671 observed data. For example May et al. (2015) used a neutral individual-based model to quantitatively
672 predict patterns observed in a 50 ha tropical forest plot, including beta-diversity (eq. 2). Surprisingly,
673 the model was able to match five emerging patterns simultaneously, but was unable to match the
674 species-area relationship and beta-diversity simultaneously, pointing to missing processes. The ability
675 to explain patterns of ISA can thus provide useful information for validating theories and mechanistic
676 models.

677 CONCLUSION

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

690 CODE AND DATA

691 Code and data used for this study are archived in R package ‘spasm’. The version submitted with this
692 publication is at Zenodo (10.5281/zenodo.3944504). The latest development version is at GitHub
693 (<https://github.com/petrkeil/spasm>).

694 ACKNOWLEDGEMENTS

695 The contributions of PK, JMC and DM were facilitated by support from German Centre for
696 Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the German Research
697 Foundation (FZT 118). We thank to Nathan Kraft and an anonymous referee for helpful comments.

- 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.
- Baddeley, A., E. Rubak, and R. Turner. 2016. *Spatial point patterns: methodology and applications with R*. CRC Press, Taylor & Francis Group, Boca Raton, FL.
- 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*:ele.13525.
- Blois, J. L., N. J. Gotelli, A. K. Behrensmeyer, J. T. Faith, S. K. Lyons, J. W. Williams, K. L. Amatangelo, A. Bercovici, A. Du, J. T. Eronen, G. R. Graves, N. Jud, C. Labandeira, C. V. Looy, B. McGill, D. Patterson, R. Potts, B. Riddle, R. Terry, A. Tóth, A. Villaseñor, and S. Wing. 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., S. R. Supp, L. H. Antão, A. Bates, H. Bruelheide, J. M. Chase, F. Moyes, A. Magurran, B. McGill, I. H. Myers-Smith, M. Winter, A. D. Bjorkman, D. E. Bowler, J. E. K. Byrnes, A. Gonzalez, J. Hines, F. Isbell, H. P. Jones, L. M. Navarro, P. L. Thompson, M. Vellend, C. Waldock, and M. Dornelas. 2019. The geography of biodiversity change in marine and terrestrial assemblages. *Science* 366:339–345.
- 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.
- Braun-Blanquet, J. 1964. *Pflanzensoziologie: Grundzüge der Vegetationskunde*. Springer, Wien.
- Brown, J. H. 1995. *Macroecology*. University of Chicago Press.
- 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, IL.
- Bruehlheide, H., B. Jiménez-Alfaro, U. Jandt, and F. M. Sabatini. 2020. Deriving site-specific species pools from large databases. *Ecography*:ecog.05172.
- Calatayud, J., E. Andivia, A. Escudero, C. J. Melián, R. Bernardo-Madrid, M. Stoffel, C. Aponte, N. G. Medina, R. Molina-Venegas, X. Arnan, M. Rosvall, M. Neuman, J. A. Noriega, F. Alves-Martins, I. Draper, A. Luzuriaga, J. A. Ballesteros-Cánovas, C. Morales-Molino, P. Ferrandis, A. Herrero, L. Pataro, L. Juen, A. Cea, and J. Madrigal-González. 2020. Positive associations among rare species and their persistence in ecological assemblages. *Nature Ecology & Evolution* 4:40–45.
- Cody, M. L., and J. M. Diamond, editors. 1979. *Ecology and evolution of communities*. Belknap Press, Cambridge, MA.
- 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, NY.
- 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. *The 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. 2019. spacemakeR: Spatial modelling. R-Forge.
- Dray, S., R. Pélessier, P. Couteron, M.-J. Fortin, P. Legendre, P. R. Peres-Neto, E. Bellier, R. Bivand, F. G. Blanchet, M. De Cáceres, A.-B. Dufour, E. Heegaard, T. Jombart, F. Munoz, J. Oksanen, J. Thioulouse, and H. H. Wagner. 2012. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* 82:257–275.
- 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://dynamicsecology.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, MA.
- 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* 281:20140922–20140922.
- Goodman, L. A., and W. H. Kruskal. 1979. *Measures of association for cross classifications*. Springer, New York, NY.
- 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 DC.
- 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* 107:5030–5035.
- Gotelli, N. J., E. M. Hart, and A. M. Ellison. 2015. *EcoSimR: Null model analysis for ecological data*. CRAN.
- 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.
- 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.
- 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, NJ.
- 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, NY.
- 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. *Scientific Data* in press.
- 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.
- Karger, D. N., A. F. Cord, M. Kessler, H. Kreft, I. Kühn, S. Pompe, B. Sandel, J. Sarmiento Cabral, A. B. Smith, J.-C. Svenning, H. Tuomisto, P. Weigelt, and K. Wesche. 2016. Delineating probabilistic species pools in ecology and biogeography: Probabilistic species pools. *Global Ecology and Biogeography* 25:489–501.
- 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.
- Lebart, L., and J. P. Fénelon. 1971. *Statistique et informatique appliquées*. Dunod, Paris.
- 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.
- 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.
- Lyons, S. K., K. L. Amatangelo, A. K. Behrensmeyer, A. Bercovici, J. L. Blois, M. Davis, W. A. DiMichele, A. Du, J. T. Eronen, J. Tyler Faith, G. R. Graves, N. Jud, C. Labandeira, C. V. Looy, B. McGill, J. H. Miller, D. Patterson, S. Pineda-Munoz, R. Potts, B. Riddle, R. Terry, A. Tóth, W. Ulrich, A. Villaseñor, S. Wing, H. Anderson, J. Anderson, D. Waller, and N. J. Gotelli. 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. *Proc. R. Soc. 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.
- McGlenn, 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 DC.
- 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., L. N. Hudson, S. L. Hill, S. Contu, I. Lysenko, R. A. Senior, L. Börger, D. J. Bennett, A. Choimes, B. Collen, J. Day, A. De Palma, S. Díaz, S. Echeverria-Londoño, M. J. Edgar, A. Feldman, M. Garon, M. L. K. Harrison, T. Alhousseini, D. J. Ingram, Y. Itescu, J. Kattge, V. Kemp, L. Kirkpatrick, M. Kleyer, D. L. P. Correia, C. D. Martin, S. Meiri, M. Novosolov, Y. Pan, H. R. P. Phillips, D. W. Purves, A. Robinson, J. Simpson, S. L. Tuck, E. Weiher, H. J. White, R. M. Ewers, G. M. Mace, J. P. W. Scharlemann, and A. Purvis. 2015. Global effects of land use on local terrestrial biodiversity. *Nature* 520:45–50.
- Norberg, A., N. Abrego, F. G. Blanchet, F. R. Adler, B. J. Anderson, J. Anttila, M. B. Araújo, T. Dallas, D. Dunson, J. Elith, S. D. Foster, R. Fox, J. Franklin, W. Godsoe, A. Guisan, B. O'Hara, N. A. Hill, R. D. Holt, F. K. C. Hui, M. Husby, J. A. Kålås, A. Lehikoinen, M. Luoto, H. K. Mod, G. Newell, I. Renner, T. Roslin, J. Soinen, W. Thuiller, J. Vanhatalo, D. Warton, M. White, N. E. Zimmermann, D. Gravel, and O. Ovaskainen. 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.
- Økland, 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*. R package version 2.5-5, <https://cran.r-project.org> (accessed 28 Nov 2019).
- 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.
- 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.
- 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. *Proc. SPIE*. San Diego, CA.
- 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. *Quästiió* 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, IL.
- 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. *The 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 competition 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 Pub. House, Fairland, MD.
- Simpson, E. H. 1949. Measurement of diversity. *Nature* 163:688–688.
- 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. *The 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*. 2. ed. Cambridge University Press, Cambridge.
- Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. *The 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.
- Taylor, L. R. 1961. Aggregation, variance and the mean. *Nature* 189:732–735.
- 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.
- Tóth, A. B., S. K. Lyons, W. A. Barr, A. K. Behrensmeier, J. L. Blois, R. Bobe, M. Davis, A. Du, J. T. Eronen,

- J. T. Faith, D. Fraser, N. J. Gotelli, G. R. Graves, A. M. Jukar, J. H. Miller, S. Pineda-Munoz, L. C. Soul, A. Villaseñor, and J. Alroy. 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* 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. 2007a. How individual species structure diversity in tropical forests. *Proceedings of the National Academy of Sciences* 104:19029–19033.
- Wiegand, T., S. Gunatilleke, and N. Gunatilleke. 2007b. Species associations in a heterogeneous Sri Lankan dipterocarp forest. *The 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: Biological Sciences* 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, FL.
- 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.

700 **Table 1** Brief summary of processes that generate CSA and ISA, biodiversity patterns that are linked
 701 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 limitations, <i>conspecific interactions</i> among individuals, interplay between <i>niche width</i> and spatial aggregation of environmental conditions	dispersal limits common to 2 or more species , <i>interspecific interactions</i> among individuals, interplay between <i>niche overlap</i> and aggregation of environmental conditions
Which biodiversity facets and patterns are sensitive to it?	beta diversity and its distance decay, species-area and some species-accumulation curves, altitudinal and latitudinal gradients of range size and endemism, spatial gradients of diversity (Storch et al. 2008)	still mostly unclear
Approaches to measure it.	spatial over- or under-dispersion 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

702

703 **Table 2** Measures of ISA for binary community data. From more than 70 measures (Hubálek 1982,
 704 Legendre and Legendre 2012, Keil 2019) we have included those that have either been popular,
 705 recommended based on suitable properties, or that represent a distinct approach to ISA.

