

Jaccard dissimilarity in stochastic community models based on the species-independence assumption

Ryosuke Iritani^{1,2,*}, Vicente J. Ontiveros^{3,4}, David Alonso⁴, José A. Capitán⁵, William Godsoe⁶, and Shinichi Tatsumi⁷

¹RIKEN Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS), RIKEN, Wako, Saitama, Japan

²Department of Biology, Faculty of Sciences, The University of Tokyo

*lambtani@gmail.com

³Institute of Aquatic Ecology, University of Girona, Girona, Spain

⁴Theoretical and Computational Ecology, Center for Advanced Studies of Blanes (CEAB-CSIC), Spanish Council for Scientific Research, Acces Cala St. Francesc 14, Blanes E-17300 Spain

⁵Complex Systems Group, Department of Applied Mathematics, Universidad Politécnica de Madrid, Avenida Juan de Herrera, 6, Madrid E-28040 Spain

⁶BioProtection Research Centre, Lincoln University, Lincoln, New Zealand

⁷Hokkaido Research Center, Forestry and Forest Products Research Institute, Sapporo, Hokkaido, Japan

April 14, 2023

- Conception and design: RI, VJO, WG, DA, and ST;
- Acquisition of data: not applicable;
- Analysis: RI, WG, DA, JAC, and ST;
- Interpretation: RI and ST;
- Drafting: all;
- Revision: all;
- All authors approved the content of the manuscript for submission.
- ORCID iD:
 - RI** 0000-0002-2396-1109
 - VJO** 0000-0001-8477-2574
 - DA** 0000-0002-8888-1644
 - JAC** 0000-0002-6245-0088
 - WG** 0000-0003-1697-6916
 - ST** 0000-0002-1789-1685
- No conflict of interest to be declared exists.
- Running head: *Beta-diversity in a heterogeneous world*
- Keywords: Beta-diversity; Heterogeneity; Jaccard index; Null models; Similarity coefficient; Species Distribution Models; Stochasticity.

Abstract

Understanding how the spatial variation in species composition (beta-diversity) varies with biotic and abiotic conditions is one of the ultimate goals in biology. Theory predicts that beta-diversity is a consequence of two factors, species-level differences (defined as the variations among species in the probabilities that species are present in the landscape) and spatial heterogeneity (defined as the difference, between two sites, in the probabilities with which species are present). At present, however, the importance of each factor is unclear. Here, we take a probabilistic and combinatorial approach to examine the effects of species differences and spatial heterogeneity on the degree to which species assemblages in two spatial locations differ in species compositions. We first derived analytical and approximation formulae of the expectation and variance of the pairwise beta-diversity, based on the assumption that the presence probabilities of species are independent of each other. Though it seems intuitive that greater species differences leads to greater beta-diversity, our methods predict that the reverse is likely to occur under some circumstances; strikingly, when space is homogeneous, beta-diversity decreases with increasing species differences. This prediction suggests that policy making for increasing species traits-variation would, without adequately managing environmental heterogeneity, induce biotic homogenization, resulting in undesired outcomes. Second, we illustrate our method using data from five woodpecker species in Switzerland, showing that the woodpecker species' joint distributions change considerably with time, and also that such changes are basically explained by the changes in the incidences of some of the species. The new framework can improve our understanding of how pairwise beta-diversity responds to species differences and spatial heterogeneity.

1 Introduction

Beta-diversity (the spatial variation in species compositions) links diversity across scales (Whittaker 1972; Anderson *et al.* 2010; Chase *et al.* 2019; Poggiato *et al.* 2021). Beta-diversity varies with fundamental processes such as dispersal, environmental filtering and species interactions (Vellend 2010; Anderson *et al.* 2010; Socolar *et al.* 2016; Maynard *et al.* 2017; Legendre 2019; Thompson *et al.* 2020). Understanding the patterns of beta-diversity is thus considered as one of the ultimate goals in ecology. Reductions in beta-diversity, known as biotic homogenization (Olden & Poff 2003; Olden & Rooney 2006; Olden *et al.* 2018), have been caused by various types of global changes, such as urbanization (McKinney 2006), species invasion (Powell *et al.* 2013), climate change (Veech & Crist 2007), land-use change (Vellend *et al.* 2007), and interactions thereof (Karp *et al.* 2017). A consequence of biotic homogenization is the reduction of ecosystem functioning across the globe (Hautier *et al.* 2017; Mori *et al.* 2018; Albrecht *et al.* 2021; Wang *et al.* 2021). Understanding how beta-diversity changes in response to variations in biotic and abiotic conditions should lead to better management, conservation, and urban planning in our modern society (Crowther *et al.* 2015).

Beta-diversity is often measured by using pairwise indices based on empirical presence-absence (incidence) data (Koleff *et al.* 2003). Even for such simple, incidence-based beta-diversity, how biotic and abiotic variables influence the beta-diversity has not been conceptually established. That is, we know little about "beta-diversity patterns", the dependence of beta-diversity on biotic and abiotic factors. Indeed, results from previous work on beta-diversity patterns have been mixed. For example, theory shows that dispersal, tends to homogenize local communities and thereby reduce beta-diversity (Loreau 2000; Mouquet & Loreau 2003; Thompson *et al.* 2020), or to have rather opposing effects on beta-diversity (Lu *et al.* 2019; Lu 2021). Meanwhile, experimental work suggests that dispersal may promote beta-diversity (Vannette & Fukami 2017). Disturbances like fire also can have opposing effects on beta-diversity (Altermatt *et al.* 2011; Myers *et al.* 2015; Ojima & Jiang 2016). To better interpret complicated beta-diversity patterns, it is necessary to establish a general framework that can incorporate variable biotic and abiotic conditions.

Among other factors, species differences and spatial heterogeneity both have profound effects on beta-diversity (Peres-Neto *et al.* 2001; Legendre & Cáceres 2013). A rich amount of work has showed that spatial heterogeneity is a significant determinant of beta-diversity (Veech & Crist 2007; Báldi 2008; Jankowski *et al.* 2009; Allouche *et al.* 2012; Bar-Massada & Wood 2013; Heino *et al.* 2013; Pomara *et al.* 2013; Astorga

30 *et al.* 2014; Heino *et al.* 2014; Bar-Massada 2015a,b; Royan *et al.* 2015; Bar-Massada & Belmaker 2017;
31 Zorzal-Almeida *et al.* 2017; Ben-Hur & Kadmon 2020; Shinohara *et al.* 2022), but species difference can also
32 have a strong impact on beta-diversity (Silva *et al.* 2018; Santos *et al.* 2021). For example, dispersal mode
33 (passive vs. active) and body size (small vs. large) are both strong determinants of beta-diversity patterns as
34 revealed by meta-analyses (Soininen *et al.* 2007; De Bie *et al.* 2012). Similarly, Cao *et al.* (2021) showed that
35 beta-diversity may co-vary with niche-related characteristics. Recent studies have made use of functional
36 traits that reflect species' dispersal abilities and niche characteristics to explicitly explain metacommunity
37 processes such as dispersal limitation and environmental filtering (Kattge *et al.* 2020; Schrader *et al.* 2022).
38 It is, therefore, becoming increasingly important to explicitly consider the effects of both species differences
39 and spatial heterogeneity on beta-diversity (Peres-Neto *et al.* 2001; Gotelli & Ulrich 2011); however, we
40 know little about such effects on beta-diversity, and thus predicting the effects of global changes in species
41 traits and spatial heterogeneity on beta-diversity remains challenging

42 To discern beta-diversity patterns under variable conditions, it is of great use to regard both incidence-
43 data and resulting beta-diversity as stochastically varying quantities ("stochastic variables"), because
44 by doing so, we can assess how the probability distributions of beta-diversity varies with mechanistic
45 factors that determine species presence probabilities (Baroni-Urbani 1980; Real & Vargas 1996; Gotelli &
46 Ulrich 2011; Hui & McGeoch 2014; Chung *et al.* 2019). In this article, we explore the details of pairwise
47 compositional dissimilarity between two species assemblages (Jaccard dissimilarity) under the influence of
48 species differences and spatial heterogeneity. The logical starting point is akin to one of the most successful
49 null models in ecology, the Theory of Island Biogeography (TIB; MacArthur & Wilson 1963; MacArthur &
50 Wilson 1967), which is the pioneering work on the stochastic treatment of species richness and composition.
51 Following recent work on TIB (Alonso *et al.* 2015; Ontiveros *et al.* 2021), our work assumes that species
52 incidences (presence or absence) are stochastic variables that are independent from each other both
53 within and between sites, which nullifies any correlation between them, the so-called *species independence*
54 *assumption*. We thereby examine the effect of species differences and spatial heterogeneity on the expected
55 value of pairwise dissimilarity of species composition (Jaccard 1908, 1912; Veech 2012; Arita 2017; Keil *et al.*
56 2021). This allows for the consideration of various realistic factors that drive differences in local biotas,
57 such as the distance to the mainland, and specific spatial niche partitioning.

58 to analyze the effects of species differences on beta diversity, we obtain the exact and approximated
59 expression for the expectation and variance of Jaccard dissimilarity under the species independence
60 assumption. Second, we examine how the expectation and variance in beta-diversity respond to species
61 differences and spatial heterogeneity. Specifically, we numerically generate species presence probabilities for
62 a pair of sites, and assess how varying species-differences and spatial heterogeneity influence beta-diversity
63 patterns. We find that species differences can have opposing effects on beta-diversity. Counterintuitively,
64 we find that species differences result in lower beta-diversity when site heterogeneity is scarce. While
65 traditional analyses of beta-diversity focus on sites censused completely, many contemporary datasets are
66 based on statistical models such as Species Distribution Models (SDMs; Elith & Leathwick 2009; Guisan *et al.*
67 2017; Zurell *et al.* 2020). Indeed, recent work has attempted to improve the statistical power of local species
68 richness estimation in SDMs under heterogeneity (Calabrese *et al.* 2013). We therefore examine how our
69 approach can be applied to such statistical models. The temporal Jaccard dissimilarity (Legendre 2019;
70 Magurran *et al.* 2019; Figure 1) is designed to project how the local species composition across regions is
71 expected to change with time, particularly in response to global environmental changes. We combined our
72 method with SDMs and assessed the expected changes in the compositions of woodpecker species across
73 Switzerland. We believe that the approach be a starting point to provide further insights to the extent and
74 intensity of future compositional change, and to help us allocate resources for tasks such as monitoring,
75 conservation, or restoration.

Table 1: Summary of notation used in the main text.

Notation	Definition	Note
i	Species label	$i = 1, 2, \dots$, or S
S	The total number of species in the mainland	“species pool size”
j	Site label, with $j = 1$ or 2	“sites” may be spatial or temporal
$x_{i,j}$	Incidence of species i in site j	0 (absence) or 1 (presence)
$\mathbf{X}_{(S,2)}$	Incidence table of size S -by-2	Abbreviated to \mathbf{X}
$:=$	Defining a quantity	
\equiv	Identity	usually with respect to “all i ”
Ω	Set of incidences (absence 0 and presence 1)	$:= \{0, 1\}$
$\mathbf{x}_{\circ,j}$	Column vector of configuration in site j	
$ \mathbf{x}_{\circ,j} $	The number of species present in site j	$= \sum_{i=1}^S x_{i,j}$
$\mathbf{x}_{i,\circ}$	Row vector of species i 's incidence in space	
$p_{i,j}$	Probability of i present in j (i.e., $\text{Prob}[x_{i,j} = 1]$)	$a_{i,j} = 1 - p_{i,j}$ for probability of absence
$b_{i,\circ}$	Probability of i present in both sites 1 and 2	$\text{Prob}[x_{i,1} = x_{i,2} = 1]$; b for “both”
$d_{i,\circ}$	Probability of i absent from both sites 1 and 2	$\text{Prob}[x_{i,1} = x_{i,2} = 0]$; d for “double-absence”
$P_{\mathbf{X}}$	Probability that a table \mathbf{X} is observed	$= \prod_{i=1}^S \prod_{j=1}^2 p_{i,j}^{x_{i,j}} a_{i,j}^{1-x_{i,j}}$
$\beta_{\mathbf{X}}^J$	Jaccard dissimilarity for an incidence table \mathbf{X}	
$\gamma_{\mathbf{X}}$	The total number of species present in the landscape for table \mathbf{X}	“Gamma-diversity”
$\mathbf{E}[\beta^J]$	Expectation of Jaccard dissimilarity	$\mathbf{E}[\beta^J \gamma > 0]$ for conditional expectation
$\mathbf{V}[\beta^J]$	Variance of Jaccard dissimilarity	$\mathbf{Std}[\beta^J] := \sqrt{\mathbf{V}[\beta^J]}$
β_{heur}^J	Approximation of $\mathbf{E}[\beta^J \gamma > 0]$	“Heuristic approximation”
w	Species difference	$0 \leq w < 2$; Eqn (5)
h	Spatial heterogeneity	$0 \leq h \leq 1$; Eqn (6)
μ_j	Average presence probability in site j	

76 2 Methods and Results

77 Model

78 Jaccard dissimilarity is defined as the proportion of the number of ‘unique species’ (by which in this
79 literature we mean the number of species that are present in only one of the sites) to that of species present
80 in at least one of the sites (Jaccard 1908, 1912), which measures a compositional difference between two sites
81 (Box 1). Note that the two sites are either a pair of spatially segregated sites (at the same time-point) or pair of
82 temporally segregated sites (the same local place), the latter approach known as the ‘temporal beta-diversity’
83 (Legendre 2019; Magurran *et al.* 2019; Figure 1B). As such, we consider a landscape (metacommunity)
84 consisting of either two spatially or temporally segregated sites (Figure 1B), but we simply say “two sites” in
85 the present manuscript. We consider a matrix of binary elements, in which species are listed from top to
86 bottom (labelled $i = 1, 2, \dots, S$), and sites are listed from left to right (from $j = 1$ to 2). That is, an incidence
87 table of S species with 2 sites is written as \mathbf{X} . Following Chase *et al.*’s (2011) definition, we define “species
88 pool” as the collection of species that can possibly inhabit either site within a reasonable time period during
89 which the incidence data census is performed, and therefore defined S as the maximum number of species
90 that may be present in at least one of the sites (with a positive probability; we refer to S as “species pool
91 size”). We do not consider a species that can never be present. The formal approach to derive Jaccard
92 dissimilarity and their moments is encapsulated in Box 1.

93 We write x_{ij} for (i, j) -th element of an incidence table \mathbf{X} , with $x_{ij} = 1$ if species i is present in j , and
94 $x_{ij} = 0$ otherwise absent. We write p_{ij} for the probability that species i is present in site j (and this
95 probabilistic approach for spatial species-incidence is the part of TIB; MacArthur & Wilson 1963; MacArthur
96 & Wilson 1967; Real *et al.* 2016; Carmona & Pärtel 2020). Statistically, the sum $\sum_{i=1}^S x_{ij}$, which represents
97 the local species richness (alpha-diversity), follows the Poisson-Binomial distribution with parameters
98 $(p_{1j}, p_{2j}, \dots, p_{Sj})$. The key assumption is that p_{ij} may take different values depending on species i and
99 site j but are independent of each other (the species independence assumption). There are numerous
100 mechanisms by which presence probabilities are site- or species-specific. For example, differences in
101 colonization ability, tolerance to disturbance, ecological niches, and environmental filtering can jointly
102 influence presence probabilities.

103 The sum $\sum_{i=1}^S x_{ij}$, which represents the species richness (alpha-diversity), follows the Poisson-Binomial
104 distribution with parameters $(p_{1j}, p_{2j}, \dots, p_{Sj})$. For brevity, we write: (i) $a_{ij} := 1 - p_{ij}$ for the probability
105 that species i is absent in j ; (ii) $b_{i,o} := p_{i,1}p_{i,2}$ for the probability that species i is present in both sites 1 and 2
106 (probability of “double-presence”); (iii) $d_{i,o} := a_{i,1}a_{i,2}$ for the probability that species i is absent in both sites
107 1 and 2 (probability of “double-absence”). As a result, $1 - b_{i,o} - d_{i,o}$ represents the probability that species i
108 is unique to either site 1 or 2. Note that we do not distinguish in which sites (1 or 2) the unique species
109 are present. For example, colonization ability, stress-tolerance, distance from continents, environmental
110 filtering, and disturbance frequency can in concert influence presence probabilities of species in each site.
111 Therefore, we we generally assume that the presence probabilities of the species in each site are generally
112 different.

113 To analyze the probability distribution of beta-diversity, we present a novel derivation using tools
114 from stochastic analyses and combinatorics. These tools allow us to exactly compute the expectation and
115 variance of Jaccard dissimilarity based on the species-independence assumption. Numerical computation
116 can take a long time, because of combinatorial calculations associated with species-specific and site-specific
117 presence probabilities $(p_{1,1}, p_{1,2}, p_{2,1}, \dots, p_{S,1}, p_{S,2})$. To make the formula accessible to as broad researchers as
118 possible, we present a fast computable approximation (below).

138

Box 1: Derivation of the expectation of Jaccard dissimilarity

We write $\mathbf{X} \in \Omega^S \otimes \Omega^2$ (with \otimes for a direct product between sets) to indicate that the incidence table \mathbf{X} is a matrix with S rows and 2 columns, each of whose elements is either 0 for absence or 1 for presence. We also write $\mathbf{x}_{o,j} = (x_{1,j}, x_{2,j}, \dots, x_{S,j})^\top$ for a column vector within the incidence matrix (\mathbf{X}) in a site j , with \top for transpose. Therefore, $\mathbf{x}_{o,j} \in \Omega^S$. We write $|\mathbf{x}_{o,j}| := \sum_{i=1}^S x_{i,j}$ for alpha-diversity in a site j . In addition, we write $\langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle := \sum_{i=1}^S x_{i,1}x_{i,2}$ for the number of species present in both sites 1 and 2. Note that the number of species present in the landscape (gamma-diversity) is given by $|\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}| - \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle$.

Jaccard dissimilarity index (Jaccard 1908, 1912) for a given table \mathbf{X} , denoted by $\beta_{\mathbf{X}}^J$, is defined as:

$$\beta_{\mathbf{X}}^J = \frac{(\# \text{ unique species})}{(\# \text{ species present in at least one of the sites})} = \frac{(|\mathbf{x}_{o,1}| - \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle) + (|\mathbf{x}_{o,2}| - \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle)}{|\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}| - \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle}, \quad (1)$$

which represents the proportion of the number of unique species (present only one of the two sites) to the number of present species (present at least one of the two sites). Eqn (1) is written as a function of stochastic variables, $x_{i,j}$.

We aim to evaluate the first and second moments (expectation and variance) of Jaccard dissimilarity and assess how the moments vary with parameters of species differences and spatial heterogeneity. If the expectation of Jaccard dissimilarity varies in a certain way with a varying parameter, then the Jaccard dissimilarity is expected to behave accordingly and thereby forms a “beta-diversity pattern.”