Measure	Reference	Symbol	Formula or description	R code in
A) Pairwise				
C-score	(Stone and Roberts 1990)	C_{seg}	bc	bipartite::C.score EcoSimR::c_score vegan::designdist
Scaled C-score	(Ulrich and Gotelli 2013)	C_{segSc}	$\frac{bc}{n(n-1)/2}$	vegan::designdist
Togetherness	(Stone and Roberts 1992)	C_{tog}	ad	bipartite::togetherness vegan::designdist
Scaled togetherness	(Ulrich and Gotelli 2013)	C_{togSc}	$\frac{ad}{n(n-1)/2}$	vegan::designdist
Jaccard similarity	(Jaccard 1901)	C_{jacc}	$\frac{a}{a+b+c}$	betapart::betapart vegan::betadiver
Dice-Sorensen similarity	(Dice 1945, Sørensen 1948)	C_{sor}	$\frac{2a}{2a+b+c}$	betapart::beta.pair vegan::betadiver
Coefficient of association	(Forbes 1907)	C_{forbes}	$\frac{an}{(a+b)(a+c)}$	vegan::designdist
Alroy's coefficient (a new take on C_{forbes})	(Alroy 2015)	C_{alroy}	$\frac{a(z+\sqrt{z})}{(a+b)(a+c)+a\sqrt{z}+\frac{bc}{2}}$ where $z = a + b + c$	vegan::designdist
Pearson tetrachoric correlation	(Pearson and Heron 1913)	C_{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, Tóth et al. 2019)	C_{FETmp}	$\left(\sum_{x_{min} \leq x \leq x_{max}} \frac{(i!x)(n-i!j-x)}{n!j} \right) - \frac{(i!a)(n-i!b)}{2(n!j)}$ Where $i = a+b$, $j = a+c$, and x is the possible number of overlaps, with $x_{min} = \max(a-d, 0)$ and $x_{max} = \min(i, j)$. The symbol $!$ denotes the 'choose' operation.	https://github.com/anikoboth/FCW
Matching coefficient	(Sokal and Michener 1958)	C_{match}	$\frac{a+d}{n}$	vegan::designdist
B) Matrix-wise				
Variance ratio	(Schluter 1984)	C_{ratio}	$S_r^2 / \sum \sigma^2$, where S_r^2 is the variance of per-site species richness and σ^2 is the sum of per-species variances of incidences.	EcoSimR::v_ratio https://github.com/mcgfinnlab/vario

Checker score	(Gotelli 2000)	$C_{checker}$	Number of species pairs forming perfect checkerboard distributions.	EcoSimR::checker
Number of unique species combinations	(Gotelli 2000)	C_{combo}	Number of species that always co-occur.	EcoSimR::species_combo
Network connectance	(Dormann et al. 2009)	C_{com}	$\frac{F}{\gamma n}$, where F is the number of all co-occurrences	bipartite::networklevel

706 *Notes:* For pairwise indices, a is the number of sites where both species co-occur, c and b are numbers
707 of sites occupied uniquely by each species respectively, d is the number of sites where none of the
708 species occur, $n = a + b + c + d$. For matrix-wise indices, where \bar{n} and \bar{a} are the mean number of
709 occupied sites per species the mean number of species per site respectively, and S and n are numbers
710 of all species and sites in the matrix respectively.

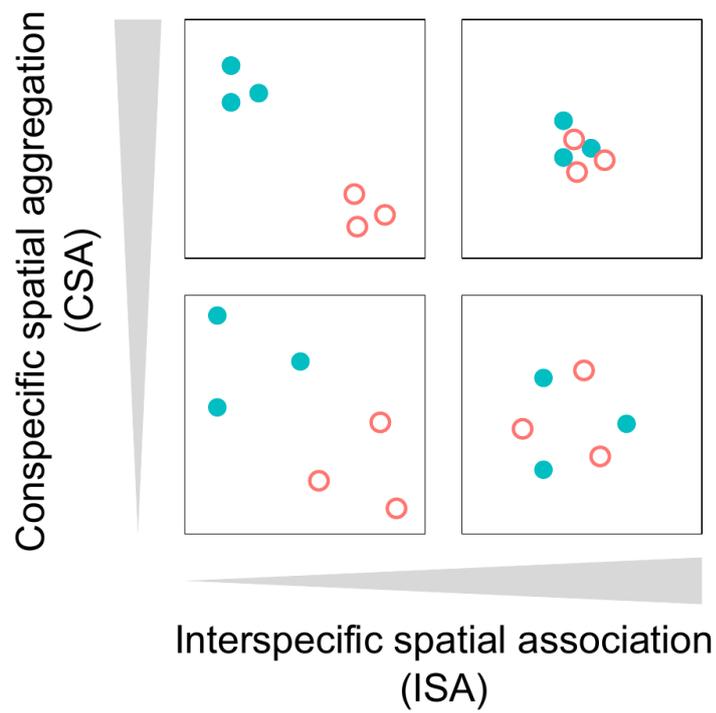
711

712 **Table 3** Select measures of ISA for abundance data. From the plethora of existing measures we have
 713 included those that have been popular, recommended, or that represent a distinct approach to ISA.
 714 Specifically, we selected three correlation-based indices, two distance-based indices, two abundance-
 715 based variants of the binary indices, and two matrix-wise indices.

Measure	Reference	Symbol	Formula or description	Notes	R code in
A) Pairwise					
Covariance	(Legendre and Legendre 2012)	CA_{cov} CA_{cov_hell}	$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$	With optional transformation (Hellinger, log, sqrt) of raw abundances.	stats::cov
Pearson correlation (scaled covariance)	(Legendre and Legendre 2012)	CA_{cov} CA_{cov_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}	$\frac{\sum_{i=1}^n x_i - y_i }{x_+ + y_+}$	One of the abundance-based variants of C_{sor} .	vegan::vegdist
Ruzicka 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 Ruzicka	(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}	The same principle as the binomial C_{ratio} .		

716 Notes: x and y are the vectors of abundance of two species, \bar{x} and \bar{y} are their means, x_+ and y_+ are
717 their sums, σ_x and σ_y are their standard deviations, x_i and y_i are abundances at site i , and n is the total
718 number of sites.
719

720
721



722
723
724
725

Figure 1 The difference between con-specific spatial association (CSA, vertical gradient) and inter-specific association (ISA, horizontal gradient), where the latter is the subject of this paper. Points are individuals, colors mark species.

Schools of thought	Typical data	Example metrics	Visual representations	Software																																				
<p>Assembly rules Cody & Diamond (1975) Ulrich & Gotelli (2010)</p> <p>Phytocoenology Ordinations Braun-Blanquet (1964) ter Braak (1988)</p>	<p>Incidence matrix</p> <table border="1"> <tr><td colspan="2"></td><td colspan="3">Sites</td></tr> <tr><td rowspan="3">Species</td><td>Sp1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>Sp2</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>Sp3</td><td>0</td><td>1</td><td>0</td></tr> </table> <p>Abundance matrix</p> <table border="1"> <tr><td colspan="2"></td><td colspan="3">Sites</td></tr> <tr><td rowspan="3">Species</td><td>Sp1</td><td>3</td><td>0</td><td>1</td></tr> <tr><td>Sp2</td><td>1</td><td>2</td><td>0</td></tr> <tr><td>Sp3</td><td>0</td><td>3</td><td>0</td></tr> </table>			Sites			Species	Sp1	1	0	1	Sp2	1	1	0	Sp3	0	1	0			Sites			Species	Sp1	3	0	1	Sp2	1	2	0	Sp3	0	3	0	<p>Species-wise (1 species) SCA metric</p> <p>Pair-wise metrics (2 species) Togetherness C_{tog} C-score C_{seg} Jaccard index C_{jacc} Covariance CA_{cov} Pearson correlation CA_{cor}</p> <p>Matrix-wise summary metrics (all species) Variance ratio C_{ratio} Connectance C_{conn} Mean pairwise metrics</p>	<p>Graph</p> <p>Ordination diagram, k-means clusters</p> <p>Corplot</p>	<p>Metrics: EcoSimR bipartite vegan spaa coocur</p> <p>Visuals: corrplot qgraph cluster ade4 vegan</p> <p>Signif. tests: EcoSimR</p>
		Sites																																						
Species	Sp1	1	0	1																																				
	Sp2	1	1	0																																				
	Sp3	0	1	0																																				
		Sites																																						
Species	Sp1	3	0	1																																				
	Sp2	1	2	0																																				
	Sp3	0	3	0																																				
<p>Geostatistics Wagner (2003)</p>	<p>Georeferenced Incidences at sites</p> <p>Georeferenced abundances at sites</p>	<p>Distance-dependent covariance, or other pairwise metrics</p>	<p>Community variogram</p>	<p>vario</p>																																				
<p>Point pattern analysis Wiegand & Moloney (2014)</p>	<p>Fully mapped sites</p> <p>Fully mapped region</p>	<p>$D_{12}(r)$ $K_{12}(r)$ Bivariate Pair Correlation Function (PCF) P-M classification Species Mingling Index Mean Compositional Information $F(r)$</p>	<p>Bivariate PCF plot</p> <p>P-M classification plot</p>	<p>Programita spatstat</p>																																				