Box 2: Expectations: exact

We denote the expectation, conditioned on at least one species being present, of Jaccard dissimilarity by $\mathbf{E}[\beta^J | \gamma > 0]$. We found that (Appendix A):

$$\mathbf{E}[\beta^J | \gamma > 0] = \int_0^1 \sum_{i=1}^S \frac{1 - b_{i,o} - d_{i,o}}{1 - \prod_{k=1}^S d_{k,o}} \prod_{\ell=1, \ell \neq i}^S (1 - z + zd_{\ell,o}) dz = \frac{\sum_{i=1}^S \sum_{\mathbf{y} \in \Omega^S} \frac{1}{|\mathbf{y}|} \frac{y_i(1 - b_{i,o} - d_{i,o})}{s C_{|\mathbf{y}|}} \prod_{\ell(\neq i)}^S d_{\ell,o}^{1 - y_\ell}}{1 - \prod_{k=1}^S d_{k,o}} \quad (2)$$

where $\mathbf{y} = (y_1, y_2, \dots, y_S) (\in \Omega^S)$ represents a binary sequence with length S and $s C_{|\mathbf{y}|}$ represents the binomial coefficient, which counts the number of ways, disregarding order, that $|\mathbf{y}| = \sum_{i=1}^S y_i (= 0, 1, \dots, S)$ species can be chosen from S species (Van Lint & Wilson 2001). Eqn (2) is conditioned on the premise that at least one of the species is present in the landscape (which occurs with probability $1 - \prod_{i=1}^S d_{i,o}$). Note that $\mathbf{E}[\beta^J | \gamma > 0]$ has a symmetry in site 1 and 2, by which swapping (permutating) $p_{i,1}$ and $p_{i,2}$ for any species i does not change the result (Baselga 2010).

As is detailed in Appendix A, Eqn (2) recovers the formula for species-equivalence cases (Chase *et al.* 2011; Lu *et al.* 2019). As suggested in previous studies, the species pool size S , which represents the maximum number of species present in the landscape, has no effects on the expected Jaccard dissimilarity conditioned of at least one of the species being present. The invariance is because species being equivalent implies that the incidence of any species occurs equally likely and independently, regardless of how many species can potentially inhabit the landscape.

158 **Approximations for expectation and variance**

159 We here present a “heuristic” approximation for the expectation of Jaccard dissimilarity, which was used in
 160 the previous studies (Kalyuzhny *et al.* 2021; Ontiveros *et al.* 2021). We write β_{heur}^J for the expectation of the
 161 numerator divided by the expectation of the denominator; that is:

$$\begin{aligned} \mathbf{E}\left[\beta^J \mid \gamma > 0\right] &\approx \beta_{\text{heur}}^J = \frac{\mathbf{E}[\# \text{ unique species}]}{\mathbf{E}[\# \text{ species present at least one of the sites}]} \\ &= \frac{\sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ})}{\sum_{i=1}^S (1 - d_{i,\circ})} \end{aligned} \quad (3)$$

162 (the derivation and interpretation are provided in Appendix B; c.f., Lynch & Walsh 1998). Mathematically,
 163 the heuristic approximation obtains by exchanging the order of taking expectation and computing fraction
 164 while guaranteeing that there is at least one species present. In the example below, we will show that
 165 Eqn (3) provides a near-identical approximation for the conditional expectation $\mathbf{E}\left[\beta^J \mid \gamma > 0\right]$; the result of
 166 stochastic simulations to verify the accuracy of the heuristic approximation is given in Appendix B.

167 We also derive exact and approximated formulae of the variance of Jaccard dissimilarity, but both
 168 involve much complication in general. Therefore, we show the results on variance for species-equivalence
 169 case and encapsulated the detailed analyses for general cases with species differences in the Appendix D.

170 **Application 1: Theoretical analysis**

171 **Special case: identical species**

172 We demonstrate a special case where the species are identical in any sense but the landscape can be
 173 heterogeneous; that is: $p_{i,j} \equiv p_{\circ,j}$ for each $j = 1, 2$. We consequently write $b_{\circ,j} \equiv b$ and $d_{\circ,j} \equiv d$ for double-
 174 presence and double-absence (respectively; note that $p_{\circ,1}$ and $p_{\circ,2}$ may differ). Substituting those into the
 175 expected Jaccard dissimilarity index (Eqn (2)) yields:

$$\mathbf{E}\left[\beta^J \mid \gamma > 0\right] \Bigg|_{\text{species equivalence}} = \frac{1 - d - b}{1 - d} = \beta_{\text{heur}}^J. \quad (4)$$

176 That is, the heuristic approximation and the exact conditional expectation completely agree with each
 177 other. Eqn (4) also recovers the results of previous studies that assume species’ presence probabilities are
 178 identical (Chung *et al.* 2019; Lu *et al.* 2019; Lu 2021; Ontiveros *et al.* 2021). In other words, the sufficient
 179 condition for the exact and approximated formulae of the expectation to be equal is that species are equal
 180 in presence probabilities.

181 We numerically investigate the dependence of the expected Jaccard dissimilarity on the presence
 182 probabilities in sites 1 and 2 (Figure 2A). We found that the expected Jaccard dissimilarity becomes larger
 183 as the presence probabilities decrease. More specifically, the expected Jaccard dissimilarity increases with
 184 increasing probability of double-presence (i.e., when the presence probabilities of both species become
 185 large).

186 We also investigate the dependence of the variance of Jaccard dissimilarity on the pair of presence
 187 probabilities $(p_{\circ,1}, p_{\circ,2})$ (Figure 2B). We find that: (i) when $p_{\circ,1}$ and $p_{\circ,2}$ are small, the variance tends to take
 188 large values; (ii) when the presence probability in one site is large but the other is small, the variance of
 189 Jaccard dissimilarity is small but the change in the variance with increasing the difference between $p_{\circ,1}$
 190 and $p_{\circ,2}$ appears to be very small (< 0.05 ; Figure 2B); (iii) when both values are large, the variance tends to
 191 be small; and (iv) when both probabilities are at intermediate values, the dependence of the variance on
 192 presence probabilities appears to be small.

193 Effects of species difference and spatial heterogeneity

194 We examine how species differences and spatial heterogeneity jointly influence the expectation of Jaccard
195 dissimilarity, by means of generating series of synthetic presence probabilities. For incidence probabilities
196 (p_{1j}, \dots, p_{Sj}) with $j = 1, 2$, we first define species difference w and spatial heterogeneity h by:

$$w := \frac{1}{\mu_1 + \mu_2} \sum_{j=1}^2 \frac{1}{S} \sum_{i=1}^S |p_{ij} - \mu_j| \quad (5)$$

$$h := \frac{1}{S} \sum_{i=1}^S |p_{i,1} - p_{i,2}|, \quad (6)$$

197 where $\mu_j := \sum_{i=1}^S p_{ij} / S$ represents the average of presence probabilities in site j (for $j = 1, 2$). The first
198 parameter w represents the degree to which species presence probabilities (plotted against species labels) are
199 not equal (uneven; see Appendix C for the intuitive interpretation). Also, h represents the degree to which
200 two sites differ in the presence probabilities, averaged per species. Biologically, the spatial heterogeneity h
201 can reflect various processes; for example, if a given species is favored by the environment in site 1 but is
202 filtered out by site 2, then it is reasonable to say that spatial heterogeneity exists for that species 1. Similarly,
203 if site 1 is accessible for some species but not for other species, interspecific differences in dispersal ability
204 can also generate variation in presence probabilities and thereby increase species difference w . Note that
205 both quantities are functions of the presence probabilities. Importantly, these quantities can be empirically
206 estimated by calculating the presence probabilities p_{ij} by averaging long time-series data (assuming that
207 the community incidence is at equilibrium; MacArthur & Wilson 1963; MacArthur & Wilson 1967) and
208 then inserting those values in to Eqns (5) and (6)

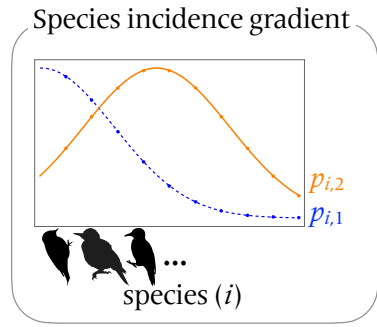
209 We assess how varying species difference and spatial heterogeneity jointly alter $\mathbf{E}[\beta^J]$. We first consider a
210 case with $p_{i,1} = p_{i,2}$ for all species, that is, the case where two sites are homogeneous. In this case, increasing
211 the species difference w reduces beta diversity (Figure 3A-1). This is the case also in the presence of a slight
212 difference between two sites (Figure 3A-2). Hence, when there is at most low spatial heterogeneity, species
213 differences systematically reduce beta-diversity.

214 As the degree of site heterogeneity h becomes larger, whether species differences result in larger or
215 smaller beta-diversity becomes less consistent. We found both cases where species differences increase
216 and decrease beta-diversity (Figure 3B-1, B-2). The difference between Figure 3B-1 and B-2 is that while
217 Figure 3B-1 assumes that two sites are heterogeneous evenly among species ($|p_{i,1} - p_{i,2}|$ all equal), Figure 3B-
218 2 uses different values $|p_{i,1} - p_{i,2}|$ (although both are on average equal). That is, the effects of species
219 differences on beta-diversity depend on the patterns of species differences and spatial heterogeneity.

220 To examine the robustness of this prediction, we work on broader ranges of parameters (w and h) by
221 generating various values of $\{p_{ij}\}$ and plotting the expectation of Jaccard dissimilarity against species
222 difference with varying spatial heterogeneity. We first generate a dataset consisting of various shapes of the
223 presence-probability curves (Appendix C, Eqn C50) and corresponding species difference (w), spatial hetero-
224 geneity (h), the expected Jaccard dissimilarity, and the average presence probability for all species among
225 two sites ($\mu := (\mu_1 + \mu_2) / 2$; this notation is used only here). We then extract a subset of the dataset with
226 varying ranges. From this sub-dataset, depending on the range of μ (with $\mu \in [0.1, 0.2], [0.3, 0.4], [0.5, 0.6],$
227 or $[0.7, 0.8]$; we choose mutually exclusive intervals to avoid excess numbers of points to be considered),
228 we separately plot the dependence of the expected Jaccard dissimilarity on species difference.

229 We find that the resulting patterns are consistent with those in Figure 3: in spatially homogeneous
230 environments, increasing species difference tends to reduce beta-diversity, and this trend becomes reversed
231 or obscured as the spatial heterogeneity increases (Figure 4). Notably, the average presence probability can
232 constrain the range of the expected value of Jaccard dissimilarity (Figure 4), in accordance with Figure 2

(A) Varying presence probabilities



- calculate ↓
- Species difference w (eqn 5)
 - Site heterogeneity h (eqn 6)

(B) Compositional dissimilarity for ...

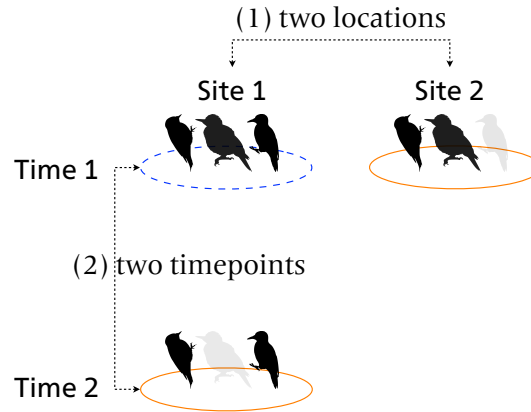
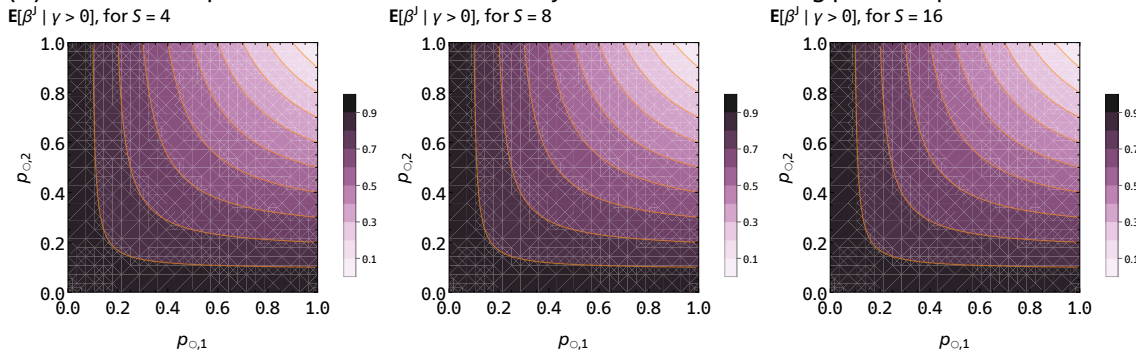


Figure 1: Schematic illustration of the framework. (A) The presence probabilities $(p_{1,1}, p_{1,2}, p_{2,1}, \dots, p_{S,2})$ take different values. Of our interest are the effects of the differences (measured by w) and site heterogeneity (measured by h) on compositional dissimilarity, both of which are computed from given $p_{1,1}, \dots, p_{S,1}, p_{1,2}, \dots, p_{S,2}$. (B) The compositional dissimilarity (defined by the Jaccard dissimilarity) can be applied to spatial and temporal changes. When j represents a spatial site, we compute the expected Jaccard dissimilarity from the compositions of the two site (for example 110 and 011). We can carry out the same calculation for time point labels $j = 1, 2$. Note the symbol \top for transpose. The woodpecker pictures are from <http://phylopic.org/> (CC0).

(A) Conditional expectation of Jaccard dissimilarity increases with reducing presence probabilities



(B) Conditional standard deviation of Jaccard dissimilarity increases with reducing presence probabilities

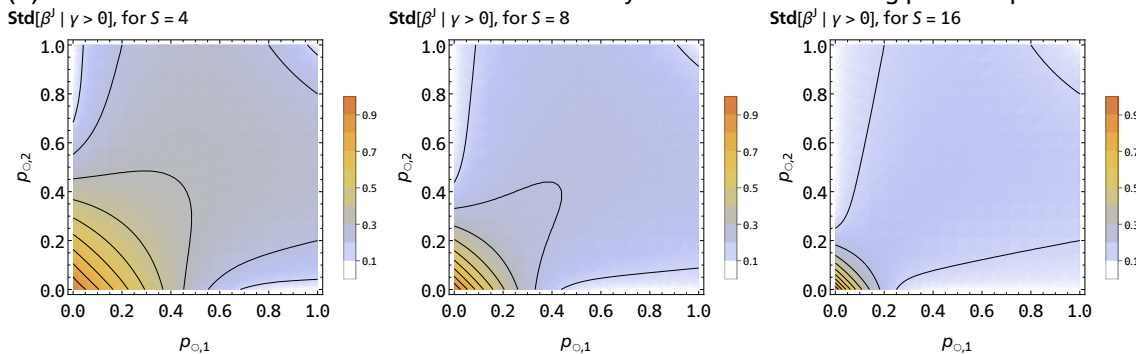
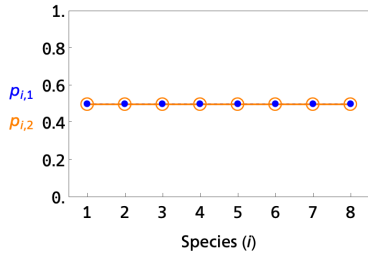


Figure 2: The conditional expectation of Jaccard dissimilarity $E[\beta^j | \gamma > 0]$ (panel A) and the conditional standard deviation of Jaccard dissimilarity $\text{Std}[\beta^j | \gamma > 0] := \sqrt{\text{V}[\beta^j | \gamma > 0]}$ (B), plotted on $p_{o,1}$ (horizontal axis), $p_{o,2}$ (vertical axis) for varying species pool sizes, S . (A) The expected Jaccard dissimilarity is lower when the probability of double-presence is larger (right top zones), and higher when the probability of being present in only one of the sites is higher (left top and right bottom zones). The effect of S on the expectation is negligibly small, and thus increasing S (from left to right panels) does not appear to change the overall trend. (B) The standard deviation of Jaccard dissimilarity is lower when the presence probabilities take extreme values (four corners). As S increases, the standard deviation tends to be less dependent on intermediate values of presence probabilities, as seen from the observation that over the wide region of the rectangle, the standard deviation is relatively unchanged with presence probabilities.

(A-1) Sites are homogeneous \Rightarrow beta decreases with increasing species difference w

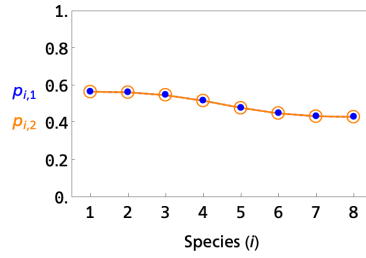
$w = 0, h = 0.$



Exact ($E[\beta^j | \gamma > 0]$) = 0.666667

Heuristic approx (β_{heur}^j) = 0.666667

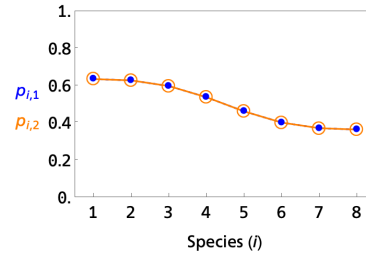
$w = 0.1, h = 0.$



Exact ($E[\beta^j | \gamma > 0]$) = 0.661062

Heuristic approx (β_{heur}^j) = 0.661545

$w = 0.2, h = 0.$

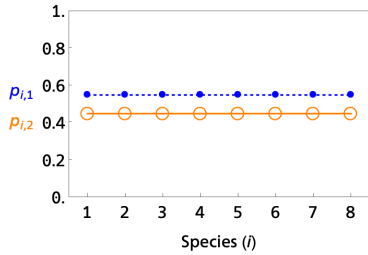


Exact ($E[\beta^j | \gamma > 0]$) = 0.644037

Heuristic approx (β_{heur}^j) = 0.645942

(A-2) Sites are weakly heterogeneous \Rightarrow beta decreases with increasing w

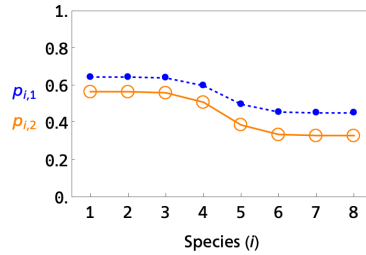
$w = 0, h = 0.1$



Exact ($E[\beta^j | \gamma > 0]$) = 0.671096

Heuristic approx (β_{heur}^j) = 0.671096

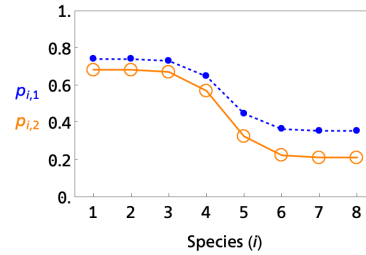
$w = 0.1, h = 0.1$



Exact ($E[\beta^j | \gamma > 0]$) = 0.653477

Heuristic approx (β_{heur}^j) = 0.655013

$w = 0.2, h = 0.1$

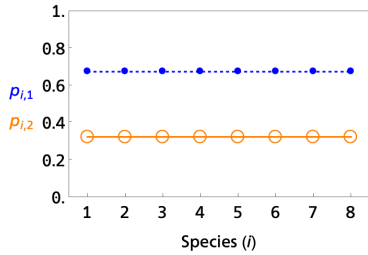


Exact ($E[\beta^j | \gamma > 0]$) = 0.59844

Heuristic approx (β_{heur}^j) = 0.604339

(B-1) Sites are strongly heterogeneous \Rightarrow beta may decrease with increasing w