727

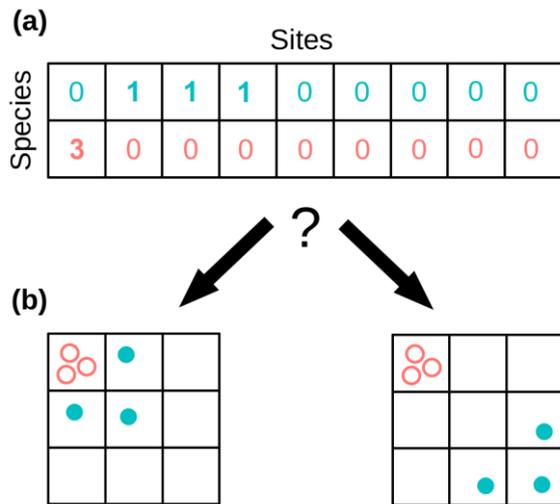
728

729

730

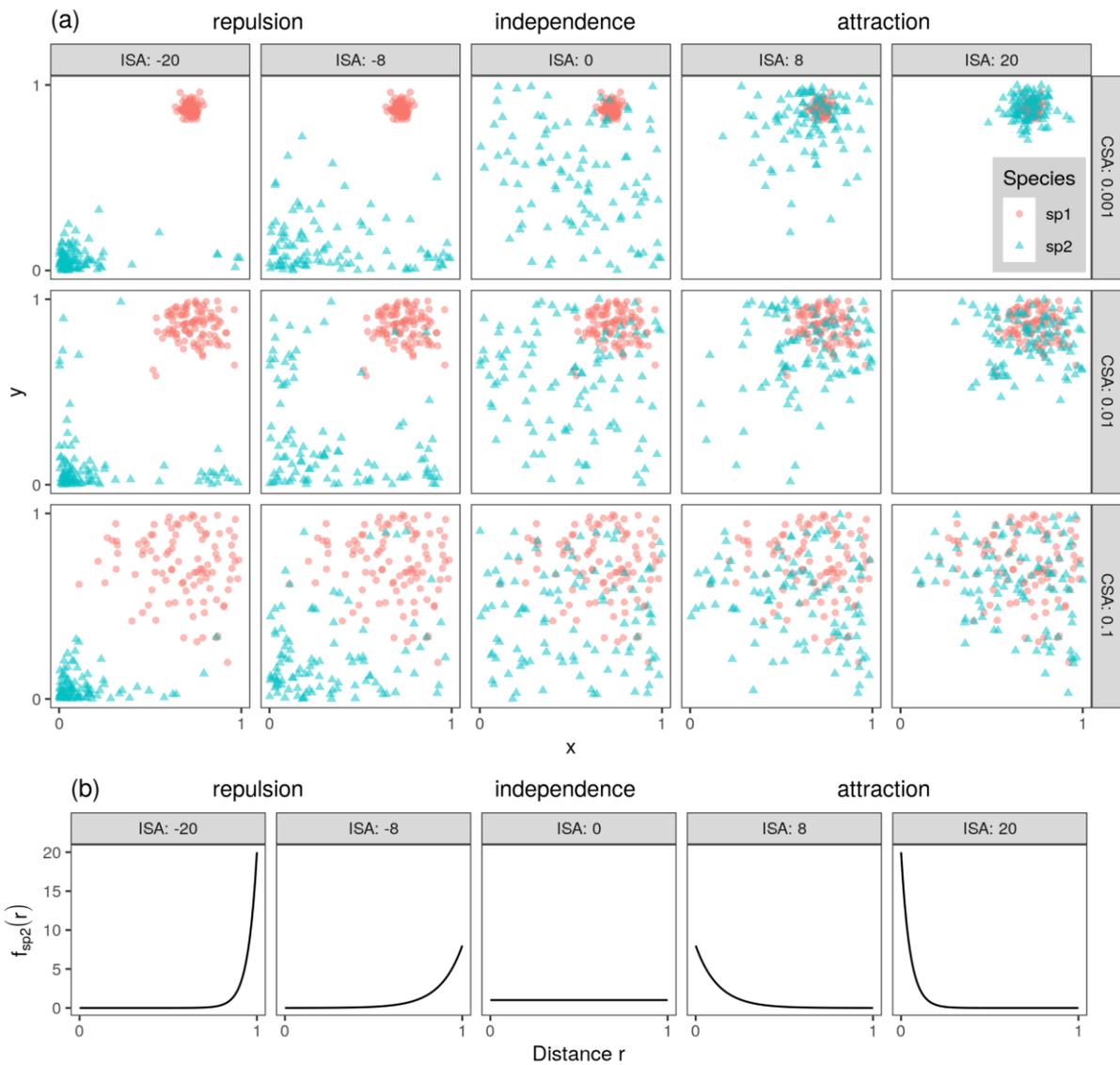
731

Figure 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 consists of R packages.

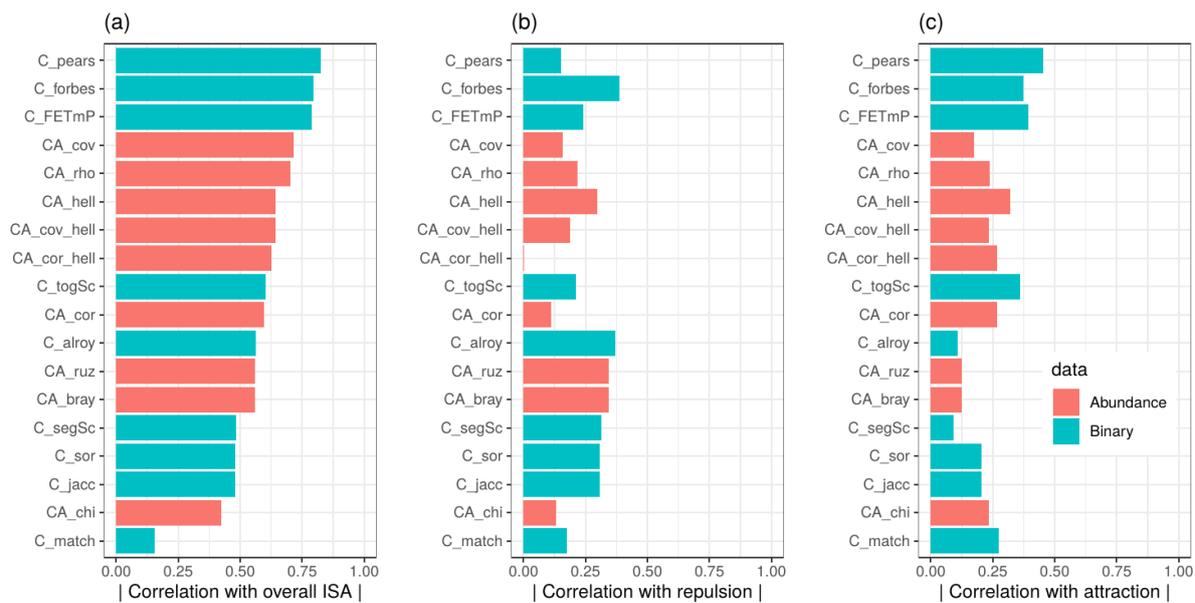


732

733 **Figure 3** The fundamental importance of spatial distance in ISA. A single community matrix (a) can
 734 potentially reflect two different spatial arrangements of the community (b), each with a different
 735 magnitude of ISA - the two species in the left spatial matrix are more attracted, while on the right they
 736 are more segregated, yet this spatial arrangement is not reflected in the community matrix.



746
 747 **Figure 5** (a) Simulated spatial distributions of individuals (points) of two species (sp1 and sp2) in a
 748 square domain under 3 levels of con-specific aggregation (CSA) of sp1 and 5 levels of inter-specific
 749 aggregation (ISA). (b) Truncated exponential probability density function [$f_{sp2}(r)$ (Keil 2014, 2019)]
 750 describes how likely we are to observe an individual of sp2 at a given distance from any individual of
 751 sp1. This $f_{sp2}(r)$ is convenient since its shape depends on a single parameter (here called ISA) which
 752 represents various magnitudes of inter-specific repulsion (left) and attraction (right) and their
 753 relationship with distance.
 754



755

756

757

758

759

760

761

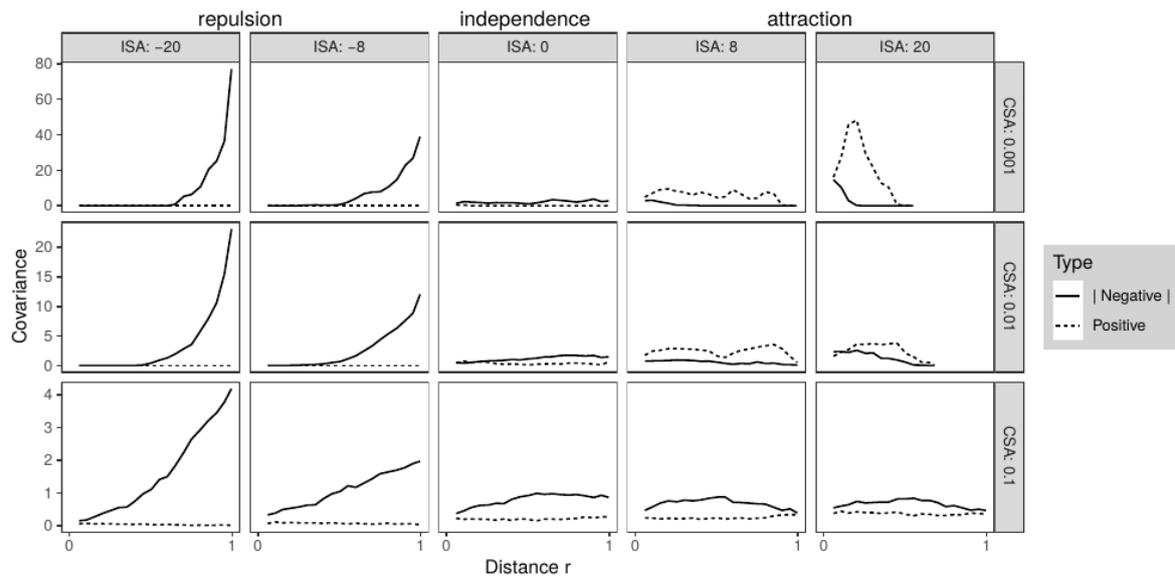
762

763

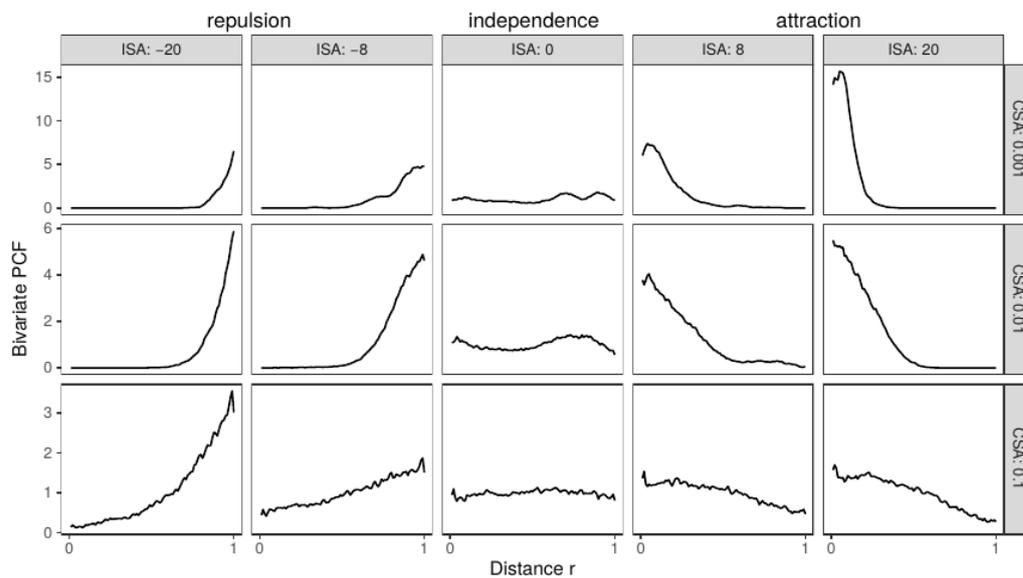
764

Figure 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 ($ISA > 0$). Given are absolute values of Spearman correlation coefficients. Note that some metrics very well separate negative from positive associations (i.e. they have good correlation with overall ISA), but within these two categories they have weak correlation with the underlying ISA parameter, e.g. covariance-based measures (CA_{cov} , CA_{cov_hell}). See Fig. S6 for details of these relationships.

(a) Community variograms

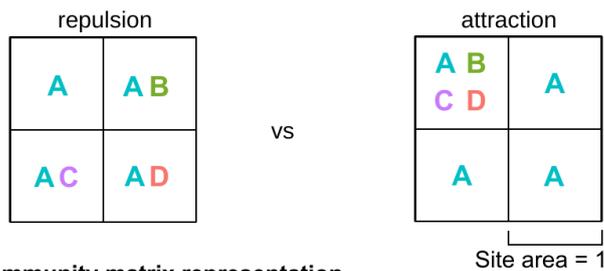


(b) Bivariate pair correlation functions

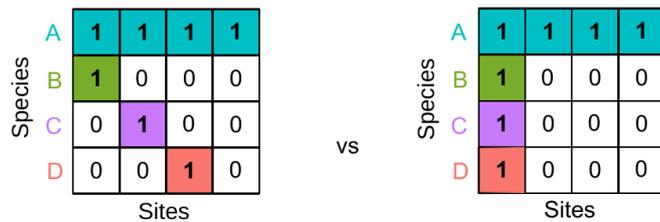


765
 766 **Figure 7.** Ability of two spatially explicit approaches to recover ISA patterns of the two species from
 767 Figure 4b. (a) Community variograms calculated on abundance data obtained by aggregation of the
 768 point pattern in a 20 x 20 pixel grid. Also, 200 points were simulated for each species, instead of the
 769 100 points in Figure 5. (b) Bivariate pair correlation function calculated directly from the point
 770 patterns.

(a) Spatial representation of presences of 4 species (A, B, C, D) at 4 sites



(b) Community matrix representation



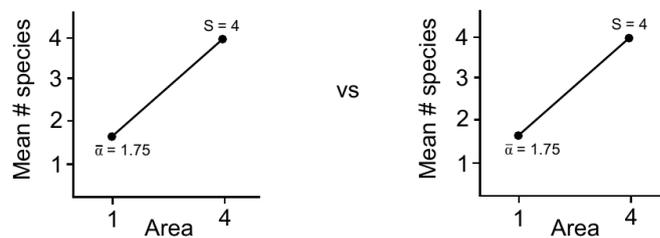
(c) ISA

mean $C_{jacc} = 0.13$ vs mean $C_{jacc} = 0.63$

(d) Beta diversity

Whittaker = 2.29 vs Whittaker = 2.29
 mean $\beta_{jacc} = 0.42$ vs mean $\beta_{jacc} = 0.63$

(e) Species-area relationships



771

772 **Figure 8** Insensitivity of species-area relationships and Whittaker's index to ISA, as also mentioned
 773 by Plotkin et al. (2000) and Storch (2016). (a) Two communities consist of 4 square sites each, with 4
 774 species (A, B, C, D) either present or absent. The communities differ in the magnitude of pairwise
 775 ISA; species in the left one are segregated, those in the right one are attracted to each other. (b) These
 776 communities can be described by spatially implicit community matrices Y from which metrics of ISA
 777 (c) can be calculated. While the Whittaker index remains constant in both communities, the mean
 778 pairwise C_{jacc} correctly reflects ISA. (d) The same logic applies when beta diversity measures such as
 779 β_w and β_{jacc} , where the former is simply a ratio of gamma diversity (total richness of a community) and
 780 mean alpha diversity (mean richness across sites), which is also the slope of slope of SAR, which
 781 remains constant even though ISA changes (e).

Supplementary material

This is supplementary material for manuscript “*Measurement and analysis of interspecific spatial associations as a facet of biodiversity*” by Petr Keil, Thorsten Wiegand, Anikó B. Tóth, Daniel J. McGlenn, and Jonathan M. Chase.

Appendix S1 - Glossary

Several attempts have been made to clarify the terminology of spatial associations and resemblance in the sites vs. species context (Hubálek 1982, Legendre and Legendre 2012), but the consensus is not entirely clear and many of the offered definitions are author- or study-specific. Here we provide what we see as the signal across the literature that we have reviewed:

- **Inter-specific** and **hetero-specific** are used synonymously for relationships among individuals of two or more species. Inter-specific has also been more popular than hetero-specific (Fig. S1).
- **Intra-specific** and **con-specific** are used synonymously for relationships among individuals within a single species.
- **Association** is a general term that describes mutual arrangement of spatial positions of two or more species, irrespectively to its direction, range of values, or particular interpretation.
- **Resemblance** is more general than association. Legendre and Legendre (2012) use association for relationships among species, while resemblance is the more general term applicable to both compositional similarity among sites and association among species.
- **Co-occurrence** can be used for occurrence of two or more species at a given site, or it can be used for a magnitude of the observed inter-specific association. It is usually reserved for binary (presence/absence) data.
- **Similarity** (and its mathematical complement, **dissimilarity**) are measures of resemblance that always take values between 0 and 1, are not metric, but can be converted/re-scaled to metric distances.
- **Distances** are measures of resemblance that have a minimum at 0, but can have various upper bounds (including infinity), depending on the constraints of the space in which they are measured.
- **Correlation-based metrics of resemblance** are centered around 0 (no relationship), with values > 0 for positive relationship, and < 0 for negative relationship. They can be converted/re-scaled to take values between 0 and 1.
- **Repulsion or segregation vs attraction or aggregation.** These usually refer to negative vs positive spatial associations among species, but they are sometimes used for con-specific patterns. A typical example is aggregation, which is often used for spatial clumping among conspecific individuals. We thus urge authors to clearly state which type (inter- vs intra-specific) they use, particularly in titles and abstracts of papers.
- **Biotic interaction or inter-specific interaction** is an effect of individuals of one species on individuals of other species (it can also be mutual). Examples are predation, mutualism, parasitism, or competition. Interactions may influence interspecific spatial associations, but usually together with other factors.

Appendix S2 – Literature search

We performed two literature searches in order to (i) assess the relative representation of ISA in papers dealing with biodiversity, and (ii) assess the relative representation of ISA in three specific ecological journals. Below we provide details on these searches.

Web of Science search

The goal of our first search was to assess which terms, metrics, and ecological patterns are most often associated with biodiversity papers, and how the relative frequency of ISA-related terms in biodiversity papers compares with the other topics. We searched Clarivate Web of Science (WoS) using the formula “biodiversity AND TERM” in the “Topic” field of the WoS search engine, where TERM is one of the search terms described below. When using the “Topic” field the engine searches for the TERM in titles, abstracts, author keywords, and Keywords Plus. We restricted the search to all papers published between 1996 and 2019. The search was done on 19 Sep 2019, with the exception of “trophic” and “competiti*” terms that were added on 15 June 2020. This is the list of the search TERMS:

- **TERMs associated with ISA:** attraction*, C-score, ecosim*, repulsion*, segregation*, variance NEAR/0 ratio*, cooccurr* OR co-occurr*, checker*, JSMD OR (joint NEAR/0 species), co-distrib* OR codistrib*, *coenos* or *cenos*
- **TERMs associated with biotic interactions and networks:** trophic, competiti*, network*, biotic NEAR/0 interact*
- **TERMs associated with classic biodiversity patterns:** beta NEAR/0 diversit*, evenness, jaccard, nestedness, simpson, species-area, species NEAR/0 richness, species NEAR/0 turnover, relative NEAR/0 abundance*, SAD, extinct*, shannon, invasi* OR invad*, geograph* NEAR/0 range*, functional, phylogenetic
- **General TERMs:** heterospecific* OR hetero-specific*, conspecific* OR con-specific*, inter-specific* OR interspecific*, intra-specific* OR intraspecific*
- **TERMs associated with spatially explicit approaches:** point NEAR/0 pattern* OR point-pattern*, variogram* or semivari*, pair NEAR/0 correl*, K-function, kriging, spatial* NEAR/0 autocor*, nearest NEAR/0 neighbor*, distance NEAR/0 decay*

Data generated by this search are at:

https://github.com/petrkeil/spasm/blob/master/data/wos_trends.csv

The code that we used to produce the figures is at:

https://github.com/petrkeil/spasm/blob/master/analyses/wos_trends_analysis.r

Results. In Figure S1, we show that terms associated with biotic interactions (e.g. networks, competition, trophic interactions), as well as classic biodiversity patterns, are more frequent by about an order of magnitude in biodiversity literature than terms associated with ISA. Even general ISA-related terms such as “co-occurrence” are considerably less common than some quite specific biodiversity terms such as “shannon” or “evenness”. In addition, in the general terms, the prefixes “inter-” and “intra-” are more common than “con-” and “hetero-”, which is in line with our decision to use the term “inter-specific associations”, rather than “hetero-specific associations”. Finally and unsurprisingly, we found that terms associated with spatially explicit analyses are rare, relatively to other terms. The reason is that spatially explicit analyses require considerably more data and expertise than implicit ones.