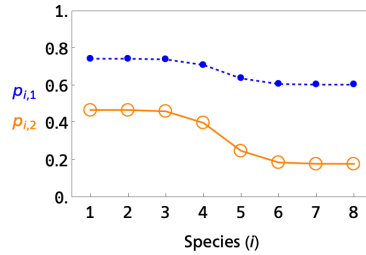
$w = 0, h = 0.35$



Exact ($E[\beta^j | \gamma > 0]$) = 0.718975

Heuristic approx (β_{heur}^j) = 0.718975

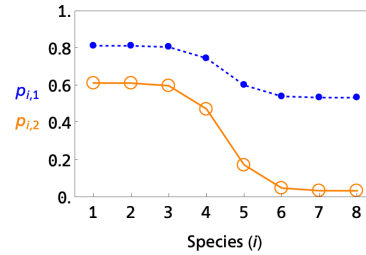
$w = 0.1, h = 0.35$



Exact ($E[\beta^j | \gamma > 0]$) = 0.704116

Heuristic approx (β_{heur}^j) = 0.705752

$w = 0.2, h = 0.35$

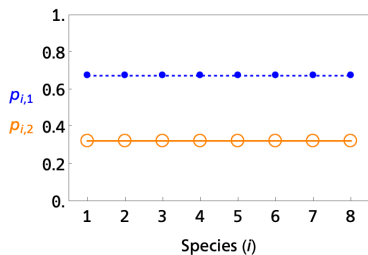


Exact ($E[\beta^j | \gamma > 0]$) = 0.657895

Heuristic approx (β_{heur}^j) = 0.66439

(B-2) Sites are strongly heterogeneous \Rightarrow beta may increase with increasing w

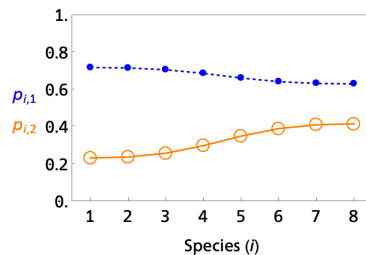
$w = 0, h = 0.35$



Exact ($E[\beta^j | \gamma > 0]$) = 0.718975

Heuristic approx (β_{heur}^j) = 0.718975

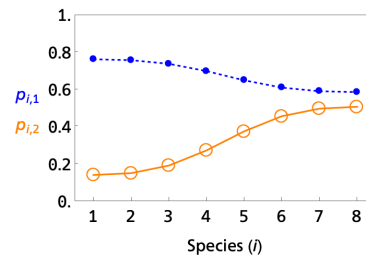
$w = 0.1, h = 0.35$



Exact ($E[\beta^j | \gamma > 0]$) = 0.723095

Heuristic approx (β_{heur}^j) = 0.723094

$w = 0.2, h = 0.35$



Exact ($E[\beta^j | \gamma > 0]$) = 0.735304

Heuristic approx (β_{heur}^j) = 0.735295

Figure 3: The beta-diversity patterns under species differences and spatial heterogeneity. Setup: blue dashed curves plot $p_{i,1}$ and orange $p_{i,2}$ respectively (referred to as species incidence curves). In all panels, the average presence probability per species per site is set one half: $(\mu_1 + \mu_2) / 2 = 0.5$. Species differences and site-heterogeneity are calculated from the incidence curves, each indicated. (A-1) When sites are completely homogeneous (i.e., when $p_{i,1} = p_{i,2}$ for all species $i = 1, \dots, S$), increasing species difference results in lower beta-diversity. (A-2) When sites are weakly heterogeneous $h = 0.05$, species difference reduces beta-diversity, as in (A). (B) When two sites are strongly heterogeneous with $h = 0.40$ in this example, increasing species differences can either decrease (B-1) or increase (B-2) beta-diversity. The data points $(p_{1,1}, p_{1,2}, \dots, p_{8,2})$ and corresponding indices (the conditional expectation and approximation) are encapsulated in Appendix C.

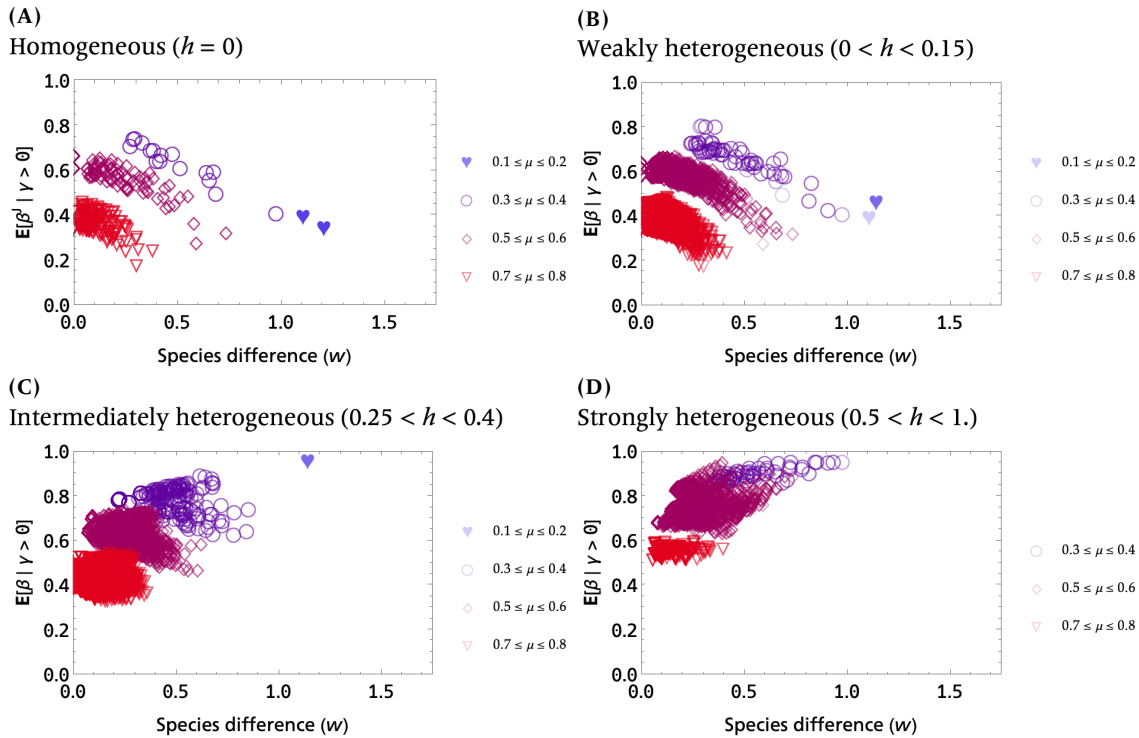


Figure 4: Dependence of the conditional expectation on species difference with varying spatial heterogeneity h (four panels) and varying mean presence probability (plotted in blue, purple, dark purple, and red), with $\mu = (\mu_1 + \mu_2) / 2$. When two sites are completely homogeneous, increasing species difference tends to decrease the expected Jaccard dissimilarity (panel A). When the degree of spatial heterogeneity is small, this trend is robust (panel B). As spatial heterogeneity increases, the dependence of beta-diversity on species difference becomes obscured, or even reversed (panels C and D).

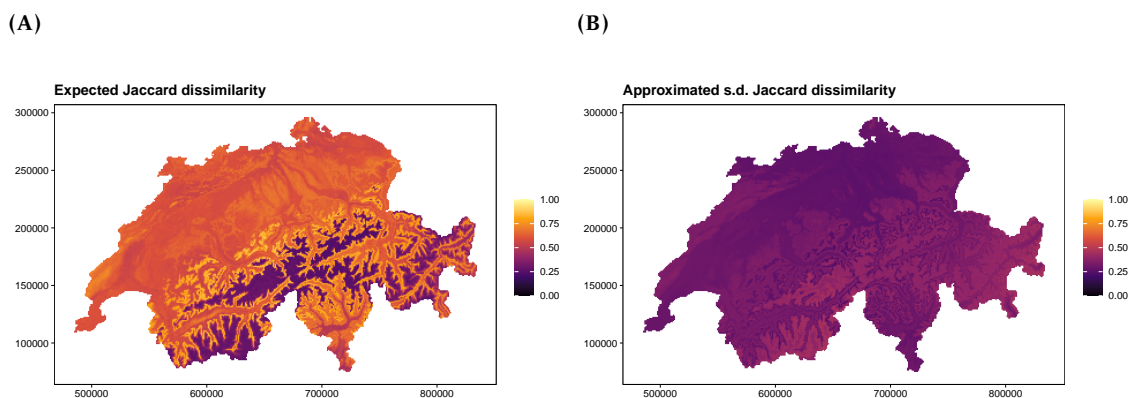


Figure 5: Outcomes of species distribution models, using a climatic variable. We quantified the expected, compositional dissimilarity of five woodpecker species at two time points, current and future, over the region of Switzerland. That is, we assessed the expectation and standard deviation of temporal Jaccard dissimilarity (Legendre 2019; Magurran *et al.* 2019). We used occupancy estimations for current and future climatic conditions over Switzerland. (A) Expectation. Compositional changes are expected to be high in the upper limit of the current distribution and lowlands. (B) Standard deviation (approximated). The standard deviation tends to be small, which is consistent with the analytical prediction in Figure 2.

233 Application 2: Temporal Jaccard dissimilarity with Species Distribution Models

234 We provide a case application of our method using the Species Distribution Models (SDMs; Elith & Leathwick
235 2009; Guisan *et al.* 2017; Zurell *et al.* 2020, with the details described in Appendix E). Generally, SDMs seek
236 to estimate the probability that each species is present at a given site given information on the environment
237 found at that site. Our approach enables quantification of the changes in species distribution on a location-
238 by-location basis, based on temporal Jaccard dissimilarity (the proportion of the number of species that are
239 present only one of two time points relative to the number of species that are present in at least one of the
240 two time points; Legendre 2019; Magurran *et al.* 2019, Figure 1). Note that the temporal data systematically
241 comes with correlations between compositions at two time points ('temporal autocorrelation'), but our
242 analysis takes advantage of the site-permutation symmetry in Jaccard dissimilarity so that we can omit the
243 correlations between two compositions.

244 Based on Schmid *et al.* (1998, 2018) and Zurell *et al.* (2019b, 2020), we use data of five woodpecker
245 species, *Picus viridis*, *P. canus*, *Dendrocopos major*, *D. minor*, and *Dryocopus martius* in Switzerland. These
246 species have common evolutionary history but use different habitats (Benz *et al.* 2006; Pasinelli 2007;
247 Pons *et al.* 2010). For example, *P. canus* and *D. minor* occur at lowlands, while *P. viridis* is more widely
248 found across Switzerland (Appendix E). The variation in geographic habitat use arguably reflects species
249 differences, making the system ideal for an application of temporal Jaccard dissimilarity. Note also that
250 incidence-based SDMs assume that species-independence is uncorrelated, and therefore are reconciled
251 with our species-independence assumption.

252 We examine how dissimilar woodpeckers' metacommunity will be under future climatic conditions
253 (Representative Concentration Pathway 4.5) across Switzerland in 2050, compared with the sub-community
254 under the current climate, at a scale of 1×1 [km]. Our methods predict that the expectations of temporal
255 Jaccard dissimilarity are unexpectedly high (Figure 5), indicating significant future changes in the distribu-
256 tion of the woodpeckers. This result is primarily explained by species dynamics in lowland sites where
257 some species thrived and others failed: *P. canus*, that will decrease its occupancy rate near the rivers and
258 will increase in surrounding areas (SI Figure 4), and *D. minor*, whose occupancy is expected to increase in
259 lowlands and valleys (SI Figure 4). Dissimilarity in hillsides is expected to be moderate due to a general
260 increment in richness (SI Figure 4). These results are consistent with a general trend of Switzerland forest
261 birds moving to higher grounds as a response to environmental change (Maggini *et al.* 2014). Our results
262 demonstrate the dramatic effects of climate change on temporal beta-diversity of a bird metacommunity.

263 3 Discussion

264 We derived the formulae of the expectation and variance of Jaccard dissimilarity index as a measure of
265 compositional dissimilarity between two sites for nonidentical species in heterogeneous landscapes. We
266 showed that species differences may have nonmonotonic effects on Jaccard dissimilarity (Figure 3). When
267 sites are relatively similar in their species presence probabilities, species difference reduces Jaccard dissimi-
268 larity. When sites are different in their presence probabilities, species differences can cause varying effects
269 on Jaccard dissimilarity, making robust prediction challenging. Our work allows us to implement empirical
270 incidence data of presence probabilities and assess the long term estimations for Jaccard dissimilarity.
271 In addition, our method enables quantitative comparison of predicted outcomes with observed species
272 distributions. Hence, the present work offers a powerful framework to theoretically and quantitatively
273 investigate spatial variations in species composition.

274 Our results suggest that knowledge of the species pool will be important for some properties but not
275 others. When all species are identical, we found that the conditional expectation of Jaccard dissimilarity
276 (given that at least one species is likely present) is independent of species pool size, as was revealed in the
277 literature (Lu *et al.* 2019; Kalyuzhny *et al.* 2021; Lu 2021). In contrast, we found that the variance more
278 substantially depends on the species pool size even when the species presence probabilities are equal
279 in each pair of sites (Figure 2B). Specifically, the variance becomes smaller when the species pool size

280 is larger, suggesting that large species pool sizes may reduce the uncertainty in statistical inferences for
281 Jaccard dissimilarity. However, the uncertainty may become larger when the baseline presence probabilities
282 are smaller (Figure 2B), suggesting that the presence of opportunistic species (species with the presence
283 probabilities are very low but not zero) may influence the robust inference of Jaccard dissimilarity, and also
284 that rarely observed species can increase the uncertainty of estimations for Jaccard dissimilarity (Wolda
285 1981; Colwell & Coddington 1994; Plotkin & Muller-Landau 2002; Chao *et al.* 2004). These predictions
286 are consistent with the importance of species that are currently present but can be potentially present
287 (dark diversity; Pärtel *et al.* 2011; Carmona & Pärtel 2020), and our work suggests that dealing with the
288 uncertainty is the key to predict the dark diversity.

289 Comparison with simulated data showed that the approximation of the Jaccard dissimilarity agrees
290 well with the exact formula, with its accuracy higher with the increasing species pool size (Appendix
291 B). Notably, the heuristic approximation is a closed form with respect to the species-wise average of the
292 presence probabilities: both the numerator and denominator can be written as functions of the averages of
293 double-presence and double-absence probabilities, the results suggest that within-site species difference is
294 likely unimportant. Because the heuristic approximation is computationally much cheaper than calculating
295 the exact expectation of Jaccard dissimilarity in a brute force approach, we suggest that, when the species
296 pool size is relatively small, exact evaluation of the Jaccard dissimilarity is suitable and should be compared
297 with the approximated value which can be computed faster. When the species pool size is large, we suggest
298 using our approximations and exact values both, e.g., by applying Gauß' fast Fourier Transforms (Cooley
299 & Tukey 1965; Heideman *et al.* 1984). When the species pool size is unknown, the Eqn (3) is beneficial,
300 because the interpretation is clear and the expectations in the numerator and denominator can be assessed
301 separately.

302 Using data on woodpecker distributions, we projected shifts in habitat use potentially helping practi-
303 tioners to identify locations of interest for conservation and management (Nieto-Lugilde *et al.* 2017). This
304 approach required only information on the expectation and variance. Likely mechanisms of the species
305 differences and temporal heterogeneity in this system include colonization abilities, habitat selection, and
306 species-specific tolerance to environmental challenges. By specifying possible mechanisms, we can obtain
307 deeper insight into the processes by which biotic homogenization occurs. For instance, partitioning beta-
308 diversity into colonization and extinction components can be of great use to better understand dynamic
309 processes of beta-diversity (Tatsumi *et al.* 2021). Our study can be used in conjunction with data such as
310 SDMs for better understanding and management of the spatio-temporal dynamics of biodiversity. In previ-
311 ous studies using SDMs, mapping of dissimilarity has been carried out in a spatial (Godsoe 2013; Godsoe &
312 Case 2014) or temporal (Ferrier *et al.* 2007; Fitzpatrick *et al.* 2013). In this vein, we provide an analytical
313 approach that avoids relying on simulations, consequently speeding up the estimation of dissimilarity
314 and its standard deviation, applicable to any method under the species independence assumption, e.g., in
315 dynamical metapopulation models. Therefore, we improve accessibility to dissimilarity analyses.

316 Our key finding is that beta-diversity will only decrease with species differences depending when
317 space is completely or weakly homogeneous (Figure 3A-1 and A-2, and Figure 4A and B). When space is
318 very heterogeneous, To facilitate biological interpretations for the prediction, we first list species in the
319 descending order of presence probability in site 1 (i.e., $p_{1,1} \geq p_{2,1} \geq \dots \geq p_{S,1}$). We then plotted the curve of
320 $p_{i,j}$ against $i = 1, 2, \dots, S$ and termed this curve as a "species-incidence curve" for site j ; we here interpret that
321 a higher presence probability of a species reflects a suitability for that species. Given these settings, under
322 no spatial heterogeneity (i.e., $p_{i,1} = p_{i,2}$ for all species $i = 1, \dots, S$, and thus when $h = 0$), the two species-
323 incidence curves synchronize perfectly. In this case, increasing species differences increases the probability
324 of some species being either (i) double-absent or (ii) present in both sites. In these cases, an increase in
325 species differences reduces beta-diversity. This is because an increase in species difference indirectly makes
326 the two incidence curves flat to gradually uneven (from left to right; Figure 3A-1) and thereby both sites
327 become either more suitable or unsuitable for all species; note, however, that this result assumes that
328 the average presence probability in each site remains constant. Biologically, spatial homogeneity tends
329 to constrain the variation in species-incidence curves in a way that species differences make both sites

330 more suitable or less suitable for all species. In other words, species differences under homogeneity are
331 unlikely to foster uniquely present species. This prediction is robust against small increases in spatial
332 heterogeneity (Figure 3A-2, Figure 4B). From the perspective of estimating beta-diversity, the results suggest
333 that using common, averaged presence probabilities for all species may result in systematic overestimation
334 of beta-diversity. In a weakly heterogeneous metacommunity, beta-diversity is predicted to decrease with
335 species differences due to the synchronization of species incidence curves between two sites.

336 With increasing spatial heterogeneity, however, the effect of species difference on beta-diversity becomes
337 obscured (Figure 3B-1, B-2, Figure 4C and D). This is because in the presence of spatial heterogeneity, species
338 incidence curves are allowed to become asynchronous. As the degree of spatial heterogeneity becomes large,
339 two species incidence curves are necessarily asynchronised (or form a “nestedness” or “turnover pattern”,
340 *sensu* Harrison *et al.* 1992, by which species incidences tend to be spatially segregated). Let us compare
341 Figure 3B-1, in which beta-diversity decreases with species difference, versus B-2 in which beta-diversity
342 otherwise increases. We can see that in panel B-1, the effect of increasing species difference on beta-diversity
343 is similar to that in panel A-1 and A-2 with the two incidence curves synchronizing. This is because, in
344 Figure 3B-1, as species difference increases, both sites become more suitable for species 1, 2, and 3 but less
345 suitable for species 4, 5, and 6. Therefore this pattern can be understood in the similar way as those in
346 panels A-1 and A-2. In contrast, panel B-2 exhibits the asynchronization pattern with increasing species
347 difference. In this case, increasing the species difference can make the nestedness pattern more pronounced;
348 for example, as species difference increases, site 1 becomes more suitable but site 2 less suitable for species
349 1. Critically, as species difference increases, site 1 (and site 2) becomes more (and less) suitable for species
350 1, 2, and 3 but less (and more, respectively) suitable for species 4, 5, and 6, leading to increased turnover of
351 species. As these interpretations are not explicitly clear from Figure 4, we suggest careful assessments of the
352 mechanisms of species incidence across space to better predict the dependence of beta-diversity on species
353 difference in heterogeneous environments. Hence, heterogeneous environments make the prediction and
354 management of beta-diversity even more challenging.