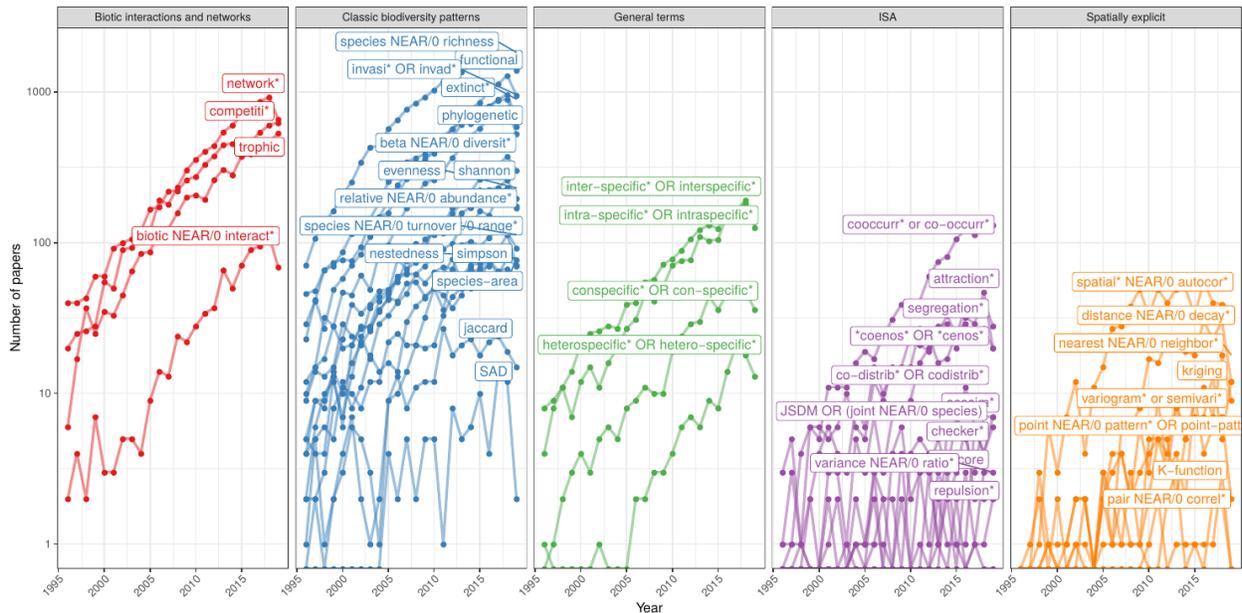


Figure S1 Temporal trends in frequency of particular TERMS in biodiversity-related papers. Y-axis shows the total number of papers published in a given year that are listed on Clarivate Web of Science, and that contain the given search TERM AND “biodiversity” in the “Topic” field. Note the \log_{10} scale of the y-axis. Left plot (red lines) shows terms related to biotic interactions and networks among species. Second to the left plot (blue lines) shows terms associated with classic biodiversity patterns in space and time. Second to the right plot (green lines) shows general terms used to distinguish between a within-species and across-species analyses. Second to the right plot (purple lines) shows terms related to ISA, and the right plot (orange) shows terms related to spatially explicit analytical approaches.

Manual search through *Ecology*, *Ecography*, and *American Naturalist*

In order to alleviate potential concerns about our selection of TERMS in the previous search, we conducted a second search that entirely avoids the use of WoS or search terms. One specific concern was that “biodiversity” as a term was coined as late as in 1988, which could have affected the trends. Thus, our goal here was to assess the popularity of ISA in a broader ecological context, rather than in the context of biodiversity only. We chose three ecological journals that have been important in the field for many decades and represent the breadth of current ecological research. These were:

- **Ecology** (Ecological Society of America), since it is a general and widely read ecological journal representing all facets of the discipline, from theoretical to applied.
- **American Naturalist** (The American Society of Naturalists), since it has emphasized theoretical papers, it also encompasses evolution, and it was prominent in the 1980s when ISA gained a particular prominence in ecology.
- **Ecography** (Nordic Society Oikos) since unlike the previous journals, it is not American, and it embraces pattern-oriented and macroecological approaches that may be under-represented in the more process- and hypotheses-oriented journals.

From these journals, we downloaded titles and abstracts of all papers published in 1995, 1999, 2003, 2007, 2011, 2015, and 2019, representing 3,858 papers in total. We read each of them and characterized the papers using 12 binary variables, as follows:

- **ISA.** Papers that describe, or explain, static patterns of ISA, as defined in this paper.
- **CSA.** Papers that describe, or explain, static patterns of conspecific spatial aggregation, which includes geographic range limits, abundance-occupancy relationships, occupancy-area relationships, fractal occupancy patterns, spatial autocorrelation of species distributions, measurements of deviations from complete spatial randomness, and patterns of endemism.
- **Spec_interaction.** Papers that deal with biotic interactions in a broad sense, which includes competition, mutualism, trophic interactions (herbivory, predation), parasitism, defense mechanisms, trophic networks, pollination, commensalism, complementarity, or facilitation.
- **Spec_coexistence.** Papers that deal with mechanisms of species coexistence.
- **Comm_assembly.** Papers that deal with community assembly. This is a rather vague category, and so we marked this for all papers that simply use the term community assembly in the abstract or title, or that deal with priority effects, assembly rules, community filters, environmental filters, metacommunities, and related subjects.
- **Relative_commonness.** Papers that describe, or explain, patterns of relative species abundances, which includes range-size distributions, species-abundance distributions, or occupancy distributions. We use the term commonness since it encompasses both abundance and range size.
- **Diversity.** Papers that describe, or explain, patterns of biological diversity within discrete spatial units, measured by various diversity measures such as species richness, species density, Simpson and Shannon indices, evenness, alpha or gamma diversity, functional diversity, genetic diversity, or phylogenetic diversity.
- **Species_accumulation.** Papers that describe, or explain, patterns of species accumulation with area, volume, sample size, or number of individuals, which includes species-area and species-volume relationships, ratios or regional to local richness, and rarefaction curves.
- **Beta.** Papers that describe, or explain, spatial patterns of species compositional dissimilarity among sites, which includes beta diversity, species turnover, nestedness, and distance-decay of similarity.
- **Beta_related.** Papers encompassing the two above-mentioned categories, i.e. both species accumulation and beta diversity.
- **Comm_comparison.** All papers that describe, or explain, composition and diversity across sites. This is a very broad category that encompasses all papers from the aforementioned three categories.
- **Div_comp_gradient.** All papers that describe diversity, beta diversity, and community composition as a function of environmental or spatial (geographic) gradients.

Results. Figure S2 shows that, among the 3,858 papers examined, there are fewer ISA-related papers than papers in other categories. The only categories that are similarly marginal are community assembly, relative commonness, and species accumulation. The most common topics among papers are species interactions, followed by a suite of topics that focus on comparisons of species diversity and composition among sites. Even beta diversity, which often requires the same data as an analysis of ISA, has been a more common topic than ISA.

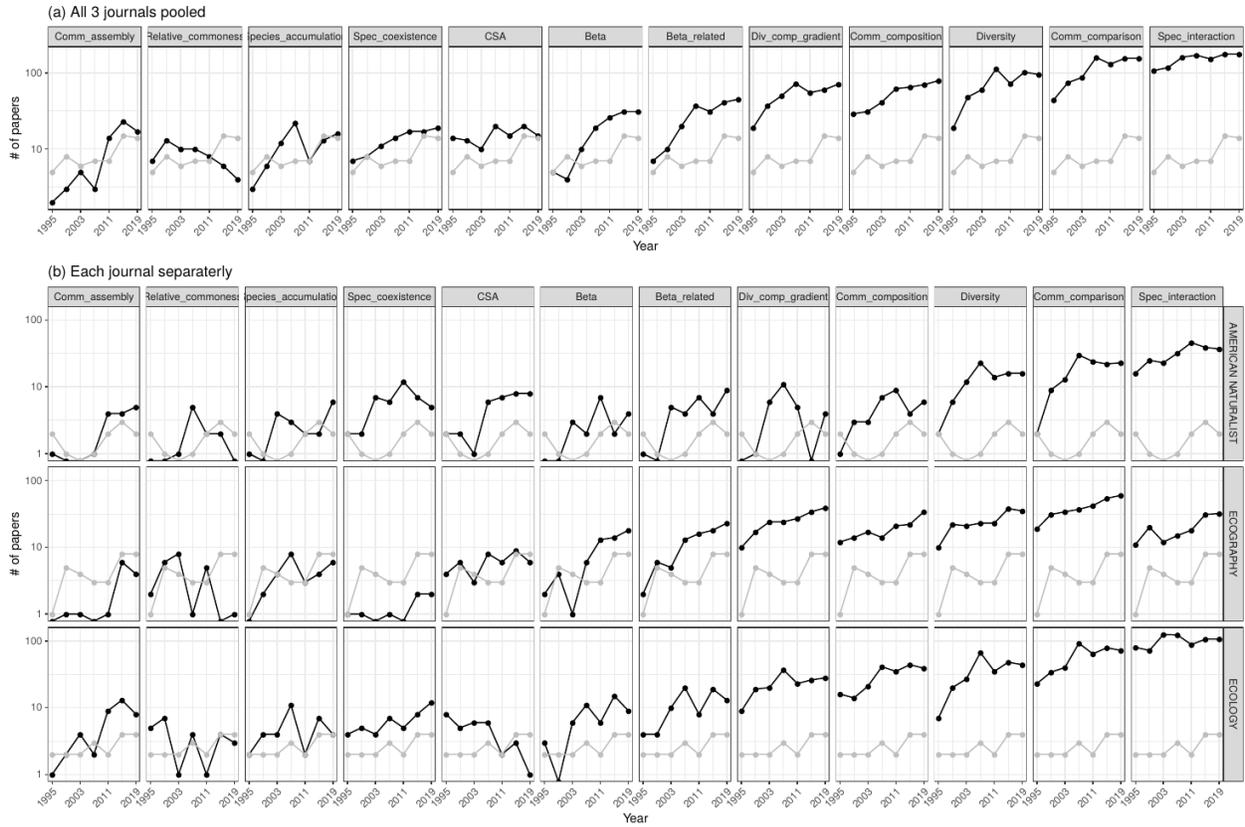


Figure S2 Temporal trends in counts of papers dealing with a given subject, published in Ecology, American Naturalist, and Ecography. Grey lines are number of papers dealing with ISA, black lines correspond to the topic in the header of each facet. Panel (a) gives counts in all three journals pooled together, panel (b) separates them by journal.