355 Our study has critical implications for conservation. Generally, beta-diversity is a key factor for ecosys-
356 tem functioning from local to global scales (Socolar *et al.* 2016; Mori *et al.* 2018). Local ecosystem functioning
357 may be driven by species’ functional dissimilarity like niches; Godoy *et al.* 2020). For example, Loiseau
358 *et al.* (2016) pointed out that conservation policy designed to protect taxonomic diversity cannot be fully
359 reconciled with functional diversity management. Indeed our finding predicts that with the environmental
360 homogeneity, increasing local species differences lead to lower beta-diversity (Figure 3A-1 and A-2, and
361 Figure 4A and B). Given that the species difference is associated with functional diversity (or transformed
362 quantity thereof; Palacio *et al.* 2022), the present prediction suggests that a conservation policy aiming to
363 maintain high beta-diversity be traded-off against the local, functional diversity. This trade-off becomes
364 more complicated when the spatial heterogeneity is strong, by which beta-diversity may or may not
365 decrease with species-differences, thereby making the prediction of beta-diversity rather difficult. One
366 promising approach is thus to identify species traits and environmental factors for maintaining a balance
367 (evenness) in the likeliness of species presence and absence, which may produce new interesting questions.
368 Moving forward, open questions include: how does incidence-based beta-diversity respond to changes in
369 functional diversity in colonization ability and extinction tolerance? How does functional diversity, in turn,
370 respond against the reduction in compositional dissimilarity (biotic homogenization)?

371 To conclude, we have derived the analytic formula of the expectation and variance of Jaccard dissim-
372 ilarity index incorporating different species in a pair of heterogeneous sites, and revealed the opposing
373 effects of species differences on beta-diversity. Assuming that species incidences are uncorrelated with
374 each other, we found that species differences in presence probabilities may lead to a complex pattern
375 (Figure 3). This work will help researchers better understand the probabilistic, or stochastic, nature of
376 Jaccard dissimilarity (Real & Vargas 1996). Future studies may explore the effects of species associations
377 on the probabilistic properties of Jaccard dissimilarity, and also carry out occupancy dynamics analyses,
378 beyond pairwise dissimilarity analyses (MacKenzie *et al.* 2018). One of the most promising approaches is a
379 process-based approach (Pilowsky *et al.* 2022), by which we can incorporate further complications that

380 influence beta-diversity. Our method can incorporate additional realities to track and manage the changes
381 in species distributions under global changes.

382 4 Acknowledgement

383 The authors thank two anonymous reviewers, Ryosuke Nakadai, Naoto Shinohara, and Akira Terui, for
384 helpful comments. We thank JSPS-KAKENHI (grant numbers 19K22457, 19K23768, and 20K15882 to RI, and
385 21K14880 to ST) for funding. The authors also thank Margarita Salas grant funded by the Spanish Ministry
386 of Universities and the “European Union - Next GenerationEU” to VJO, CRISIS (PGC2018-096577-B-I00) to
387 DA and JAC, UNIQUE (PID2021-127202NB-C21) to DA, and PRIORITY (PID2021-127202NB-C22) to JAC, all
388 funded by MCIN/AEI/10.13039/501100011033 and “ERDF A way of making Europe”. WG was supported
389 by the Te Pūnaha Matatini centre of research excellence. RI was inspired by some of the questions and
390 answers on Cross Validated (Stack Exchange), in calculating the expectation of reciprocals (c.f., Cressie *et al.*
391 1981; Lange 2010).

392 References

- 393 Albrecht, J., Peters, M. K., Becker, J. N., Behler, C., Classen, A., Ensslin, A., Ferger, S. W., Gebert, F., Gerschlauser,
394 F., Helbig-Bonitz, M., Kindeketa, W. J., Kühnel, A., Mayr, A. V., Njovu, H. K., Pabst, H., Pommer, U., Röder,
395 J., Rutten, G., Costa, D. S., Sierra-Cornejo, N., Vogeler, A., Vollstädt, M. G. R., Dulle, H. I., Eardley, C. D.,
396 Howell, K. M., Keller, A., Peters, R. S., Kakengi, V., Hemp, C., Zhang, J., Manning, P., Mueller, T., Bogner, C.,
397 Böhning-Gaese, K., Brandl, R., Hertel, D., Huwe, B., Kiese, R., Kleyer, M., Leuschner, C., Kuzyakov, Y., Naus,
398 T., Tschapka, M., Fischer, M., Hemp, A., Steffan-Dewenter, I., & Schleuning, M. (2021). Species richness is
399 more important for ecosystem functioning than species turnover along an elevational gradient. *Nature*
400 *Ecology & Evolution*, **5**.12, pp. 1582–1593. DOI: [10.1038/s41559-021-01550-9](https://doi.org/10.1038/s41559-021-01550-9) (cit. on p. 2).
- 401 Allouche, O., Kalyuzhny, M., Moreno-Rueda, G., Pizarro, M., & Kadmon, R. (2012). Area–heterogeneity
402 tradeoff and the diversity of ecological communities. *Proceedings of the National Academy of Sciences*,
403 **109**.43, pp. 17495–17500. DOI: [10.1073/pnas.1208652109](https://doi.org/10.1073/pnas.1208652109) (cit. on p. 2).
- 404 Alonso, D., Pinyol-Gallemlí, A., Alcoverro, T., & Arthur, R. (2015). Fish community reassembly after a coral
405 mass mortality: higher trophic groups are subject to increased rates of extinction. *Ecology Letters*, **18**.5,
406 pp. 451–461. DOI: [10.1111/ele.12426](https://doi.org/10.1111/ele.12426) (cit. on p. 3).
- 407 Altermatt, F., Schreiber, S., & Holyoak, M. (2011). Interactive effects of disturbance and dispersal directionality
408 on species richness and composition in metacommunities. *Ecology*, **92**.4, pp. 859–870. DOI: [10.1890/10-1095.1](https://doi.org/10.1890/10-1095.1) (cit. on p. 2).
- 410 Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell,
411 H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C., & Swenson, N. G. (2010).
412 Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*,
413 **14**.1, pp. 19–28. DOI: [10.1111/j.1461-0248.2010.01552.x](https://doi.org/10.1111/j.1461-0248.2010.01552.x) (cit. on p. 2).
- 414 Arita, H. T. (2017). Multisite and multispecies measures of overlap, co-occurrence, and co-diversity. *Ecogra-*
415 *phy*, **40**.6, pp. 709–718. DOI: [10.1111/ecog.01942](https://doi.org/10.1111/ecog.01942) (cit. on p. 3).
- 416 Arnold, B. C. (2012). *Majorization and the Lorenz order: A brief introduction*. **43**. Springer Science & Business
417 Media (cit. on p. 33).
- 418 Astorga, A., Death, R., Death, F., Paavola, R., Chakraborty, M., & Muotka, T. (2014). Habitat heterogeneity
419 drives the geographical distribution of beta diversity: the case of New Zealand stream invertebrates.
420 *Ecology and Evolution*, **4**.13, pp. 2693–2702. DOI: [10.1002/ece3.1124](https://doi.org/10.1002/ece3.1124) (cit. on p. 2).
- 421 Báldi, A. (2008). Habitat heterogeneity overrides the species–area relationship. *Journal of Biogeography*, **35**.4,
422 pp. 675–681. DOI: [10.1111/j.1365-2699.2007.01825.x](https://doi.org/10.1111/j.1365-2699.2007.01825.x) (cit. on p. 2).

- 423 Bar-Massada, A. (2015a). Complex relationships between species niches and environmental heterogeneity
424 affect species co-occurrence patterns in modelled and real communities. *Proceedings of the Royal Society*
425 *B: Biological Sciences*, **282**.1813, p. 20150927. DOI: [10.1098/rspb.2015.0927](https://doi.org/10.1098/rspb.2015.0927) (cit. on p. 3).
- 426 — (2015b). Immigration rates and species niche characteristics affect the relationship between species
427 richness and habitat heterogeneity in modeled meta-communities. *PeerJ*, **3**, e832. DOI: [10.7717/peerj.](https://doi.org/10.7717/peerj.832)
428 **832** (cit. on p. 3).
- 429 Bar-Massada, A. & Belmaker, J. (2017). Non-stationarity in the co-occurrence patterns of species across
430 environmental gradients. *Journal of Ecology*, **105**.2, pp. 391–399. DOI: [10.1111/1365-2745.12713](https://doi.org/10.1111/1365-2745.12713) (cit. on
431 p. 3).
- 432 Bar-Massada, A. & Wood, E. M. (2013). The richness-heterogeneity relationship differs between heterogeneity
433 measures within and among habitats. *Ecography*, **37**.6, pp. 528–535. DOI: [10.1111/j.1600-0587.2013.](https://doi.org/10.1111/j.1600-0587.2013.00590.x)
434 **00590.x** (cit. on p. 2).
- 435 Baroni-Urbani, C. (1980). A statistical table for the degree of coexistence between two species. **44**.3, pp. 287–
436 289. DOI: [10.1007/bf00545229](https://doi.org/10.1007/bf00545229) (cit. on p. 3).
- 437 Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology*
438 *and Biogeography*, **19**.1, pp. 134–143. DOI: [10.1111/j.1466-8238.2009.00490.x](https://doi.org/10.1111/j.1466-8238.2009.00490.x) (cit. on p. 6).
- 439 Ben-Hur, E. & Kadmon, R. (2020). Heterogeneity–diversity relationships in sessile organisms: a unified
440 framework. *Ecology Letters*, **23**.1, pp. 193–207. DOI: [10.1111/ele.13418](https://doi.org/10.1111/ele.13418) (cit. on p. 3).
- 441 Benz, B. W., Robbins, M. B., & Peterson, A. T. (2006). Evolutionary history of woodpeckers and allies (Aves:
442 Picidae): Placing key taxa on the phylogenetic tree. *Molecular Phylogenetics and Evolution*, **40**.2, pp. 389–
443 399. DOI: [10.1016/j.ympev.2006.02.021](https://doi.org/10.1016/j.ympev.2006.02.021) (cit. on p. 12).
- 444 Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2013). Stacking species distribution models and
445 adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**.1, pp. 99–
446 112. DOI: [10.1111/geb.12102](https://doi.org/10.1111/geb.12102) (cit. on p. 3).
- 447 Cao, K., Condit, R., Mi, X., Chen, L., Ren, H., Xu, W., Burslem, D. F. R. P., Cai, C., Cao, M., Chang, L.-W., Chu, C.,
448 Cui, F., Du, H., Ediriweera, S., Gunatilleke, C. S. V., Gunatilleke, I. U. A. N., Hao, Z., Jin, G., Li, J., Li, B., Li, Y.,
449 Liu, Y., Ni, H., O'Brien, M. J., Qiao, X., Shen, G., Tian, S., Wang, X., Xu, H., Xu, Y., Yang, L., Yap, S. L., Lian, J.,
450 Ye, W., Yu, M., Su, S.-H., Chang-Yang, C.-H., Guo, Y., Li, X., Zeng, F., Zhu, D., Zhu, L., Sun, I.-F., Ma, K., &
451 Svenning, J.-C. (2021). Species packing and the latitudinal gradient in beta-diversity. *Proceedings of the*
452 *Royal Society B: Biological Sciences*, **288**.1948. DOI: [10.1098/rspb.2020.3045](https://doi.org/10.1098/rspb.2020.3045) (cit. on p. 3).
- 453 Carmona, C. P. & Pärtel, M. (2020). Estimating probabilistic site-specific species pools and dark diversity
454 from co-occurrence data. *Global Ecology and Biogeography*, **30**.1, pp. 316–326. DOI: [10.1111/geb.13203](https://doi.org/10.1111/geb.13203)
455 (cit. on pp. 5, 13).
- 456 Chao, A., Chazdon, R. L., Colwell, R. K., & Shen, T.-J. (2004). A new statistical approach for assessing
457 similarity of species composition with incidence and abundance data. *Ecology Letters*, **8**.2, pp. 148–159.
458 DOI: [10.1111/j.1461-0248.2004.00707.x](https://doi.org/10.1111/j.1461-0248.2004.00707.x) (cit. on p. 13).
- 459 Chao, A. & Ricotta, C. (2019). Quantifying evenness and linking it to diversity, beta diversity, and similarity.
460 *Ecology*, **100**.12. DOI: [10.1002/ecy.2852](https://doi.org/10.1002/ecy.2852) (cit. on p. 33).
- 461 Chase, J. M., Kraft, N. J. B., Smith, K. G., Vellend, M., & Inouye, B. D. (2011). Using null models to disentangle
462 variation in community dissimilarity from variation in α -diversity. **2**.2, art24. DOI: [10.1890/es10-00117.1](https://doi.org/10.1890/es10-00117.1)
463 (cit. on pp. 5, 6).
- 464 Chase, J. M., McGill, B. J., Thompson, P. L., Antão, L. H., Bates, A. E., Blowes, S. A., Dornelas, M., Gonzalez, A.,
465 Magurran, A. E., Supp, S. R., Winter, M., Bjorkman, A. D., Bruelheide, H., Byrnes, J. E. K., Cabral, J. S., Elahi,
466 R., Gomez, C., Guzman, H. M., Isbell, F., Myers-Smith, I. H., Jones, H. P., Hines, J., Vellend, M., Waldock, C.,
467 & O'Connor, M. (2019). Species richness change across spatial scales. *Oikos*, **128**.8, pp. 1079–1091. DOI:
468 [10.1111/oik.05968](https://doi.org/10.1111/oik.05968) (cit. on p. 2).
- 469 Chung, N. C., Miasojedow, B., Startek, M., & Gambin, A. (2019). Jaccard/Tanimoto similarity test and
470 estimation methods for biological presence-absence data. **20**.S15. DOI: [10.1186/s12859-019-3118-5](https://doi.org/10.1186/s12859-019-3118-5)
471 (cit. on pp. 3, 7).

- 472 Colwell, R. K. & Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **345**.1311, pp. 101–118
473 (cit. on p. 13).
474
- 475 Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series.
476 **19.90**, pp. 297–301. DOI: [10.1090/s0025-5718-1965-0178586-1](https://doi.org/10.1090/s0025-5718-1965-0178586-1) (cit. on p. 13).
- 477 Cressie, N., Davis, A. S., & Folks, J. L. (1981). The moment-generating function and negative integer moments.
478 *The American Statistician*, **35.3**, pp. 148–150. DOI: [10.1080/00031305.1981.10479334](https://doi.org/10.1080/00031305.1981.10479334) (cit. on p. 15).
- 479 Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., Smith, J. R., Hintler, G.,
480 Duguid, M. C., Amatulli, G., Tuanmu, M.-N., Jetz, W., Salas, C., Stam, C., Piotta, D., Tavani, R., Green, S.,
481 Bruce, G., Williams, S. J., Wiser, S. K., Huber, M. O., Hengeveld, G. M., Nabuurs, G.-J., Tikhonova, E.,
482 Borchardt, P., Li, C.-F., Powrie, L. W., Fischer, M., Hemp, A., Homeier, J., Cho, P., Vibrans, A. C., Umunay,
483 P. M., Piao, S. L., Rowe, C. W., Ashton, M. S., Crane, P. R., & Bradford, M. A. (2015). Mapping tree density
484 at a global scale. *Nature*, **525**.7568, pp. 201–205. DOI: [10.1038/nature14967](https://doi.org/10.1038/nature14967) (cit. on p. 2).
- 485 De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel, H., Denys, L.,
486 Vanhecke, L., Gucht, K. Van der, *et al.* (2012). Body size and dispersal mode as key traits determining
487 metacommunity structure of aquatic organisms. *Ecology Letters*, **15.7**, pp. 740–747 (cit. on p. 3).
- 488 Elith, J. & Leathwick, J. R. (2009). Species Distribution Models: ecological explanation and prediction
489 across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40.1**, pp. 677–697. DOI:
490 [10.1146/annurev.ecolsys.110308.120159](https://doi.org/10.1146/annurev.ecolsys.110308.120159) (cit. on pp. 3, 12).
- 491 Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse
492 and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*,
493 **13.3**, pp. 252–264. DOI: [10.1111/j.1472-4642.2007.00341.x](https://doi.org/10.1111/j.1472-4642.2007.00341.x) (cit. on p. 13).
- 494 Fitzpatrick, M. C., Gotelli, N. J., & Ellison, A. M. (2013). MaxEnt versus MaxLike: empirical comparisons
495 with ant species distributions. *Ecosphere*, **4.5**, pp. 1–15. DOI: [10.1890/es13-00066.1](https://doi.org/10.1890/es13-00066.1) (cit. on p. 13).
- 496 Godoy, O., Gómez-Aparicio, L., Matías, L., Pérez-Ramos, I. M., & Allan, E. (2020). An excess of niche differences
497 maximizes ecosystem functioning. *Nature Communications*, **11.1**. DOI: [10.1038/s41467-020-17960-5](https://doi.org/10.1038/s41467-020-17960-5)
498 (cit. on p. 14).
- 499 Godsoe, W. (2013). Inferring the similarity of species distributions using Species' Distribution Models.
500 *Ecography*, **37.2**, pp. 130–136. DOI: [10.1111/j.1600-0587.2013.00403.x](https://doi.org/10.1111/j.1600-0587.2013.00403.x) (cit. on p. 13).
- 501 Godsoe, W. & Case, B. S. (2014). Accounting for shifts in the frequency of suitable environments when
502 testing for niche overlap. *Methods in Ecology and Evolution*, **6.1**, pp. 59–66. DOI: [10.1111/2041-210x.12307](https://doi.org/10.1111/2041-210x.12307)
503 (cit. on p. 13).
- 504 Gotelli, N. J. & Ulrich, W. (2011). Statistical challenges in null model analysis. *Oikos*, **121.2**, pp. 171–180. DOI:
505 [10.1111/j.1600-0706.2011.20301.x](https://doi.org/10.1111/j.1600-0706.2011.20301.x) (cit. on p. 3).
- 506 Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: with applica-*
507 *tions in R*. Cambridge University Press (cit. on pp. 3, 12).
- 508 Harrison, S., Ross, S. J., & Lawton, J. H. (1992). Beta diversity on geographic gradients in Britain. *Journal of*
509 *Animal Ecology*, **61.1**, p. 151. DOI: [10.2307/5518](https://doi.org/10.2307/5518) (cit. on p. 14).
- 510 Hautier, Y., Isbell, F., Borer, E. T., Seabloom, E. W., Harpole, W. S., Lind, E. M., MacDougall, A. S., Stevens, C. J.,
511 Adler, P. B., Alberti, J., Bakker, J. D., Brudvig, L. A., Buckley, Y. M., Cadotte, M., Caldeira, M. C., Chanton,
512 E. J., Chu, C., Daleo, P., Dickman, C. R., Dwyer, J. M., Eskelinen, A., Fay, P. A., Firn, J., Hagenah, N.,
513 Hillebrand, H., Iribarne, O., Kirkman, K. P., Knops, J. M. H., Pierre, K. J. L., McCulley, R. L., Morgan, J. W.,
514 Pärtel, M., Pascual, J., Price, J. N., Prober, S. M., Risch, A. C., Sankaran, M., Schuetz, M., Standish, R. J.,
515 Virtanen, R., Wardle, G. M., Yahdjian, L., & Hector, A. (2017). Local loss and spatial homogenization of
516 plant diversity reduce ecosystem multifunctionality. *Nature Ecology & Evolution*, **2.1**, pp. 50–56. DOI:
517 [10.1038/s41559-017-0395-0](https://doi.org/10.1038/s41559-017-0395-0) (cit. on p. 2).
- 518 Heideman, M., Johnson, D., & Burrus, C. (1984). Gauss and the history of the fast Fourier transform. **1.4**,
519 pp. 14–21. DOI: [10.1109/massp.1984.1162257](https://doi.org/10.1109/massp.1984.1162257) (cit. on p. 13).

- 520 Heino, J., Grönroos, M., Ilmonen, J., Karhu, T., Niva, M., & Paasivirta, L. (2013). Environmental heterogeneity
521 and β diversity of stream macroinvertebrate communities at intermediate spatial scales. *Freshwater*
522 *Science*, **32**.1, pp. 142–154. DOI: 10.1899/12-083.1 (cit. on p. 2).
- 523 Heino, J., Melo, A. S., & Bini, L. M. (2014). Reconceptualising the beta diversity-environmental heterogeneity
524 relationship in running water systems. *Freshwater Biology*, **60**.2, pp. 223–235. DOI: 10.1111/fwb.12502
525 (cit. on p. 3).
- 526 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated
527 climate surfaces for global land areas. *International Journal of Climatology*, **25**.15, pp. 1965–1978. DOI:
528 10.1002/joc.1276 (cit. on p. 37).
- 529 Hoover, E. M. (1936). The measurement of industrial localization. *The Review of Economics and Statistics*,
530 **18**.4, p. 162. DOI: 10.2307/1927875 (cit. on p. 31).
- 531 Hubbard, J. (1959). Calculation of partition functions. *Physical Review Letters*, **3**.2, pp. 77–78. DOI: 10.1103/
532 physrevlett.3.77 (cit. on p. 35).
- 533 Hui, C. & McGeoch, M. A. (2014). Zeta diversity as a concept and metric that unifies incidence-based
534 biodiversity patterns. *The American Naturalist*, **184**.5, pp. 684–694. DOI: 10.1086/678125 (cit. on p. 3).
- 535 Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, **44**, pp. 223–270
536 (cit. on pp. 3, 5, 6).
- 537 — (1912). The Distribution of the flora in the alpine zone. *New Phytologist*, **11**.2, pp. 37–50. DOI: 10.1111/j.
538 1469-8137.1912.tb05611.x (cit. on pp. 3, 5, 6).
- 539 Jankowski, J. E., Ciecka, A. L., Meyer, N. Y., & Rabenold, K. N. (2009). Beta diversity along environmental
540 gradients: implications of habitat specialization in tropical montane landscapes. *Journal of Animal*
541 *Ecology*, **78**.2, pp. 315–327. DOI: 10.1111/j.1365-2656.2008.01487.x (cit. on p. 2).
- 542 Jost, L. (2006). Entropy and diversity. *Oikos*, **113**.2, pp. 363–375. DOI: 10.1111/j.2006.0030-1299.14714.x
543 (cit. on p. 33).
- 544 — (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, **88**.10, pp. 2427–
545 2439. DOI: 10.1890/06-1736.1 (cit. on p. 33).
- 546 Kalyuzhny, M., Flather, C. H., Shnerb, N. M., & Kadmon, R. (2021). A framework for quantifying deviations
547 from dynamic equilibrium theory. *Global Ecology and Biogeography*, **31**.1, pp. 183–195. DOI: 10.1111/
548 geb.13409 (cit. on pp. 7, 12).
- 549 Karp, D. S., Frishkoff, L. O., Echeverri, A., Zook, J., Juárez, P., & Chan, K. M. A. (2017). Agriculture erases
550 climate-driven β -diversity in Neotropical bird communities. *Global Change Biology*, **24**.1, pp. 338–349.
551 DOI: 10.1111/gcb.13821 (cit. on p. 2).
- 552 Kattge, J., Bönsch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D., Aakala, T.,
553 Abedi, M., *et al.* (2020). TRY plant trait database – enhanced coverage and open access. *Global Change*
554 *Biology*, **26**.1, pp. 119–188. DOI: 10.1111/gcb.14904 (cit. on p. 3).
- 555 Keil, P., Wiegand, T., Tóth, A. B., McGlenn, D. J., & Chase, J. M. (2021). Measurement and analysis of interspe-
556 cific spatial associations as a facet of biodiversity. *Ecological Monographs*, **91**.3. DOI: 10.1002/ecm.1452
557 (cit. on p. 3).
- 558 Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence-absence data. *Journal*
559 *of Animal Ecology*, **72**.3, pp. 367–382. DOI: 10.1046/j.1365-2656.2003.00710.x (cit. on p. 2).
- 560 Lange, K. (2010). *Applied Probability*. 2nd ed. Springer Texts in Statistics. Springer (cit. on p. 15).
- 561 Legendre, P. (2019). A temporal beta-diversity index to identify sites that have changed in exceptional ways
562 in space–time surveys. *Ecology and Evolution*, **9**.6, pp. 3500–3514. DOI: 10.1002/ece3.4984 (cit. on pp. 2,
563 3, 5, 11, 12).
- 564 Legendre, P. & Cáceres, M. D. (2013). Beta diversity as the variance of community data: dissimilarity
565 coefficients and partitioning. *Ecology Letters*, **16**.8, pp. 951–963. DOI: 10.1111/ele.12141 (cit. on p. 2).
- 566 Leinster, T. (2021). *Entropy and Diversity: The Axiomatic Approach*. Cambridge University Press (cit. on p. 33).
- 567 Loiseau, N., Legras, G., Kulbicki, M., Mérigot, B., Harmelin-Vivien, M., Mazouni, N., Galzin, R., & Gaertner, J.
568 (2016). Multi-component β -diversity approach reveals conservation dilemma between species and

- 569 functions of coral reef fishes. *Journal of Biogeography*, **44.3**, pp. 537–547. DOI: 10.1111/jbi.12844 (cit. on
570 p. 14).
- 571 Loreau, M. (2000). Are communities saturated? On the relationship between alpha, beta and gamma
572 diversity. *Ecology Letters*, **3.2**, pp. 73–76. DOI: 10.1046/j.1461-0248.2000.00127.x (cit. on p. 2).
- 573 Lu, M. (2021). Complex relationships between beta diversity and dispersal in meta-community models.
574 *Ecography*, **44.12**, pp. 1769–1780. DOI: 10.1111/ecog.05937 (cit. on pp. 2, 7, 12).
- 575 Lu, M., Vasseur, D., & Jetz, W. (2019). Beta diversity patterns derived from island biogeography theory. *The*
576 *American Naturalist*, **194.3**, E52–E65. DOI: 10.1086/704181 (cit. on pp. 2, 6, 7, 12, 25).
- 577 Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA (cit. on
578 p. 7).
- 579 MacArthur, R. H. & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, **17.4**,
580 p. 373. DOI: 10.2307/2407089 (cit. on pp. 3, 5, 8).
- 581 MacArthur, R. H. & Wilson, E. O. (1967). *The theory of island biogeography*. **1**. Princeton University Press
582 (cit. on pp. 3, 5, 8).
- 583 MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2018). *Occupancy*
584 *estimation and modeling: inferring patterns and dynamics of species occurrence*. 2nd ed. Elsevier (cit. on
585 p. 14).
- 586 Maggini, R., Lehmann, A., Zbinden, N., Zimmermann, N. E., Bolliger, J., Schröder, B., Foppen, R., Schmid, H.,
587 Beniston, M., & Jenni, L. (2014). Assessing species vulnerability to climate and land use change: the
588 case of the Swiss breeding birds. *Diversity and Distributions*, **20.6**, pp. 708–719. DOI: 10.1111/ddi.12207
589 (cit. on p. 12).
- 590 Magurran, A. E., Dornelas, M., Moyes, F., & Henderson, P. A. (2019). Temporal β diversity—a macroecological
591 perspective. *Global Ecology and Biogeography*, **28.12**, pp. 1949–1960. DOI: 10.1111/geb.13026 (cit. on
592 pp. 3, 5, 11, 12).
- 593 Marshall, A. W., Olkin, I., & Arnold, B. C. (1979). *Inequalities: theory of majorization and its applications*. **143**.
594 Springer (cit. on p. 33).
- 595 Maynard, D. S., Crowther, T. W., & Bradford, M. A. (2017). Competitive network determines the direction of
596 the diversity–function relationship. *Proceedings of the National Academy of Sciences*, **114.43**, pp. 11464–
597 11469. DOI: 10.1073/pnas.1712211114 (cit. on p. 2).
- 598 McKinney, M. L. (2006). Urbanization as a major cause of biotic homogenization. *Biological Conservation*,
599 **127.3**, pp. 247–260. DOI: 10.1016/j.biocon.2005.09.005 (cit. on p. 2).
- 600 McVinish, R. & Lester, R. J. G. (2020). Measuring aggregation in parasite populations. *Journal of The Royal*
601 *Society Interface*, **17.165**, p. 20190886. DOI: 10.1098/rsif.2019.0886 (cit. on p. 33).
- 602 Mori, A. S., Isbell, F., & Seidl, R. (2018). β -diversity, community assembly, and ecosystem functioning. *Trends*
603 *in Ecology & Evolution*, **33.7**, pp. 549–564. DOI: 10.1016/j.tree.2018.04.012 (cit. on pp. 2, 14).
- 604 Mouquet, N. & Loreau, M. (2003). Community Patterns in Source-Sink Metacommunities. *The American*
605 *Naturalist*, **162.5**, pp. 544–557. DOI: 10.1086/378857 (cit. on p. 2).
- 606 Myers, J. A., Chase, J. M., Crandall, R. M., & Jiménez, I. (2015). Disturbance alters beta-diversity but not the
607 relative importance of community assembly mechanisms. *Journal of Ecology*, **103.5**, pp. 1291–1299. DOI:
608 10.1111/1365-2745.12436 (cit. on p. 2).
- 609 Nieto-Lugilde, D., Maguire, K. C., Blois, J. L., Williams, J. W., & Fitzpatrick, M. C. (2017). Multiresponse
610 algorithms for community-level modelling: Review of theory, applications, and comparison to species
611 distribution models. *Methods in Ecology and Evolution*, **9.4**, pp. 834–848. DOI: 10.1111/2041-210x.12936
612 (cit. on p. 13).
- 613 Ojima, M. N. & Jiang, L. (2016). Interactive effects of disturbance and dispersal on community assembly.
614 *Oikos*, **126.5**, pp. 682–691. DOI: 10.1111/oik.03265 (cit. on p. 2).
- 615 Olden, J. D., Comte, L., & Giam, X. (2018). The Homogocene: a research prospectus for the study of biotic
616 homogenisation. *NeoBiota*, **37**, pp. 23–36. DOI: 10.3897/neobiota.37.22552 (cit. on p. 2).
- 617 Olden, J. D. & Poff, N. L. (2003). Toward a Mechanistic Understanding and Prediction of Biotic Homogeniza-
618 tion. *The American Naturalist*, **162.4**, pp. 442–460. DOI: 10.1086/378212 (cit. on p. 2).

- 619 Olden, J. D. & Rooney, T. P. (2006). On defining and quantifying biotic homogenization. *Global Ecology and*
620 *Biogeography*, **15**.2, pp. 113–120. DOI: 10.1111/j.1466-822x.2006.00214.x (cit. on p. 2).
- 621 Ontiveros, V. J., Capitán, J. A., Casamayor, E. O., & Alonso, D. (2021). The characteristic time of ecological
622 communities. *Ecology*, **102**.2. DOI: 10.1002/ecy.3247 (cit. on pp. 3, 7, 28).
- 623 Palacio, F. X., Callaghan, C. T., Cardoso, P., Hudgins, E. J., Jarzyna, M. A., Ottaviani, G., Riva, F., Graco-Roza, C.,
624 Shirey, V., & Mammola, S. (2022). A protocol for reproducible functional diversity analyses. *Ecography*,
625 **2022**.11. DOI: 10.1111/ecog.06287 (cit. on p. 14).
- 626 Pärtel, M., Szava-Kovats, R., & Zobel, M. (2011). Dark diversity: shedding light on absent species. *Trends in*
627 *Ecology & Evolution*, **26**.3, pp. 124–128. DOI: 10.1016/j.tree.2010.12.004 (cit. on p. 13).
- 628 Pasinelli, G. (2007). “Nest site selection in middle and great spotted woodpeckers *Dendrocopos medius* and
629 *D. major*: implications for forest management and conservation”. *Topics in Biodiversity and Conservation*.
630 Springer Netherlands, pp. 457–472. DOI: 10.1007/978-1-4020-6320-6_30 (cit. on p. 12).
- 631 Peres-Neto, P. R., Olden, J. D., & Jackson, D. A. (2001). Environmentally constrained null models: site
632 suitability as occupancy criterion. *Oikos*, **93**.1, pp. 110–120. DOI: 10.1034/j.1600-0706.2001.930112.x
633 (cit. on pp. 2, 3).
- 634 Pilowsky, J. A., Colwell, R. K., Rahbek, C., & Fordham, D. A. (2022). Process-explicit models reveal the
635 structure and dynamics of biodiversity patterns. *Science Advances*, **8**.31. DOI: 10.1126/sciadv.abj2271
636 (cit. on p. 14).
- 637 Plotkin, J. B. & Muller-Landau, H. C. (2002). Sampling the species composition of a landscape. *Ecology*, **83**.12,
638 pp. 3344–3356 (cit. on p. 13).
- 639 Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the interpretations
640 of joint modeling in community ecology. *Trends in Ecology & Evolution*, **36**.5, pp. 391–401. DOI: 10.1016/
641 *j.tree.2021.01.002* (cit. on p. 2).
- 642 Pomara, L. Y., Ruokolainen, K., & Young, K. R. (2013). Avian species composition across the Amazon River:
643 the roles of dispersal limitation and environmental heterogeneity. *Journal of Biogeography*, **41**.4, pp. 784–
644 796. DOI: 10.1111/jbi.12247 (cit. on p. 2).
- 645 Pons, J.-M., Olioso, G., Cruaud, C., & Fuchs, J. (2010). Phylogeography of the Eurasian green woodpecker
646 (*Picus viridis*). *Journal of Biogeography*, **38**.2, pp. 311–325. DOI: 10.1111/j.1365-2699.2010.02401.x
647 (cit. on p. 12).
- 648 Powell, K. I., Chase, J. M., & Knight, T. M. (2013). Invasive plants have scale-dependent effects on diversity
649 by altering species-area relationships. *Science*, **339**.6117, pp. 316–318. DOI: 10.1126/science.1226817
650 (cit. on p. 2).
- 651 Real, R., Barbosa, A. M., & Bull, J. W. (2016). Species distributions, quantum theory, and the enhancement
652 of biodiversity measures. *Systematic Biology*, syw072. DOI: 10.1093/sysbio/syw072 (cit. on p. 5).
- 653 Real, R. & Vargas, J. M. (1996). The probabilistic basis of Jaccard’s index of similarity. *Systematic Biology*, **45**.3,
654 pp. 380–385. DOI: 10.1093/sysbio/45.3.380 (cit. on pp. 3, 14).
- 655 Routledge, R. D. (1983). Evenness indices: are any admissible? *Oikos*, **40**.1, p. 149. DOI: 10.2307/3544211
656 (cit. on p. 33).
- 657 Royan, A., Reynolds, S. J., Hannah, D. M., Prudhomme, C., Noble, D. G., & Sadler, J. P. (2015). Shared
658 environmental responses drive co-occurrence patterns in river bird communities. *Ecography*, **39**.8,
659 pp. 733–742. DOI: 10.1111/ecog.01703 (cit. on p. 3).
- 660 Santos, F., Lima, M. G. M., Espinosa, S., Ahumada, J. A., Jansen, P. A., Spironello, W. R., Hurtado, J., Juen, L.,
661 & Peres, C. A. (2021). Site and species contribution to β -diversity in terrestrial mammal communities:
662 Evidence from multiple Neotropical forest sites. *Science of The Total Environment*, **789**, p. 147946. DOI:
663 10.1016/j.scitotenv.2021.147946 (cit. on p. 3).
- 664 Schmid, H., M. Kestenholz, Knaus, P., L. Rey, & Sattler, T. (2018). The state of birds in Switzerland: special
665 issue on the breeding bird atlas 2013–2016. en. DOI: 10.13140/RG.2.2.17872.15366 (cit. on pp. 12, 37).
- 666 Schmid, H., Posse, B., & Vogelwarte, S. S. (1998). *Schweizer Brutvogelatlas: Verbreitung der Brutvögel in der*
667 *Schweiz und im Fürstentum Liechtenstein 1993-1996*. Schweizerische Vogelwarte (cit. on pp. 12, 37).

- 668 Schrader, J., Wright, I. J., Kreft, H., Weigelt, P., Andrew, S. C., Abbott, I., & Westoby, M. (2022). ETIB-T: An
669 Equilibrium Theory of Island Biogeography for plant traits. *Journal of Biogeography*, **50**.1, pp. 223–234.
670 DOI: [10.1111/jbi.14526](https://doi.org/10.1111/jbi.14526) (cit. on p. 3).
- 671 Shinohara, N., Hongo, Y., Ichinokawa, M., Nishijima, S., Sawayama, S., Kurogi, H., Uto, Y., Mita, H., Ishii,
672 M., Kusano, A., & Akimoto, S. (2022). Similar fish species composition despite larger environmental
673 heterogeneity during severe hypoxia in a coastal ecosystem. *Ecology and Evolution*, **12**.5. DOI: [10.1002/
674 ece3.8884](https://doi.org/10.1002/ece3.8884) (cit. on p. 3).
- 675 Silva, P. G. da, Hernández, M. I. M., & Heino, J. (2018). Disentangling the correlates of species and site
676 contributions to beta diversity in dung beetle assemblages. *Diversity and Distributions*, **24**.11, pp. 1674–
677 1686. DOI: [10.1111/ddi.12785](https://doi.org/10.1111/ddi.12785) (cit. on p. 3).
- 678 Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity
679 conservation? *Trends in Ecology & Evolution*, **31**.1, pp. 67–80. DOI: [10.1016/j.tree.2015.11.005](https://doi.org/10.1016/j.tree.2015.11.005) (cit. on
680 pp. 2, 14).
- 681 Soininen, J., Lennon, J. J., & Hillebrand, H. (2007). A multivariate analysis of beta diversity across organisms
682 and environments. *Ecology*, **88**.11, pp. 2830–2838. DOI: [10.1890/06-1730.1](https://doi.org/10.1890/06-1730.1) (cit. on p. 3).
- 683 Tatsumi, S., Iritani, R., & Cadotte, M. W. (2021). Temporal changes in spatial variation: partitioning the
684 extinction and colonisation components of beta diversity. *Ecology Letters*, **24**.5, pp. 1063–1072. DOI:
685 [10.1111/ele.13720](https://doi.org/10.1111/ele.13720) (cit. on p. 13).
- 686 Thompson, P. L., Guzman, L. M., Meester, L. D., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S.,
687 & Chase, J. M. (2020). A process-based metacommunity framework linking local and regional scale
688 community ecology. **23**.9, pp. 1314–1329. DOI: [10.1111/ele.13568](https://doi.org/10.1111/ele.13568) (cit. on p. 2).
- 689 Van Lint, J. H. & Wilson, R. M. (2001). *A course in combinatorics*. Cambridge University Press (cit. on pp. 6,
690 26).
- 691 Vannette, R. L. & Fukami, T. (2017). Dispersal enhances beta diversity in nectar microbes. *Ecology Letters*,
692 **20**.7, pp. 901–910. DOI: [10.1111/ele.12787](https://doi.org/10.1111/ele.12787) (cit. on p. 2).
- 693 Veech, J. A. (2012). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeogra-*
694 *phy*, **22**.2, pp. 252–260. DOI: [10.1111/j.1466-8238.2012.00789.x](https://doi.org/10.1111/j.1466-8238.2012.00789.x) (cit. on p. 3).
- 695 Veech, J. A. & Crist, T. O. (2007). Habitat and climate heterogeneity maintain beta-diversity of birds among
696 landscapes within ecoregions. *Global Ecology and Biogeography*, **16**.5, pp. 650–656. DOI: [10.1111/j.1466-
697 8238.2007.00315.x](https://doi.org/10.1111/j.1466-8238.2007.00315.x) (cit. on p. 2).
- 698 Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, **85**.2,
699 pp. 183–206. DOI: [10.1086/652373](https://doi.org/10.1086/652373) (cit. on p. 2).
- 700 Vellend, M., Verheyen, K., Flinn, K. M., Jacquemyn, H., Kolb, A., Calster, H. van, Peterken, G., Graae, B. J.,
701 Bellemare, J., Honnay, O., Brunet, J., Wulf, M., Gerhardt, F., & Hermy, M. (2007). Homogenization of
702 forest plant communities and weakening of species? environment relationships via agricultural land
703 use. *Journal of Ecology*, **95**.3, pp. 565–573. DOI: [10.1111/j.1365-2745.2007.01233.x](https://doi.org/10.1111/j.1365-2745.2007.01233.x) (cit. on p. 2).
- 704 Wang, S., Loreau, M., Mazancourt, C., Isbell, F., Beierkuhnlein, C., Connolly, J., Deutschman, D. H., Doležal, J.,
705 Eisenhauer, N., Hector, A., Jentsch, A., Kreyling, J., Lanta, V., Lepš, J., Polley, H. W., Reich, P. B., Ruijven, J.,
706 Schmid, B., Tilman, D., Wilsey, B., & Craven, D. (2021). Biotic homogenization destabilizes ecosystem
707 functioning by decreasing spatial asynchrony. *Ecology*, **102**.6. DOI: [10.1002/ecy.3332](https://doi.org/10.1002/ecy.3332) (cit. on p. 2).
- 708 Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, **21**.2-3, pp. 213–251. DOI:
709 [10.2307/1218190](https://doi.org/10.2307/1218190) (cit. on pp. 2, 23, 24).
- 710 Wolda, H. (1981). Similarity indices, sample size and diversity. *Oecologia*, **50**.3, pp. 296–302 (cit. on p. 13).
- 711 Zorzal-Almeida, S., Bini, L. M., & Bicudo, D. C. (2017). Beta diversity of diatoms is driven by environmental
712 heterogeneity, spatial extent and productivity. *Hydrobiologia*, **800**.1, pp. 7–16. DOI: [10.1007/s10750-017-
713 3117-3](https://doi.org/10.1007/s10750-017-3117-3) (cit. on p. 3).
- 714 Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita,
715 G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz,
716 D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., Yates, K. L., Zimmermann, N. E., & Merow, C. (2020).