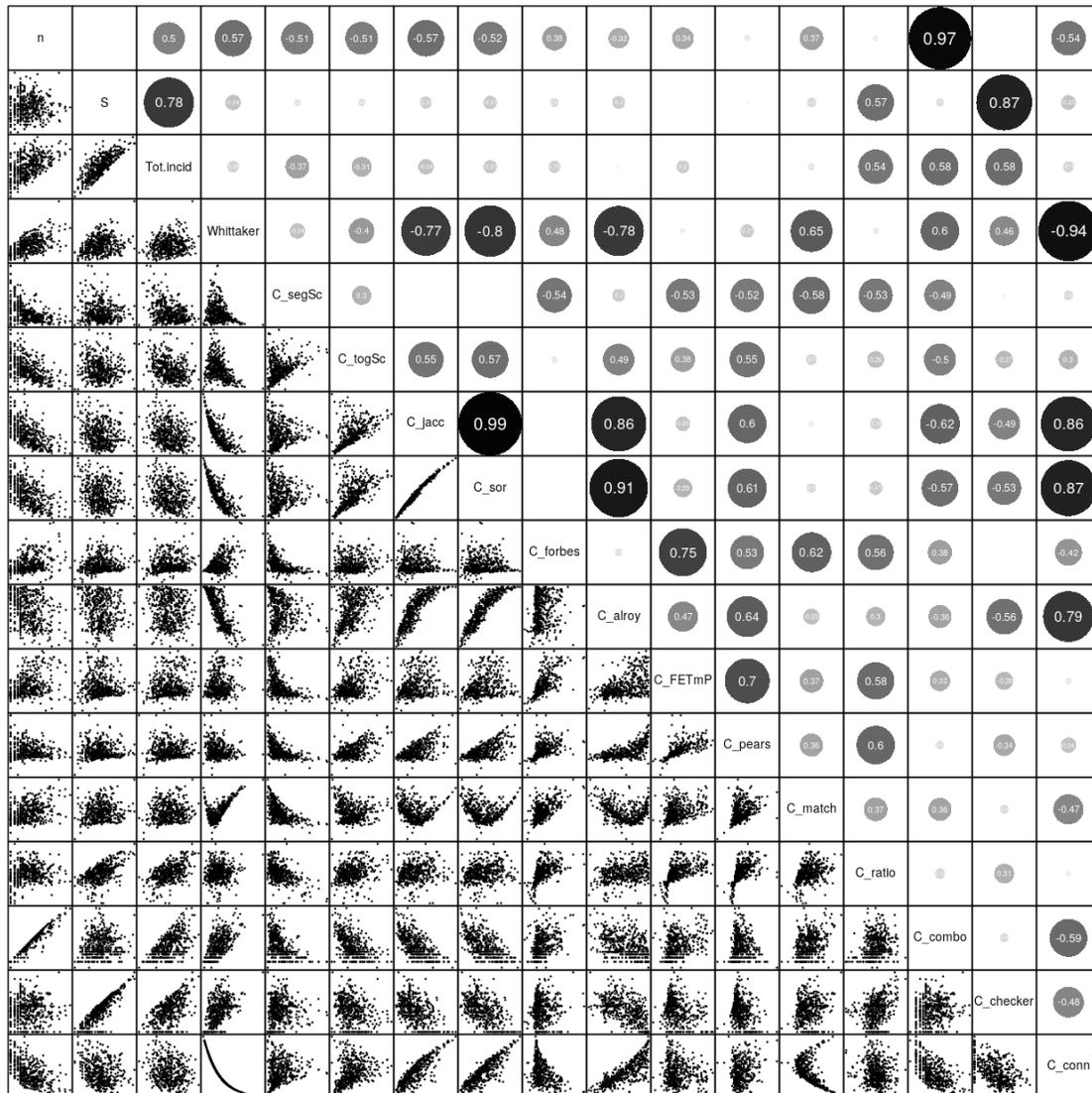


Figure S3 Correlations between ISA metrics for binary data from Table 2 applied to the empirical community matrices by Atmar and Patterson (1995) and Ulrich and Gotelli (2010). In case of the pairwise metrics, each point represents the mean value of the entire matrix. N , Gamma, and Tot.incid are numbers of sites, species, and incidences (i.e. matrix fill) respectively. C_{forbes} , C_w , C_{combo} , C_{ratio} , n , gamma, and Tot.incid were log-transformed. C_{segSc} , C_{togSc} , $C_{checker}$, were $\log(x+1)$ transformed. Numbers in the upper triangular section are Pearson correlations, with size and grey intensity proportional to the magnitude of the correlation.

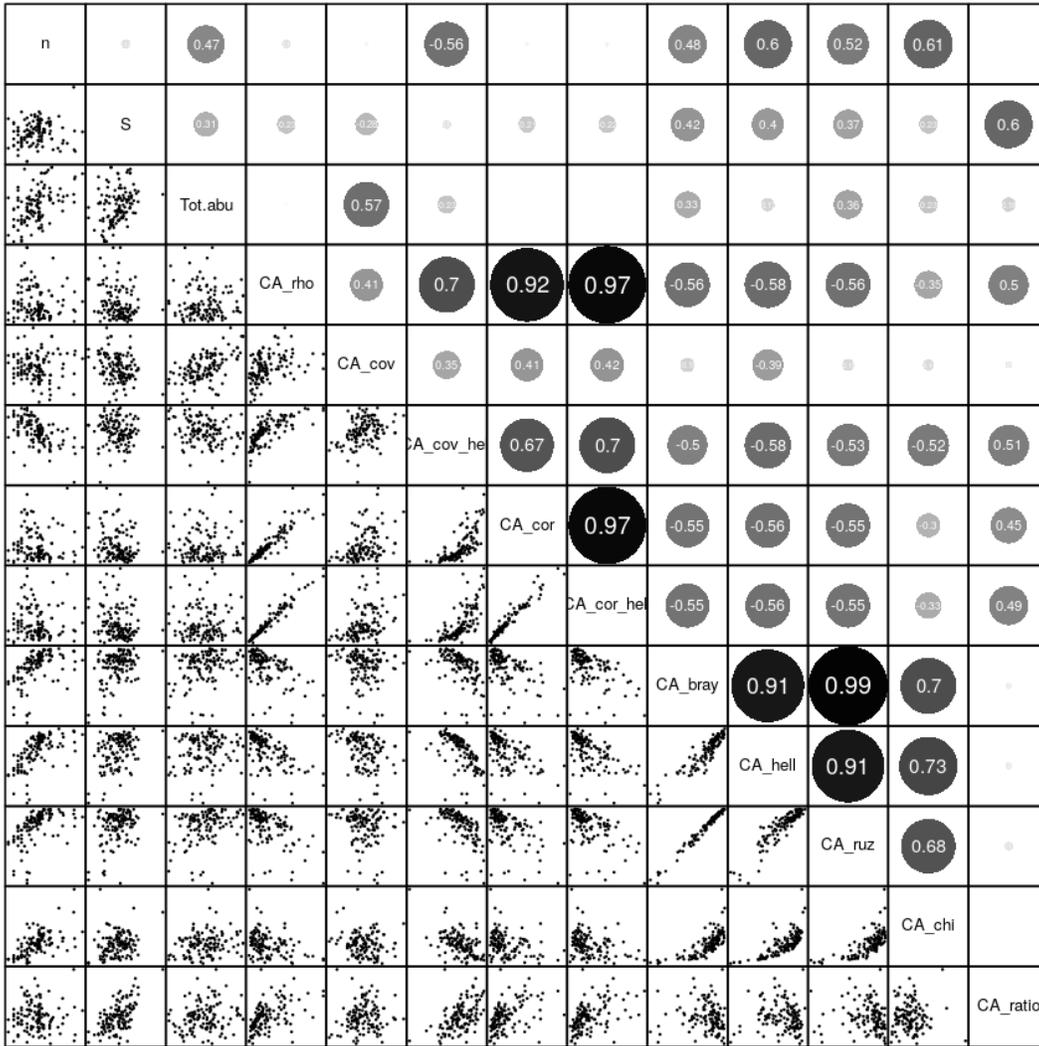


Figure S4 Correlations between ISA metrics for abundance data from Table 2 applied to the empirical community matrices by Ulrich and Gotelli (2010). In case of the pairwise metrics, each point represents a mean value of the entire matrix. *n*, gamma, and Tot.abu are numbers of sites, species, and the total number of individuals in the entire community matrix respectively. *CA_ratio*, *CA_cov*, *CA_cov_hell*, *n*, gamma, and Tot.abu were log-transformed. Numbers in the upper triangular section are Pearson correlations, with size and grey intensity proportional to the magnitude of the correlation.