- 717 A standard protocol for reporting species distribution models. *Ecography*, **43**.9, pp. 1261–1277. DOI:
718 [10.1111/ecog.04960](https://doi.org/10.1111/ecog.04960) (cit. on pp. 3, 12).
- 719 Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2019a). *Data from:*
720 *Testing species assemblage predictions from stacked and joint species distribution models*. en. DOI: [10.5061/](https://doi.org/10.5061/DRYAD.K88V330)
721 [DRYAD.K88V330](https://doi.org/10.5061/DRYAD.K88V330) (cit. on p. 37).
- 722 — (2019b). Testing species assemblage predictions from stacked and joint species distribution models.
723 *Journal of Biogeography*, **47**.1, pp. 101–113. DOI: [10.1111/jbi.13608](https://doi.org/10.1111/jbi.13608) (cit. on pp. 12, 37).

724 Notation

- 725 ○ $\Omega := \{0, 1\}$
- 726 ○ S , the species pool size, defined as the number of elements $\{i\}$ such that $\{\max_{j=1,2} p_{i,j} > 0\}$
- 727 ○ $x_{i,j} \in \Omega$: incidence
- 728 ○ $\bar{x}_{i,j} \in \Omega$: logical negation, i.e., $\bar{x}_{i,j} = 1 - x_{i,j}$
- 729 ○ $p_{i,j}$: probability that $x_{i,j} = 1$
- 730 ○ $a_{i,j}$: probability that $x_{i,j} = 0$
- 731 ○ $\mathbf{X} \in \Omega^S \otimes \Omega^2$: Incidence table of size S rows and N columns
- 732 ○ $\mathbf{x}_{o,j} := (x_{1,j}, \dots, x_{S,j})^\top (\in \Omega^S)$, referred to as “local compositional profile”
- 733 ○ $\mathbf{x}_{i,o} := (x_{i,1}, \dots, x_{i,N}) (\in \Omega^N)$, referred to as “species occurrence profile”
- 734 ○ $P_{\mathbf{X}}$: Probability that the incidence table \mathbf{X} realizes
- 735 ○ $\langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle := \sum_{i=1}^S x_{i,1}x_{i,2}$, or the inner product between local compositional profiles to count the number
- 736 of common species
- 737 ○ $|\mathbf{x}_{o,j}| := \langle \mathbf{x}_{o,j}, \mathbf{x}_{o,j} \rangle$: the total number of species present in a site j
- 738 ○ $\mathbf{X} = (\mathbf{x}_{o,1}, \mathbf{x}_{o,2})$ as we consider only two sites.

739 Appendix A Expectation of Jaccard dissimilarity

740 Definition of Jaccard dissimilarity

741 We write $\beta_{\mathbf{X}}^J$ for the Jaccard dissimilarity measure for a table \mathbf{X} , defined by:

$$\beta_{\mathbf{X}}^J := \frac{|\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}| - 2 \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle}{|\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}| - \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle}. \quad (\text{A7})$$

742 For $\mathbf{X} = \mathbf{O}$ (zero-matrix), we define $\beta_{\mathbf{O}}^J := 0$, which follows from two facts: (i) two all-zero vectors are (or

743 axiomatically should be) completely similar, and (ii) the nullification of the denominator (which is always

744 larger or equal to the numerator) should imply the nullification of the numerator (which is smaller or

745 at most equal) as well. To avoid confusion, we suppose that numerator being zero implies the Jaccard

746 dissimilarity be zero (otherwise resulting in erroneous calculations). It makes sense to exclude the zero-

747 matrix, because zero-matrix indicates that there is no species in the landscape. Therefore we will focus on

748 the conditional expectation.

749 Linking Jaccard dissimilarity to Whittaker’s (1972) beta-diversity

750 Whittaker (1972) defined beta-diversity as the ratio between region-wide diversity (gamma) to average

751 local richness (alpha). Because the region-wide diversity is $\gamma_{\mathbf{X}} := |\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}| - \langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle$ and average

752 local richness is $\alpha_{\mathbf{X}} := (|\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}|) / 2$, we have:

$$|\mathbf{x}_{o,1}| + |\mathbf{x}_{o,2}| = 2\alpha_{\mathbf{X}}, \quad (\text{A8})$$

$$\langle \mathbf{x}_{o,1}, \mathbf{x}_{o,2} \rangle = 2\alpha_{\mathbf{X}} - \gamma_{\mathbf{X}}. \quad (\text{A9})$$

753 Using the beta-diversity sensu Whittaker (1972), which is given by:

$$\beta_{\mathbf{X}}^{\text{Whittaker}} := \frac{\gamma_{\mathbf{X}}}{\alpha_{\mathbf{X}}}, \quad (\text{A10})$$

754 Jaccard dissimilarity can be rewritten as:

$$\beta_{\mathbf{X}}^J = \frac{2\gamma_{\mathbf{X}} - 2\alpha_{\mathbf{X}}}{\gamma_{\mathbf{X}}} = 2 - \frac{2}{\beta_{\mathbf{X}}^{\text{Whittaker}}}, \quad (\text{A11})$$

755 giving:

$$\beta_{\mathbf{X}}^{\text{Whittaker}} = \frac{2}{2 - \beta_{\mathbf{X}}^{\text{J}}}, \quad (\text{A12})$$

756 which is a monotonically increasing function of $\beta_{\mathbf{X}}^{\text{J}}$. Geometrically, $\beta_{\mathbf{X}}^{\text{Whittaker}}$ is a slope between two points
757 $(2, 2)$ and $(\beta_{\mathbf{X}}^{\text{J}}, 0)$, which becomes steeper as $\beta_{\mathbf{X}}^{\text{J}}$ increases. Because Whittaker's (1972) beta-diversity is a
758 monotonic transform of Jaccard dissimilarity, we can use Jaccard dissimilarity as a measure of beta-diversity.

759 Step 1: express the Jaccard dissimilarity as an integral

760 We note that any fraction of v_1 to v_2 (≥ 0) has a form of integration:

$$\frac{v_1}{v_2} \equiv \int_0^{+\infty} \frac{\partial}{\partial \xi} \exp(v_1 \xi - v_2 \theta) d\theta \Big|_{\xi=0}, \quad (\text{A13})$$

761 which yields:

$$\beta_{\mathbf{X}}^{\text{J}} \equiv \int_0^{+\infty} \frac{\partial}{\partial \xi} \exp\left(\xi \sum_{i=1}^S (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2})\right) \exp\left(-\theta \sum_{i=1}^S (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2})\right) d\theta \Big|_{\xi=0}, \quad (\text{A14})$$

762 where we assign that we do not interchange the integral with the derivative unless otherwise stated, in
763 order to remind that the integral should be defined as zero whenever the numerator is zero. We compute
764 the expectation of $\beta_{\mathbf{X}}^{\text{J}}$ (which is a stochastic variable) over the distribution $P_{\mathbf{X}}$.

765 Step 2: Independence yields product

766 Assuming the species independence, the probability that a given incidence table \mathbf{X} is observed is given by:

$$P_{\mathbf{X}} = \prod_{i=1}^S \prod_{j=1}^2 p_{i,j}^{x_{i,j}} a_{i,j}^{\bar{x}_{i,j}}, \quad (\text{A15})$$

767 as the incidence probabilities are independent across species. We can then see the following facts:

$$\begin{aligned} \sum_{\mathbf{X}} P_{\mathbf{X}} \prod_{i=1}^S (\bullet) &\equiv \prod_{i=1}^S \sum_{\mathbf{x}_{i,\circ} \in \Omega^2} p_{i,1}^{x_{i,1}} p_{i,2}^{x_{i,2}} a_{i,1}^{1-x_{i,1}} a_{i,2}^{1-x_{i,2}} (\bullet) \\ \exp\left(\sum_{i=1}^S (\bullet)\right) &\equiv \prod_{i=1}^S \exp(\bullet) \end{aligned} \quad (\text{A16})$$

768 Then we get the (unconditional) expectation as:

$$\begin{aligned} \mathbb{E}[\beta^{\text{J}}] &:= \sum_{\mathbf{X}} P_{\mathbf{X}} \beta_{\mathbf{X}}^{\text{J}} \\ &= \int_0^{+\infty} \frac{\partial}{\partial \xi} \prod_{i=1}^S \sum_{\mathbf{x}_{i,\circ} \in \Omega^2} p_{i,1}^{x_{i,1}} p_{i,2}^{x_{i,2}} a_{i,1}^{1-x_{i,1}} a_{i,2}^{1-x_{i,2}} \exp\left(\xi (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) - \theta (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2})\right) d\theta \Big|_{\xi=0}. \end{aligned} \quad (\text{A17})$$

769 Step 3: Boolean thinking

770 Let us evaluate the Boolean variable in the argument of exponential:

$$\xi (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) - \theta (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) = \begin{cases} 0, & \text{if } x_{i,1} = x_{i,2} = 0 \text{ (double-absence)} \\ -\theta, & \text{if } x_{i,1} = x_{i,2} = 1 \text{ (double-presence)} \\ \xi - \theta, & \text{otherwise (uniqueness)} \end{cases} \quad (\text{A18})$$

771 Using this can allow us to expand the summation $\sum_{x_{i,o} \in \Omega^2}$; that is:

$$\begin{aligned} & \sum_{x_{i,o} \in \Omega^2} p_{i,1}^{x_{i,1}} p_{i,2}^{x_{i,2}} a_{i,1}^{1-x_{i,1}} a_{i,2}^{1-x_{i,2}} \exp\left(\xi(x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) - \theta(x_{i,1} + x_{i,2} - x_{i,1}x_{i,2})\right) \\ & = d_{i,o} + e^{-\theta} b_{i,o} + e^{\xi-\theta} (1 - b_{i,o} - d_{i,o}) \end{aligned} \quad (\text{A19})$$

772 for all $i \in \{1, \dots, S\}$. Therefore, substituting this into Eqn (A17) results in:

$$\mathbf{E}[\beta^J] = \int_0^\infty \frac{\partial}{\partial \xi} \prod_{i=1}^S \left(d_{i,o} + e^{-\theta} b_{i,o} + e^{\xi-\theta} (1 - b_{i,o} - d_{i,o}) \right) d\theta \Bigg|_{\xi=0}. \quad (\text{A20})$$

773 Step 4: apply Leibniz rule

774 By using Leibniz rule of the derivative of a product, we can get:

$$\mathbf{E}[\beta^J] = \int_0^\infty e^{-\theta} \sum_{i=1}^S (1 - b_{i,o} - d_{i,o}) \prod_{\ell=1, \ell \neq i}^S \left(d_{\ell,o} + e^{-\theta} (1 - d_{\ell,o}) \right) d\theta. \quad (\text{A21})$$

775 By transforming the variable $z = 1 - e^{-\theta}$ with $d\theta = (1 - z) dz$, we can rewrite Eqn (A21) as:

$$\mathbf{E}[\beta^J] = \int_0^1 \sum_{i=1}^S (1 - b_{i,o} - d_{i,o}) \prod_{\ell=1, \ell \neq i}^S (1 - z + z d_{\ell,o}) dz. \quad (\text{A22})$$

776 Eqn (A22) represents the general expression for the expectation of Jaccard dissimilarity provided that
777 species incidences are uncorrelated.

778 Break to check: experiments

779 Experiment 1 | When $S = 1$, we immediately get $\mathbf{E}[\beta^J] \Big|_{S=1} = 1 - b_{1,o} - d_{1,o}$. Thus the conditional expectation
780 is $(1 - b_{1,o} - d_{1,o}) / (1 - d_{1,o})$.

781 Experiment 2 | When $S = 2$,

$$\mathbf{E}[\beta^J] \Big|_{S=2} = \int_0^1 (1 - b_{1,o} - d_{1,o}) (1 - z + z d_{2,o}) + (1 - b_{2,o} - d_{2,o}) (1 - z + z d_{1,o}) dz, \quad (\text{A23})$$

782 which is $(1 - b_{1,o} - d_{1,o}) (1 - 1/2 + d_{2,o}/2) + (1 - b_{2,o} - d_{2,o}) (1 - 1/2 + d_{1,o}/2)$. Thus
783 the conditional expectation is

$$\mathbf{E}[\beta^J | \gamma > 0] \Big|_{S=2} = \frac{1}{2(1 - d_{1,o} d_{2,o})} \left((1 - b_{1,o} - d_{1,o}) (1 + d_{2,o}) + (1 - b_{2,o} - d_{2,o}) (1 + d_{1,o}) \right) \quad (\text{A24})$$

784 Experiment 3 | When all species are equal, that is when $(p_{i,1}, p_{i,2}) \equiv (p_{o,1}, p_{o,2})$ with $p_{i,1} p_{i,2} = b$ and
785 $a_{i,1} a_{i,2} = d$,

$$\mathbf{E}[\beta^J] = \int_0^1 S(1 - b - d) (1 - z + zd)^{S-1} dz = \frac{1 - b - d}{1 - d} \cdot (1 - d^S), \quad (\text{A25})$$

786 thus recovering Lu *et al.*'s (2019) results by dividing the RHS by $1 - d^S$ the probability that
787 some species is present ($\gamma > 0$).

788 Rationale

789 First, notice that $1 - b_{i,o} - d_{i,o}$ represents the probability that species i is unique to one of the sites. Given that
790 species i is unique, we aim to identify which species are present, regardless of being unique or common. For

791 instance, given that species $i = 1$ is unique, the other species $i = 2, 3, \dots, S$, each of which is either unique,
 792 common or double-absent, we can count the number of present species and put it in the denominator by
 793 calculating an integral of the product:

$$M_1 := \int_0^1 \prod_{\ell \geq 2}^S (1 - z + z d_{\ell, o}) dz. \quad (\text{A26})$$

794 Indeed, when $S = 2$, the integral results in $M_1 = (1 + d_{2, o}) / 2$, because with probability $d_{2, o}$, species 2
 795 is absent from both sites, in which the contribution of species 1 to Jaccard dissimilarity is 1, while with
 796 probability $1 - d_{2, o}$, species 2 is present, in which case the contribution of species 1 to Jaccard dissimilarity
 797 is $1/2$ (with species 2's contribution not counted here), thus giving the expectation of $(1 + d_{2, o}) / 2$. When
 798 $S = 3$, given that species i is unique, writing 00 for double-absence of species 2 and 3 and $\overline{00}$ for non
 799 double-absence of species 2 and 3,

sp 2	sp 3	probability	# present sp (incl 1)	species 1's contribution to Jaccard
00	00	$d_{2, o} d_{3, o}$	1	1
00	$\overline{00}$	$d_{2, o} (1 - d_{3, o})$	2	$1 / 2$
$\overline{00}$	00	$(1 - d_{2, o}) d_{3, o}$	2	$1 / 2$
$\overline{00}$	$\overline{00}$	$(1 - d_{2, o}) (1 - d_{3, o})$	3	$1 / 3$

(A27)

800 The expected contribution of species 1 to Jaccard dissimilarity, conditioned on species 1 being unique, is
 801 thus given by

$$d_{2, o} d_{3, o} \cdot 1 + d_{2, o} (1 - d_{3, o}) \cdot \frac{1}{2} + (1 - d_{2, o}) d_{3, o} \cdot \frac{1}{2} + (1 - d_{2, o}) (1 - d_{3, o}) \cdot \frac{1}{3} = \frac{2 + d_{2, o} + d_{3, o} + 2d_{2, o} d_{3, o}}{6} = M_1 \quad (\text{A28})$$

802 where the second line results from calculation of M_1 for $S = 3$. From this reasoning, we can interpret
 803 Eqn (A22) as the sum of the conditional expectations of species' contribution to Jaccard dissimilarity.