Appendix S3 - Simulation procedure

Why only pairs? We refrained from simulating spatially explicit spatial association between more than two species, since the simulated patterns would need to achieve specified spatial association between each pair of species simultaneously, as well as the decay (or increase) of each of these associations with distance. Although the most recent advances in multivariate geostatistics may enable this (Genton and Kleiber 2015), these methods are still young, not widely available, and are computationally challenging to simulate cross-covariance structures of more than 3 layers (species). Further, there is the limit on the magnitude of negative associations in multi-species matrices (Brown et al. 2004) which would prevent us from exploring scenarios with strong negative ISA. As a result, we only evaluated pairwise measures of ISA. For each ISA metric we measured its correlation with the ISA parameter across all simulation setting.

Simulations. We simulated spatially explicit distributions of two species, sp1 and sp2 with abundances N1 and N2 respectively, as two point patterns in a square domain with side of 1 (Fig. 5a in the main text). One simulation proceeded as follows:

- (Fig. S5a) We chose a random point with coordinates μ_x and μ_y within the domain, with uniform probability density across the domain; this point was the center of distribution of sp1.
- (Fig. S5b) We created 2-dimensional probability density of points of sp1 as a bivariate normal distribution $f_{sp1}(\mu, \Sigma)$, where Σ is the covariance matrix with marginal variances $\sigma_x = \sigma_y = CSA$ and with covariance $\sigma_{xy} = 0$. μ is the vector of coordinates μ_x and μ_y .
- (Fig. S5c) We drew N1 points from that probability density surface. These are individuals of species 1.
- (Fig. S5d) For every location in the domain we calculated its distance r from the nearest point of sp1.
- (Fig. S5e) We transformed r using a truncated exponential function (Keil 2014) $f_{sp2}(r) = \frac{\alpha e^{\alpha r}}{e^{\alpha b} - e^{\alpha a}}$ where $r \in [a, b]$, where $a = 0$ and $b = 1$, but these truncation points can be set to any value depending on the size of the domain. We use the negative value of parameter α as the strength of interspecific association ($ISA = -\alpha$), with $ISA < 0$ being segregation, $ISA = 0$ for independence, and $ISA > 0$ being positive attraction between species.
- (Fig. 5b) We drew N2 points from from the $f_{sp2}(r)$. These are individuals of species 2.

We repeated this procedure for each combination of the following parameter values: $CSA \in \{0.001, 0.01, 0.1\}$, $N1 \in \{10, 100, 1000, 10000\}$, $N2 \in \{10, 100, 1000, 10000\}$, and for $ISA \in \{-20, -17.5, -15, \dots, 0, \dots, 15, 17.5, 20\}$, which we then aggregated to square spatial grids with $\{32, 16, 8, 4\}$ grid cells along each side. Thus, the value of 32 represents the finest (smallest) grain, and 4 the coarsest (largest) grain. Altogether, this produced 3264 simulated pairs of species, each represented by either a point pattern or a grid, and we further converted to 3264 community matrices Y , with either abundance or incidence values, and with known exact spatial position of each grid cell.

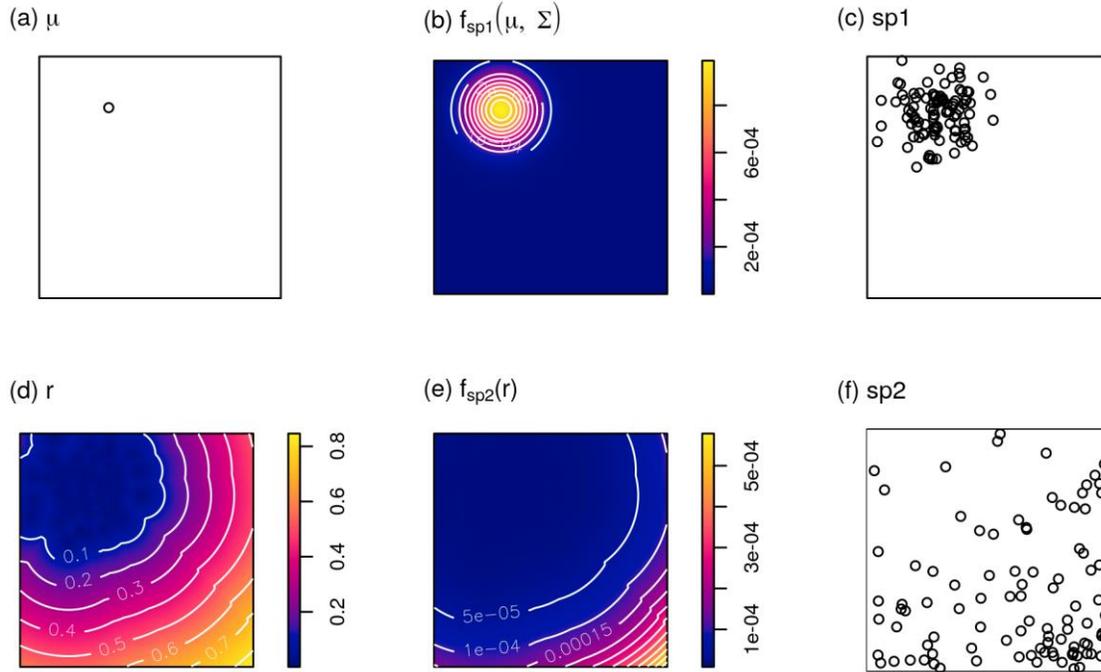


Figure S5 Simulation of two point patterns of two species, sp1 and sp2. The procedure is described in the main text. Briefly, point pattern of sp1 is simulated as a point process with bivariate normal probability density $f_{sp1}(\mu, \Sigma)$ with zero covariance and marginal variances describing the con-specific aggregation (CSA) of sp1. Point pattern of sp2 is then simulated as a point process with $f_{sp2}(r)$ describing the magnitude of ISA, where r is distance to the nearest point of sp1. This figure was also used in Keil (2019).

Analysis of the simulations. For each of the simulated community matrices Y (each with 2 species) we calculated the pairwise incidence-based ISA measures from Table 2, and the abundance-based measures from Table 3. In some of the measures we also tested their Z -score variant, calculated as $s(E_{raw} - E_{exp})/SD_{exp}$, where E_{raw} is the ISA metric (from Tables 2 and 3) calculated on observed data, E_{exp} is the null expectation of the index, and SD_{exp} is the standard deviation of the null expectation calculated over a set of community matrices obtained by 200 randomizations of Y . The randomizations were as follows: For the incidence matrices, we took the total number of incidences of each species, randomly assigned each of the incidence to a new site with uniform probabilities for each site, which corresponds to the “sim2” algorithm of EcoSimR package (Gotelli et al. 2015). For abundance-based measures, we randomly re-assigned each individual to a new site with uniform probability. Thus, in both algorithms, incidences or individuals are reshuffled only within each row of Y , irrespective of the matrix column totals. In case of the abundance data, we also ran the simulations with the IT algorithm of Ulrich and Gotelli (2010), but the results were similar. To evaluate performance of each metric, we calculated its Spearman’s Rho (rank-rank correlation) with the ISA parameter across all 4,080 simulations. All functions and scripts for the simulations and their analysis are provided at <https://github.com/petrkeil/spasm>.

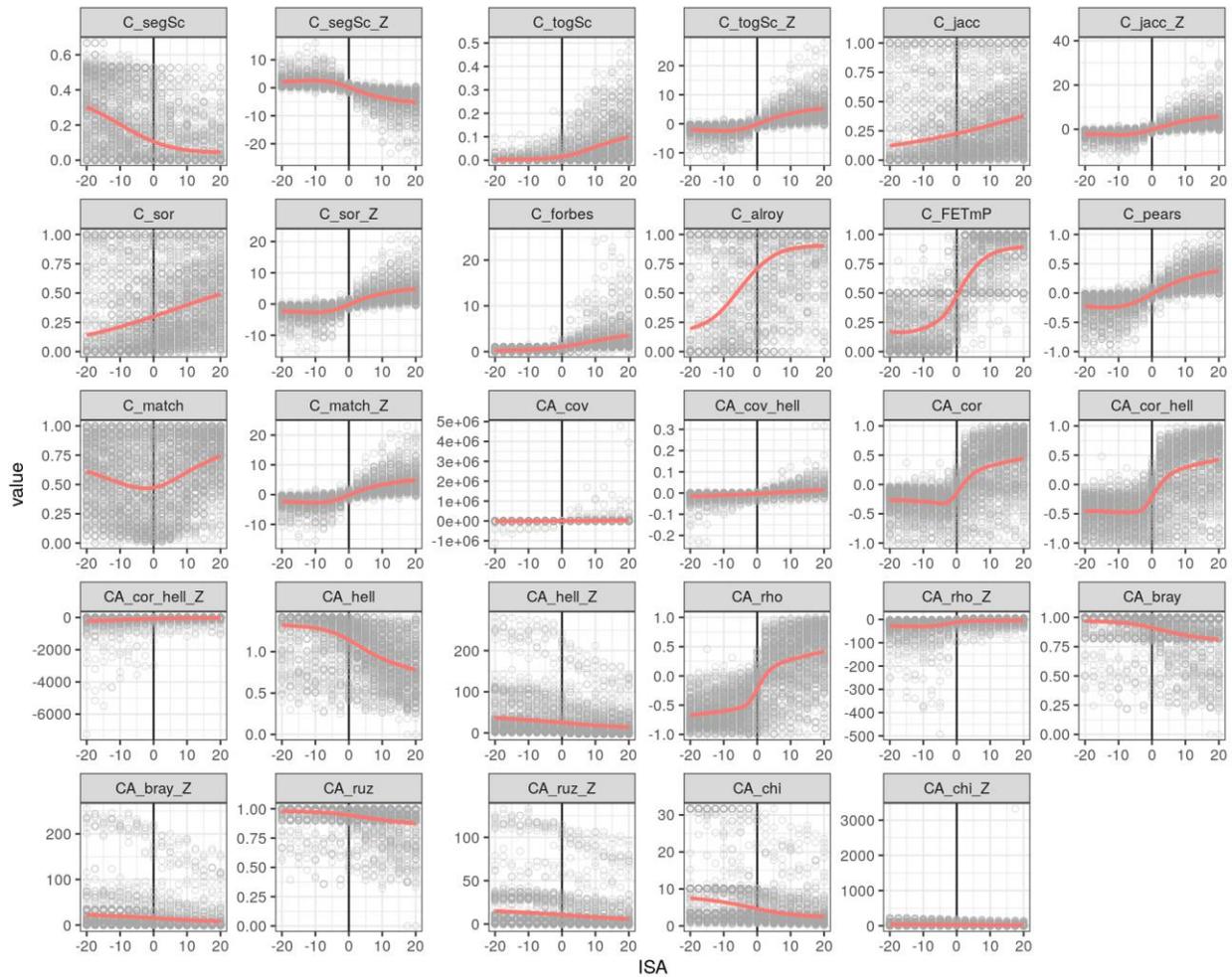


Figure S6 Relationships between the magnitude of ISA in artificial 2-species point pattern communities, and indices (measures) of ISA. These plots are the basis for Figure 6 in the main text.