804 Step 5: reach Beta function

805 Expanding the product in Eqn (A22) in terms of $1 - z$ and z , we get:

$$\begin{aligned} \mathbb{E}[\beta^J] &= \sum_{i=1}^S (1 - b_{i, o} - d_{i, o}) \int_0^1 \left(\prod_{\ell=1; (\ell \neq i)}^S (1 - z + z d_{\ell, o}) \right) dz \\ &= \sum_{i=1}^S (1 - b_{i, o} - d_{i, o}) \int_0^1 \sum_{k=1}^S \sum_{\mathbf{y} \in \Omega^S | |\mathbf{y}|=k, y_i=1} (1 - z)^{k-1} z^{S-k} \prod_{\ell=1, \ell \neq i}^S d_{i, o}^{1-y_\ell} dz \end{aligned} \quad (\text{A29})$$

806 Using the Beta function $B(k, S - k + 1) := \int_0^1 (1 - z)^{k-1} z^{S-k} dz = (k - 1)!(S - k)! / S!$, we can rewrite $\mathbb{E}[\beta^J]$ as:

$$\begin{aligned} \mathbb{E}[\beta^J] &= \sum_{i=1}^S (1 - b_{i, o} - d_{i, o}) \int_0^1 \sum_{k=1}^S \sum_{\mathbf{y} \in \Omega^S | |\mathbf{y}|=k, y_i=1} (1 - z)^{k-1} z^{S-k} \prod_{\ell=1, \ell \neq i}^S d_{i, o}^{1-y_\ell} dz \\ &= \sum_{i=1}^S (1 - b_{i, o} - d_{i, o}) \sum_{k=1}^S \sum_{\mathbf{y} \in \Omega^S | |\mathbf{y}|=k, y_i=1} \frac{(k - 1)!(S - k)!}{S!} \prod_{\ell=1, \ell \neq i}^S d_{i, o}^{1-y_\ell} \\ &= \frac{1}{S} \sum_{i=1}^S (1 - b_{i, o} - d_{i, o}) \sum_{k=1}^S \sum_{\mathbf{y} \in \Omega^S | |\mathbf{y}|=k, y_i=1} \frac{1}{S-1 C_{k-1}} \prod_{\ell=1, \ell \neq i}^S d_{i, o}^{1-y_\ell} \end{aligned} \quad (\text{A30})$$

807 where ${}_{S-1}C_{k-1}$ (with $k = |\mathbf{y}|$) represents the binomial coefficient, which counts the number of ways, disre-
 808 garding order, that $k - 1$ species can be chosen from among $S - 1$ species (Van Lint & Wilson 2001). This is the

809 exact expression of the expectation of Jaccard dissimilarity. When we consider the conditional expectation,
 810 we divide the unconditional expectation by $1 - \prod_{i=1}^S d_{i,\circ}$. An alternative expression (displayed in the main
 811 text) is given by:

$$\mathbf{E}[\beta^J | \gamma > 0] = \frac{\sum_{i=1}^S \sum_{\mathbf{y} \in \Omega^S} \frac{1}{|\mathbf{y}|} \frac{y_i(1-b_{i,\circ}-d_{i,\circ})}{s_{C_{\mathbf{y}}}} \prod_{\ell(\neq i)}^S d_{\ell,\circ}^{1-y_\ell}}{1 - \prod_{k=1}^S d_{k,\circ}} \quad (\text{A31})$$

812 which follows by rearranging the binomial coefficients.

813 **Shortcut method using the generating function**

814 By noticing that integration is the key, one can take a shortcut approach. Let $\tau(z_b, z_u, z_d) := \prod_{i=1}^S (u_{i,\circ} z_u + b_{i,\circ} z_b + d_{i,\circ} z_d)$
 815 be the joint generating function of the ternary distributions for each species to be common, unique, or
 816 double-absent ($i = 1, 2, \dots, S$). By expanding the polynomial we get the identity:

$$\tau(z_b, z_u, z_d) \equiv \sum_{\{i_u+i_b+i_d=S\}} \tau_{(i_u, i_b, i_d)} z_u^{i_u} z_b^{i_b} z_d^{i_d}, \quad (\text{A32})$$

817 where the coefficient $\tau_{(i_u, i_b, i_d)}$ represents the probability that (i) i_u species are unique, (ii) i_b species are
 818 common, and (iii) $i_d = S - i_u - i_b$ species are double-absent.¹ A vector $\mathbf{i} := (i_u, i_b, i_d)$ (with $i_u + i_b + i_d = S$)
 819 therefore represents the state, or species-implicit incidence-table, of the community, with each species
 820 categorized as either unique, common, or double-absent. What we ought to compute is then:

$$\mathbf{E}[\beta^J] = \sum_{\mathbf{i}} \frac{i_u}{i_u + i_b} \tau_{\mathbf{i}}, \quad (\text{A33})$$

821 which equals:

$$\sum_{\mathbf{i}} \frac{i_u}{i_u + i_b} \tau_{\mathbf{i}} z_u^{i_u} z_b^{i_b} z_d^{i_d} \Bigg|_{z_u=z_b=z_d=1}. \quad (\text{A34})$$

822 We wish to algebraically extract i_u (the numerator of Jaccard dissimilarity) and the reciprocal of $(i_b + i_u)$
 823 (the denominator of Jaccard dissimilarity). With the inspiration of integration, we can come up with:

$$\mathbf{E}[\beta^J] = \int_0^1 \frac{\partial}{\partial z_u} \sum_{\mathbf{i}} \tau_{(i_u, i_b, i_d)} z_u^{i_u} z_b^{i_b} z_d^{i_d} \Bigg|_{z_u=z_b} dz_b \equiv \int_0^1 \frac{\partial}{\partial z_u} \tau(z_u, z_b, 1) \Bigg|_{z_u=z_b} dz_b. \quad (\text{A35})$$

824 Using the original definition of the generating function, its derivative is given by:

$$\begin{aligned} \frac{\partial}{\partial z_u} \tau(z_b, z_u, z_d) \Bigg|_{z_u=z_b} &= \frac{\partial}{\partial z_u} \prod_{i=1}^S (u_{i,\circ} z_u + b_{i,\circ} z_b + d_{i,\circ} z_d) \Bigg|_{z_u=z_b, z_d=1} \\ &= \sum_{i=1}^S u_{i,\circ} \prod_{\ell(\neq i)} (u_{\ell} z_b + b_{\ell} z_b + d_{\ell}) \\ &= \sum_{i=1}^S u_{i,\circ} \prod_{\ell(\neq i)} ((1 - d_{\ell,\circ}) z_b + d_{\ell,\circ}), \end{aligned} \quad (\text{A36})$$

825 which thus returns us back to Eqn (A22) by integrating the last line.

826 **Appendix B Approximations**

827 **Upper bound**

828 Although the exact calculation of the expectation is correct, the computational speed may be too slow to
 829 be practical, especially when S is large. Therefore we consider approximating it. We will make use of the

¹ z_d really is unneeded but is incorporated for symmetry.

830 property of the bivariate function:

$$J(b, d) := \frac{1 - b - d}{1 - d}, \quad (\text{B37})$$

831 which is “almost” linear, on the feasible domain $\{(b, d) \in [0, 1]^2 \mid \sqrt{b} + \sqrt{d} \leq 1\}$.

832 We rewrite Eqn (A22) as:

$$\int_0^1 \sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ}) \exp\left(\sum_{\ell=1, \ell \neq i}^S \log(1 - z + zd_{\ell,\circ})\right) dz. \quad (\text{B38})$$

833 Using Jensen’s inequality,

$$\sum_{\ell=1, \ell \neq i}^S \log(1 - z + zd_{\ell,\circ}) \leq (S - 1) \log\left(1 - z + z \frac{1}{S - 1} \sum_{\ell=1, \ell \neq i}^S d_{\ell,\circ}\right), \quad (\text{B39})$$

834 where the equality achieves when $d_{\ell,\circ}$ s are all identical across species $i = 1, 2, \dots, S$. Now put:

$$h_{\setminus i} := \frac{1}{S - 1} \sum_{\ell=1, \ell \neq i}^S d_{\ell,\circ}. \quad (\text{B40})$$

835 Then we get:

$$\mathbf{E}[\beta^J] \leq \int_0^1 \sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ}) (1 - z + zh_{\setminus i})^{S-1} dz = \frac{1}{S} \sum_{i=1}^S \frac{1 - b_{i,\circ} - d_{i,\circ}}{1 - h_{\setminus i}} (1 - h_{\setminus i}^S) = \beta_+^J. \quad (\text{B41})$$

836 RHS gives a very good approximation, because $J(b, d)$ is almost linear and only very moderately concave.

837 We divide both sides by $1 - \prod_{i=1}^S d_{i,\circ}$ to get the approximation of the conditional expectation of Jaccard
838 dissimilarity.

839 Lower bound

840 The lower-bound approximation of the expected Jaccard dissimilarity is given by:

$$\begin{aligned} \mathbf{E}[\beta^J] &= \sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ}) \int_0^1 \left(\prod_{\ell=1, \ell \neq i}^S (1 - z + zd_{\ell,\circ}) \right) dz \\ &\geq \sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ}) \int_0^1 \left(1 - z + z \prod_{\ell=1, \ell \neq i}^S d_{\ell,\circ}^{\frac{1}{S-1}} \right)^{S-1} dz \\ &= \frac{1}{S} \sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ}) \frac{1 - \prod_{\ell=1, \ell \neq i}^S d_{\ell,\circ}^{\frac{S-1}{S}}}{1 - \prod_{\ell=1, \ell \neq i}^S d_{\ell,\circ}^{\frac{1}{S-1}}} = \beta_-^J, \end{aligned} \quad (\text{B42})$$

841 where the second line follows by applying the induction, and the equality holds when $d_{i,\circ}$ s are all identical.

842 ² Dividing both sides by $1 - \prod_{i=1}^S d_{i,\circ}$ yields the approximation for the conditional expectation.

843 Heuristic approximation

844 The other approximation for the conditional expectation can be obtained heuristically (Ontiveros *et al.*
845 2021):

$$\beta_{\text{heur}}^J = \frac{\sum_{i=1}^S (1 - b_{i,\circ} - d_{i,\circ})}{\sum_{i=1}^S (1 - d_{i,\circ})} = J\left(\frac{1}{S} \sum_{i=1}^S b_{i,\circ}, \frac{1}{S} \sum_{i=1}^S d_{i,\circ}\right) \quad (\text{B43})$$

²First, when $S = 2$, some convexity arguments of arithmetic and geometric means work. Second, hypothesizing that the inequality is valid for a certain $S = 2^\ell$, it is easy to prove that the inequality holds for $S = 2^{\ell+1}$. Finally, hypothesizing that the inequality is valid for a certain S , we prove the inequality is the case for $S - 1$, by using the binomial expansion of the product and applying the arithmetic-geometric means relation repeatedly. The equality achieves when all $d_{i,\circ}$ are equal. This completes the proof.

846 which represents the expected number of unique species divided by the expected number of present species.
847 Deriving this formula requires quite a bit of calculations, but if we notice:

$$1 - \prod_{i=1}^S d_{i,\circ} = - \int_0^1 \frac{d}{dz} \prod_{i=1}^S (1 - z + zd_{i,\circ}) dz = \int_0^1 \sum_{i=1}^S (1 - d_{i,\circ}) \prod_{\ell=1; \ell \neq i}^S (1 - z + zd_{\ell,\circ}) dz, \quad (\text{B44})$$

848 then we get:

$$\begin{aligned} \mathbf{E}[\beta^J | \gamma > 0] &= \frac{\mathbf{E}[\beta^J]}{1 - \prod_{i=1}^S d_{i,\circ}} \\ &= \frac{\int_0^1 \sum_{i=1}^S (1 - d_{i,\circ} - b_{i,\circ}) \prod_{\ell=1; \ell \neq i}^S (1 - z + zd_{\ell,\circ}) dz}{\int_0^1 \sum_{i=1}^S (1 - d_{i,\circ}) \prod_{\ell=1; \ell \neq i}^S (1 - z + zd_{\ell,\circ}) dz} \\ &= \frac{\sum_{i=1}^S (1 - d_{i,\circ} - b_{i,\circ}) M_i}{\sum_{i=1}^S (1 - d_{i,\circ}) M_i}, \end{aligned} \quad (\text{B45})$$

849 where we have put:

$$M_i := \int_0^1 \prod_{\ell=1; \ell \neq i}^S (1 - z + zd_{\ell,\circ}) dz \quad (\text{B46})$$

850 for $i = 1, 2, \dots, S$. If we replace the integral M_i , which shows up in both the denominator and numerator (but
851 is multiplied by different coefficients), with (say) its average over $i = 1, 2, \dots, S$, written as $M_* := \sum_{i=1}^S M_i / S$,
852 then:

$$\mathbf{E}[\beta^J | \gamma > 0] \approx \frac{\sum_{i=1}^S (1 - d_{i,\circ} - b_{i,\circ}) M_*}{\sum_{i=1}^S (1 - d_{i,\circ}) M_*} = \frac{\sum_{i=1}^S (1 - d_{i,\circ} - b_{i,\circ})}{\sum_{i=1}^S (1 - d_{i,\circ})} = \beta_{\text{heur}}^J, \quad (\text{B47})$$

853 which thus gives the heuristic approximation. The approximation is exact when $M_1 = \dots = M_S$ (which
854 is not a necessary condition). This explains why the heuristic approximation works for the conditional
855 expectation, while not for the unconditional, and also explains why the heuristic approximation may
856 be larger or smaller than the exact expectation depending on the variance in M_i s. We may observe that
857 M_i becomes increasingly small with S larger ($M_* = \mathcal{O}\left(\left(\sum_{i=1}^S d_{i,\circ}\right)^{-1}\right)$), and thus, as S increases, the
858 contribution of the replacement (M_i with M_*) to the difference between the exact and approximated
859 expectation becomes smaller.

860 The heuristic approximation may, however, be either larger or smaller than the exact conditional
861 expectation, and thus the upper and lower bounds, (β_+^J and β_-^J), both may be also recommended.

862 Altogether, we obtained the three approximations:

- 863 ○ Approximation from below: β_-^J
- 864 ○ Approximation from above: β_+^J
- 865 ○ Heuristic approximation: β_{heur}^J

866 Performances

867 We numerically compared the accuracy and precision of the approximations, in the following procedure.

868 (1) Generate two vectors (with $S = 100$):

$$\mathbf{p}_{\circ,1} = \begin{pmatrix} p_{1,1} \\ p_{2,1} \\ \vdots \\ p_{S,1} \end{pmatrix}, \mathbf{p}_{\circ,2} = \begin{pmatrix} p_{1,2} \\ p_{2,2} \\ \vdots \\ p_{S,2} \end{pmatrix}, \quad (\text{B48})$$

869 in which each value of $p_{i,1}$ is drawn from the Beta distribution with parameters 1.2 and 1.5, and $p_{i,2}$
870 drawn from the Beta distribution with parameters 1.6 and 0.8. We chose the Beta distribution to
871 generate the presence probabilities because it is the conjugate distribution of the well-known binomial
872 distribution, thus an appropriate distribution to generate probabilities of ‘success (presence)’.

873 (2) Calculate double-presence and double-absence vectors:

$$\mathbf{b}_{o,o} = \begin{pmatrix} b_{1,o} \\ b_{2,o} \\ \vdots \\ b_{S,o} \end{pmatrix} = \begin{pmatrix} p_{1,1}p_{1,2} \\ p_{2,1}p_{2,2} \\ \vdots \\ p_{S,1}p_{S,2} \end{pmatrix}, \mathbf{d}_{o,o} = \begin{pmatrix} d_{1,o} \\ d_{2,o} \\ \vdots \\ d_{S,o} \end{pmatrix} = \begin{pmatrix} a_{1,1}a_{1,2} \\ a_{2,1}a_{2,2} \\ \vdots \\ a_{S,1}a_{S,2} \end{pmatrix}, \quad (\text{B49})$$

874 which can be numerically implemented with Hadamard product.

875 (3) Compute the followings:

- 876 · $\mathbf{E}[\beta^j | \gamma > 0]$ using Eqn (A22);
- 877 · β_-^j in a brute-force approach;
- 878 · β_+^j in a brute-force approach;
- 879 · β_{heur}^j in a brute-force approach.

880 (4) Repeat the procedure 1000 times and generated the probability distributions of those four quantities
881 (but note that they are stochastic variables because we used the Beta distribution to generate $\{p_{i,j}\}$).

882 (5) We plot the distribution of the expectations of Jaccard dissimilarity measures (approximations and
883 exact value).

884 Data to generate Figure 3

885 The following $\mathbf{p}_{o,j}$ are used to generate Figure 3 of the main text.

$$\begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{pmatrix} \quad (w, h) \quad \left(\mathbf{E}[\beta^j | \gamma > 0], \beta_{\text{heur}}^j \right)$$

$$\begin{pmatrix} \{0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5\} \\ \{0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5\} \end{pmatrix} \quad (0, 0) \quad (0.666667, 0.666667)$$

$$\begin{pmatrix} \{0.432205, 0.43587, 0.450907, 0.481017, 0.518983, 0.549093, 0.56413, 0.567795\} \\ \{0.432205, 0.43587, 0.450907, 0.481017, 0.518983, 0.549093, 0.56413, 0.567795\} \end{pmatrix} \quad (0, 0.1) \quad (0.661062, 0.661545)$$

$$\begin{pmatrix} \{0.364411, 0.37174, 0.401815, 0.462035, 0.537965, 0.598185, 0.62826, 0.635589\} \\ \{0.364411, 0.37174, 0.401815, 0.462035, 0.537965, 0.598185, 0.62826, 0.635589\} \end{pmatrix} \quad (0, 0.2) \quad (0.644037, 0.645942)$$

$$\begin{pmatrix} \{0.296616, 0.30761, 0.352722, 0.443052, 0.556948, 0.647278, 0.69239, 0.703384\} \\ \{0.296616, 0.30761, 0.352722, 0.443052, 0.556948, 0.647278, 0.69239, 0.703384\} \end{pmatrix} \quad (0, 0.3) \quad (0.614943, 0.619112)$$

(B50)

$$\begin{pmatrix} \begin{pmatrix} \{0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55\} \\ \{0.45, 0.45, 0.45, 0.45, 0.45, 0.45, 0.45\} \end{pmatrix} & (0.1, 0) & (0.671096, 0.671096) \\ \begin{pmatrix} \{0.453604, 0.453752, 0.458416, 0.499943, 0.600057, 0.641584, 0.646248, 0.646396\} \\ \{0.332182, 0.332364, 0.338064, 0.388819, 0.511181, 0.561936, 0.567636, 0.567818\} \end{pmatrix} & (0.1, 0.1) & (0.653477, 0.655013) \\ \begin{pmatrix} \{0.357207, 0.357505, 0.366832, 0.449885, 0.650115, 0.733168, 0.742495, 0.742793\} \\ \{0.214364, 0.214728, 0.226128, 0.327637, 0.572363, 0.673872, 0.685272, 0.685636\} \end{pmatrix} & (0.1, 0.2) & (0.59844, 0.604339) \end{pmatrix}$$

(B51)