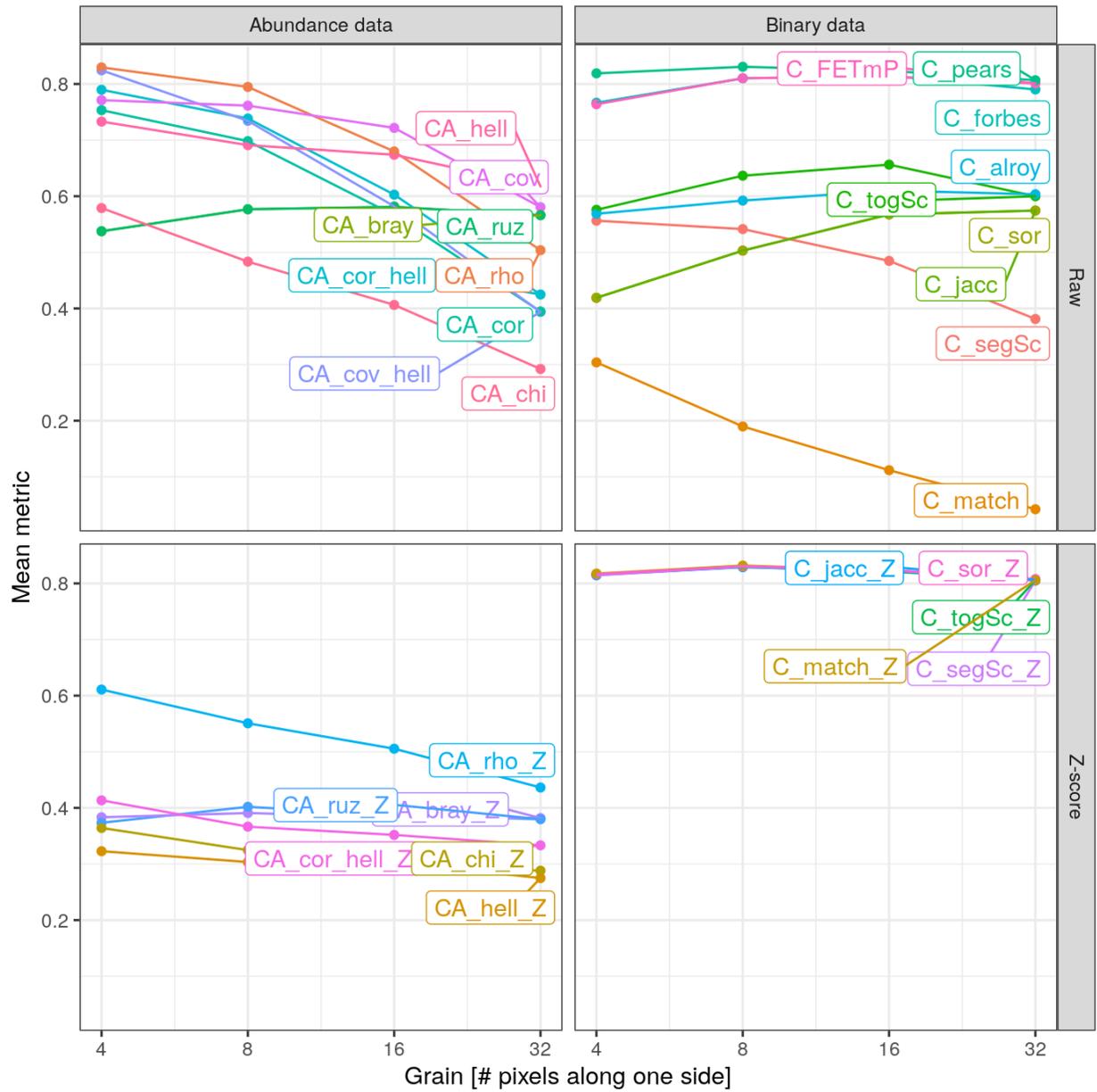


Figure S7 Relationship between the performance of ISA metrics and spatial grain that was used to aggregate point patterns. The performance is the absolute value of Spearman correlation between the parameter ISA and the average metric.

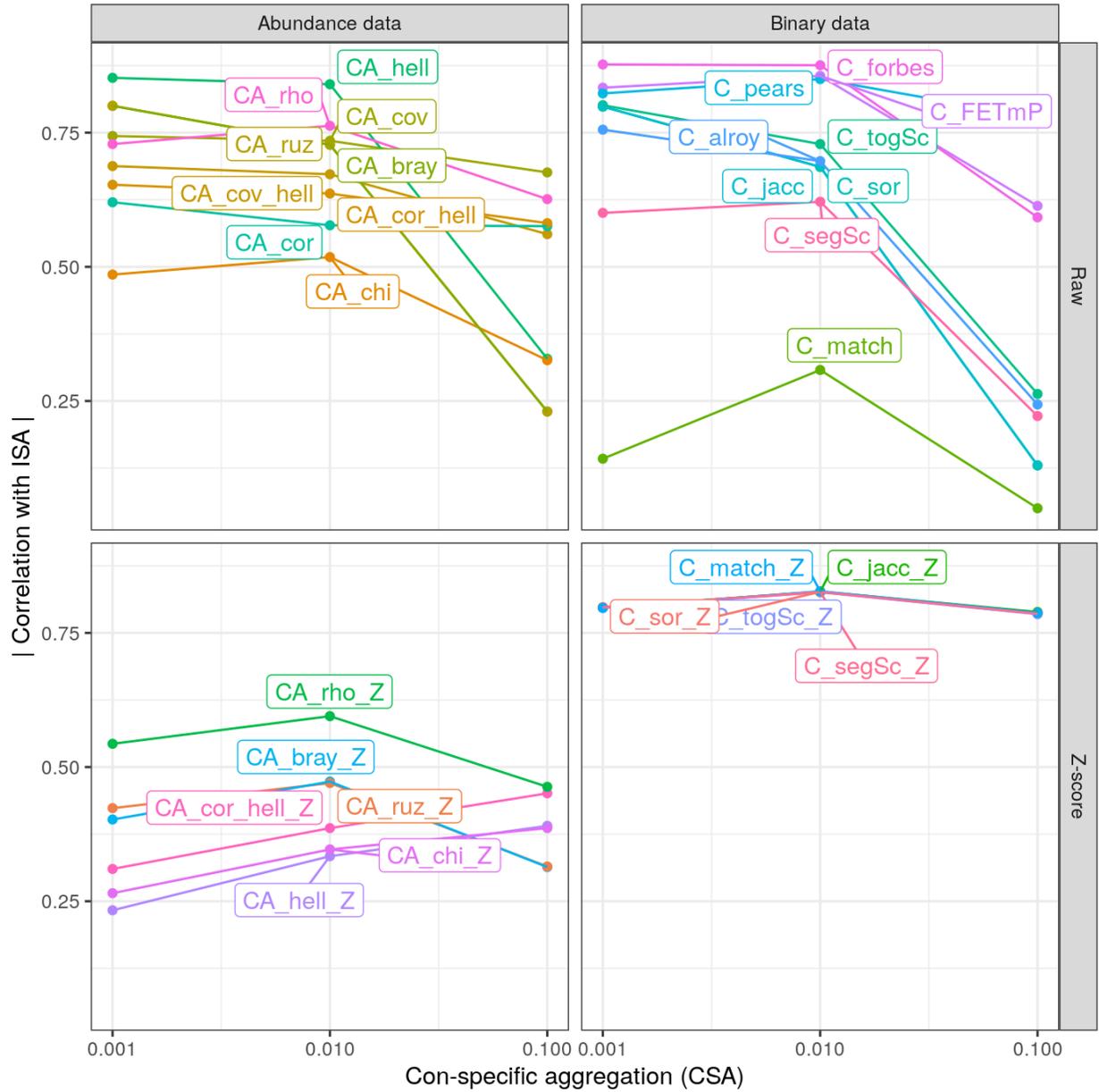


Figure S8 Relationship between the performance of ISA metrics and the magnitude of con-specific aggregation of species 1 in the simulations. The performance is the absolute value of Spearman correlation between the parameter ISA and the average metric.

Supplementary references

- Atmar, W., and B. D. Patterson. 1995. Nestedness temperature calculator. AICS Research Inc, University Park, New Mexico, and The Field Museum, Chicago.
- 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, IL.
- Genton, M. G., and W. Kleiber. 2015. Cross-covariance functions for multivariate geostatistics. *Statistical Science* 30:147–163.
- Gotelli, N. J., E. M. Hart, and A. M. Ellison. 2015. EcoSimR: Null model analysis for ecological data. CRAN.
- Hubálek, Z. 1982. Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. *Biological Reviews* 57:669–689.
- 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.
- Legendre, P., and L. Legendre. 2012. *Numerical Ecology*. Elsevier, Amsterdam.
- Ulrich, W., and N. J. Gotelli. 2010. Null model analysis of species associations using abundance data. *Ecology* 91:3384–3397.