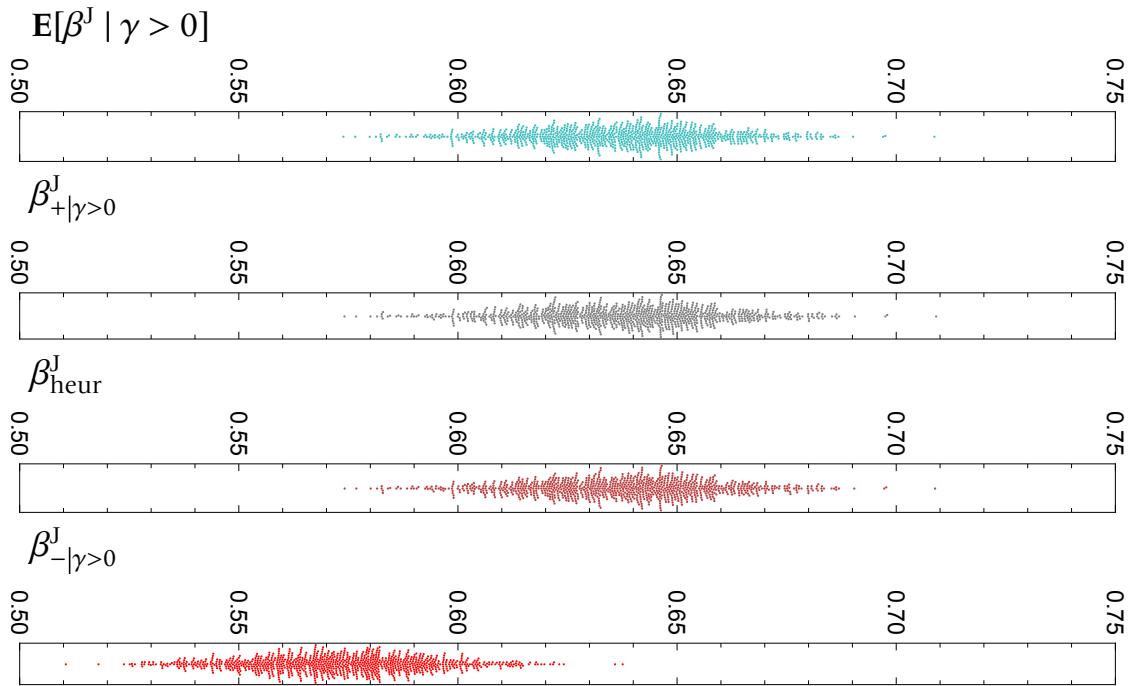
$$\begin{pmatrix} \begin{pmatrix} \{0.675, 0.675, 0.675, 0.675, 0.675, 0.675, 0.675, 0.675\} \\ \{0.325, 0.325, 0.325, 0.325, 0.325, 0.325, 0.325, 0.325\} \end{pmatrix} & (0.35, 0) & (0.718975, 0.718975) \\ \begin{pmatrix} \{0.74462, 0.744512, 0.741144, 0.711153, 0.638847, 0.608856, 0.605488, 0.60538\} \\ \{0.469595, 0.469372, 0.462376, 0.400086, 0.249914, 0.187624, 0.180628, 0.180405\} \end{pmatrix} & (0.35, 0.1) & (0.704116, 0.705752) \\ \begin{pmatrix} \{0.814239, 0.814025, 0.807288, 0.747305, 0.602695, 0.542712, 0.535975, 0.535761\} \\ \{0.614189, 0.613743, 0.599752, 0.475172, 0.174828, 0.0502477, 0.0362568, 0.0358108\} \end{pmatrix} & (0.35, 0.2) & (0.657895, 0.66439) \end{pmatrix}$$

(B52)

$$\left(\begin{array}{ccc}
 \left(\begin{array}{c} \{0.675, 0.675, 0.675, 0.675, 0.675, 0.675, 0.675, 0.675\} \\ \{0.325, 0.325, 0.325, 0.325, 0.325, 0.325, 0.325, 0.325\} \end{array} \right) & \left(\begin{array}{c} 0.35, 0. \\ \end{array} \right) & \left(\begin{array}{c} 0.718975, 0.718975 \\ \end{array} \right) \\
 \left(\begin{array}{c} \{0.719067, 0.716685, 0.70691, 0.687339, 0.662661, 0.64309, 0.633315, 0.630933\} \\ \{0.233477, 0.238424, 0.258725, 0.299374, 0.350626, 0.391275, 0.411576, 0.416523\} \end{array} \right) & \left(\begin{array}{c} 0.35, 0.1 \\ \end{array} \right) & \left(\begin{array}{c} 0.723095, 0.723094 \\ \end{array} \right) \\
 \left(\begin{array}{c} \{0.763133, 0.758369, 0.738821, 0.699677, 0.650323, 0.611179, 0.591631, 0.586867\} \\ \{0.141954, 0.151849, 0.19245, 0.273747, 0.376253, 0.45755, 0.498151, 0.508046\} \end{array} \right) & \left(\begin{array}{c} 0.35, 0.2 \\ \end{array} \right) & \left(\begin{array}{c} 0.735304, 0.735295 \\ \end{array} \right) \\
 \left(\begin{array}{c} \{0.8072, 0.800054, 0.770731, 0.712016, 0.637984, 0.579269, 0.549946, 0.5428\} \\ \{0.0504315, 0.0652731, 0.126175, 0.248121, 0.401879, 0.523825, 0.584727, 0.599568\} \end{array} \right) & \left(\begin{array}{c} 0.372887, 0.3 \\ \end{array} \right) & \left(\begin{array}{c} 0.755166, 0.755118 \\ \end{array} \right)
 \end{array} \right)$$

(B53)

886



SI Figure 1: We assessed the probability distributions of the exact and approximated Jaccard expectations.

887 Appendix C Well-definedness

888 Here we clarify some concepts that have been left unspecified in the main text.

889 Species pool

890 We first suppose that species i is *presentable* in the focal metacommunity if strictly $d_{i,o} < 1$. The collection
 891 of all presentable species is called as *species pool*. Its cardinality (i.e., the number of members of the species
 892 pool) is referred to as *species pool size* and denoted as S . By defining these, we exclude never presentable
 893 species from the species pool, $p_{i,1} = p_{i,2} = 0$.

Computation of $E[\text{Jaccard} \mid \gamma > 0]$

We formulate a Mathematica code for:

(i) the conditional expectation of Jaccard dissimilarity, denoted `jaccard[{p1, p2}]` with argument a pair of species presence probabilities $\mathbf{p1} := (p11, p21, p31, \dots, pS1)$ and $\mathbf{p2} := (p12, p22, p32, \dots, pS2)$, and

(ii) the heuristic approximation, denoted `heur[{p1, p2}]`.

The species pool size (which is written `shu`) is arbitrary and can be determined by directly defining the length of `p`'s.

```
In[3]:= Clear["Global`*"];
SetDirectory[NotebookDirectory[]];

In[5]:= jaccard[{p1_List, p2_List}] := With[{},
  shu = Length@p1;
  b1list = p1 * p2;
  d1list = (1 - p1) * (1 - p2);
  u1list = 1 - b1list - d1list;
  j = Total@Table[u1list[[i]] * (*N*) Integrate[
    
$$\frac{1}{1 - z + z * d1list[[i]]} * \text{Product}[1 - z + z * d1list[[k]], \{k, 1, shu, 1\}], \{z, 0, 1\}],
    \{i, 1, shu, 1\}] / (1 - \text{Product}[d1list[[i]], \{i, 1, shu, 1\}]);
  heur[{p1_List, p2_List}] := 
$$\frac{\text{Mean}[(p1) * (1 - p2) + (1 - p1) * (p2)]}{\text{Mean}[1 - (1 - p1) * (1 - p2)]}$$
;$$

```

For example, when $\mathbf{p1} = (p, p, p, p)$ and $\mathbf{p2} = (q, q, q, q)$:

```
In[24]:= even1 = Table[p, {dummy, 1, 4, 1}]; even2 = Table[q, {dummy, 1, 4, 1}];
jaccard[{even1, even2}] // Simplify
% == heur[{even1, even2}] // Simplify
```

```
Out[25]= 
$$\frac{p + q - 2 p q}{p + q - p q}$$

```

```
Out[26]= True
```

For another example, Fig 3B-2 bottom right panel can be recovered:

```
In[230]:= p1b = {0.763133, 0.758369, 0.738821,
  0.699677, 0.650323, 0.611179, 0.591631, 0.586867};
p2b =
  {0.141954, 0.151849, 0.19245, 0.273747, 0.376253, 0.45755, 0.498151, 0.508046};
{jaccard@({p1b, p2b}), heur@({p1b, p2b})}

Out[232]= {0.735304, 0.735295}
```

SI Figure 2: Mathematica code to compute the exact and approximated Jaccard dissimilarity expectation conditioned on $\gamma > 0$.

894 Intuitive interpretation of species difference

895 We now explain the intuition behind the species difference. To do so, we start by considering the species
896 difference for a single-site case, which is also known as the Hoover index (Hoover 1936):

$$\begin{aligned} \text{(Species difference in site 1)} &:= \omega_1 \\ &:= \frac{1}{S\mu_1} \sum_{i=1}^S |p_{i,1} - \mu_1| \\ &= \sum_{i=1}^S \left| \frac{p_{i,1}}{S\mu_1} - \frac{1}{S} \right|. \end{aligned} \quad (\text{C54})$$

897 Noticing that $p_{i,1} / (S\mu_1)$ is normalized and thus sums to unity, as well as that $1 / S$ represents the discrete
898 uniform distribution, we can see that the Hoover index represents the distance from the uniform distribu-
899 tion, meaning that the species difference in site 1 measure how far the distribution of relative presence
900 probabilities deviates from the uniform distribution (i.e., from the situation where no species difference
901 exists). We can compute a similar quantity for site 2.

902 For a two-sites case,

$$w = \frac{\mu_1}{\mu_1 + \mu_2} \omega_1 + \frac{\mu_2}{\mu_1 + \mu_2} \omega_2, \quad (\text{C55})$$

903 which is the weighted average of within-site species difference; here, the weight is given by the relative
904 value of average presence probabilities, thus giving more weight to a site that has a higher average presence
905 probability. Therefore, the species difference overall represents the deviation from the case where all
906 species have the same presence probability in each site, weighted by the relative value of average presence
907 probabilities.

908 Bounds of w and h

909 species difference

910 The species difference index is Schur-convex (Arnold 2012; McVinish & Lester 2020), thus taking the
911 minimum zero when all species are equally likely to be present in each site; that is whenever:

$$p_{1,j} = \dots = p_{S,j} > 0 \quad (\text{C56})$$

912 for $j = 1, 2$.

913 To determine an upper bound (if any), we conventionally assume that for all species i , there is a
914 minimum value of the presence probabilities $p_{i,j} \geq \varepsilon$. Then by the property of Schur-convexity of Hoover
915 index, the largest difference (inequality) occurs when:

$$\begin{aligned} p_{1,1} &= 1, & p_{i,1} &\equiv \varepsilon > 0, \forall i \geq 2, \\ p_{1,2} &= 1, & p_{i,2} &\equiv \varepsilon > 0, \forall i \geq 2, \end{aligned} \quad (\text{C57})$$

916 with species-wise permutation permitted for each j . Substituting this into w gives the maximum value,
917 which reads:

$$\max \{ w \} = \frac{2(1-\varepsilon)(S-1)}{S(\varepsilon(S-1)+1)} < 2. \quad (\text{C58})$$

918 Spatial heterogeneity

919 If and only if $p_{i,1} \equiv p_{i,2}$ for each i , spatial heterogeneity h achieves the minimum of zero. If and only if
920 $|p_{i,1} - p_{i,2}| = 1 - \varepsilon$, spatial heterogeneity h achieves the maximum of $1 - \varepsilon$.

921 **Axioms**

922 Species difference w has to satisfy a number of axioms in order to be a “difference” measure among species,
923 as does h in order to measure heterogeneity between two sites. We adopt Routledge’s (1983), Jost’s (2006,
924 2007), Chao & Ricotta’s (2019), and Leinster’s (2021) approach to develop appropriate indices. Note that the
925 presence probabilities do not represent abundances.

926 requirement 1 The first requirement for w is that, if, and only if, $p_{i,1}$ s are all equal and $p_{i,2}$ are all equal,
927 then $w = 0$.

928 requirement 2 Second, w is Schur-convex (or Schur-increasing; Marshall *et al.* 1979, Chapter 1).

929 requirement 3 The requirement for heterogeneity h is that it is a distance function between $\mathbf{p}_{o,1}$ and $\mathbf{p}_{o,2}$.

930 **Feasible ranges of spatial heterogeneity and species difference**

931 We finally examine the possible regions for (w, h) to understand the degree to which the two parameters
932 co-vary, and thereby determine the region for which they are ‘unrelated’ sensu Chao & Ricotta (2019) and
933 Leinster (2021). To describe various species-incidence curves, we use the lower- and upper-bounded Hill
934 equation, which in a generic form reads:

$$935 \quad p_{i,j} = \rho_j + (1 - \rho_j) \varphi_j \frac{\chi_j (\lambda_j (i - 1))^m + (1 - \chi_j) (1 - \lambda_j)^m (S - i)^m}{(\lambda_j (i - 1))^m + (1 - \lambda_j)^m (S - i)^m} \quad (C59)$$

936 where $\chi_j, \rho_j, \varphi_j, \lambda_j \in [0, 1]$ are all constants: χ_j represents the increase versus a decrease of presence proba-
937 bility with species labels (with $\chi = 0.50$ yielding flat curves); ρ_j represents the minimum value of presence
938 probability; φ_j represents the importance of the Hill function (the fraction); m represents the steepness
939 of the Hill function; λ_j represents the position at abscissa beyond which We then tuned these parameters
940 simultaneously and plotted the resulting (w, h) . We found that species difference and spatial heterogeneity
940 are related with each other, i.e., a value of the former restricts the range of the latter.

941 **Ecological meaning of the parameters**

942 Finally, we clarify the ecological meaning of the spatial heterogeneity and species difference parameters. In
943 the classic colonization-extinction model of island biogeography, the presence probability is determined
944 by species-specific, site-specific parameters of colonization and extinction rates. Mathematically, the
945 presence probability is a continuous function of colonization (increasing) and extinction rates (decreasing).
946 Therefore, the smaller the species difference is, the closer the colonization rates of species are (and/or
947 extinction rates). More generally, we assume that the presence probability is a continuous function of
948 species traits and environmental variables. This assumption does not necessarily mean that all variables
949 should be quantitative; qualitative (categorical) variables are also allowed so long as we can define a metric
950 (distance) in the corresponding trait space. For example, dispersal capacity measured with the presence or
951 absence of wings has a binary metric.

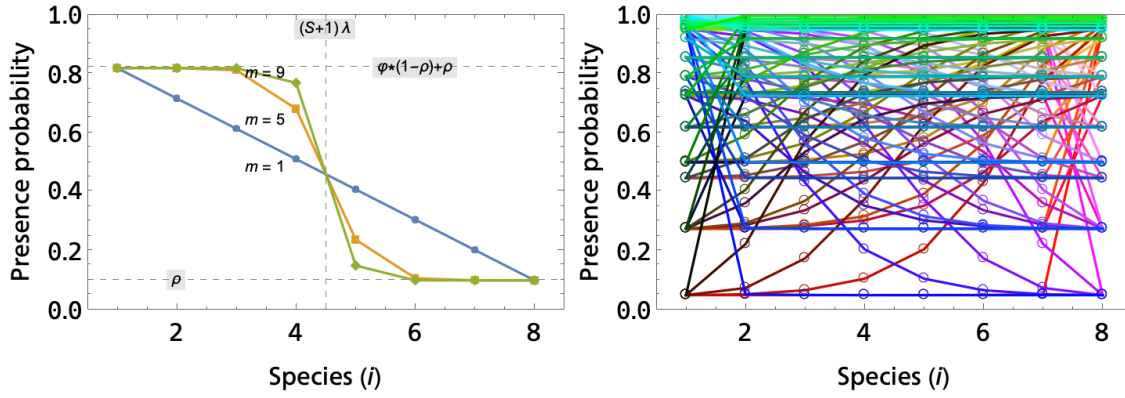
952 **Appendix D Variance of Jaccard dissimilarity**

953 **Same method as the mean**

954 To compute the variance, we use the identity for a pair of positive quantities $v_1, v_2 > 0$:

$$955 \quad \left(\frac{v_1}{v_2} \right)^2 \equiv - \int_0^{+\infty} \frac{\partial^3}{\partial \xi \partial \eta^2} \exp(v_1 \eta - (v_2 \theta + \xi \theta)) d\theta \Bigg|_{\xi=\eta=0}. \quad (D60)$$

956 One may preferably differentiate the quantity before integration (otherwise, erroneous calculation is
956 possible).



SI Figure 3: Incidence-curve generators using Eqn (C59). (A) The scaled Hill function with $\chi = 0.45$, $\lambda = 0.5$, $\rho = 0.1$, $\varphi = 0.8$, and $m = 1, 5, 9$ in Eqn (C59). (B) A series of generated incidence curves using Eqn (C59). The same scheme is used for both sites and all possible pairs of curves are considered. Fixed parameter values: $\varphi = 0.95$, $m = 2$; varied parameter ranges: $\lambda \in \{0.01, 0.255, 0.500, 0.745, 0.990\}$, $\rho \in \{0.05, 0.275, 0.500, 0.725, 0.950\}$, $\sigma \in \{0.00, 0.25, 0.50, 0.75, 1.0\}$; colors are tuned by RGB color scheme $[1 - \lambda, \rho, 1 - \sigma]$ with half opacity.

957 For Jaccard dissimilarity, we choose $v_1 = \sum_{i=1}^S (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2})$, which represents the number of
 958 unique species, and $v_2 = \sum_{i=1}^S (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2})$, which represents the number of present species (gamma
 959 diversity). That is:

$$\beta_{\mathbf{X}}^{J^2} = - \int_0^{+\infty} \frac{\partial^3}{\partial \xi \partial \eta^2} \exp \left(\eta \sum_{i=1}^S (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) \right) \exp \left(-\theta \sum_{i=1}^S (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) \right) d\theta \Big|_{\xi=\eta=0}. \quad (D61)$$

960 The expectation of $\beta_{\mathbf{X}}^{J^2}$ is given by:

$$\mathbf{E} \left[\beta^{J^2} \right] = - \int_0^{+\infty} \frac{\partial^3}{\partial \xi \partial \eta^2} e^{-\xi \theta} \prod_{i=1}^S \sum_{\mathbf{x}_{i,o}} P_{\mathbf{X}} \exp \left(\eta (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) - \theta (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) \right) d\theta \Big|_{\xi=\eta=0} \quad (D62)$$

961 By evaluating the Boolean variable,

$$\eta (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) - \theta (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) = \begin{cases} 0, & \mathbf{x}_{i,o} = (0, 0); \\ \eta - \theta, & \mathbf{x}_{i,o} = (0, 1); \\ \eta - \theta, & \mathbf{x}_{i,o} = (1, 0); \\ -\theta, & \mathbf{x}_{i,o} = (1, 1); \end{cases} \quad (D63)$$

962 the resulting expression reads:

$$\mathbf{E} \left[\beta^{J^2} \right] = - \int_0^{+\infty} \frac{\partial^3}{\partial \xi \partial \eta^2} e^{-\xi \theta} \prod_{i=1}^S \left(d_{i,o} + b_{i,o} e^{-\theta} + (1 - d_{i,o} - b_{i,o}) e^{\eta - \theta} \right) d\theta \Big|_{\xi=\eta=0}. \quad (D64)$$

963 This is the most general expression for the second moment of the Jccard dissimilarity. For brevity we write
 964 $u_{i,o} := 1 - b_{i,o} - d_{i,o}$ for the probability that species i is unique (neither double-absent nor common); also, we
 965 write $\tau_{\ell}(\theta) := d_{\ell} + (1 - d_{\ell}) e^{-\theta}$ for the moment generating function of the probability that species ℓ is present
 966 in at least one of the sites, $1 - d_{\ell,o}$; write $\psi_i(\theta, \eta) := d_{i,o} + b_{i,o} e^{-\theta} + u_{i,o} e^{\eta - \theta}$, thus with $\psi_i(\theta, 0) \equiv \tau_i(\theta)$.

967 Leibniz rule for the second η -derivatives is given by:

$$\frac{\partial^2}{\partial \eta^2} \prod_{i=1}^S \psi_i(\theta, \eta) \Big|_{\eta=0} = \left(\sum_{i=1}^S u_{i,o} e^{-\theta} \prod_{\ell(\neq i)} \tau_{\ell}(\theta) \right) + \sum_{i=1}^S u_{i,o} e^{-\theta} \sum_{\substack{k=1; \\ k \neq i}}^S u_k e^{-\theta} \prod_{\substack{\ell=1; \\ \ell \neq i, k}}^S \tau_{\ell}(\theta), \quad (D65)$$

968 using which we get:

$$\mathbf{E}[\beta^J]^2 = -\frac{\partial}{\partial \xi} \int_0^{+\infty} e^{-\xi\theta} \left(\sum_{i=1}^S u_{i,\circ} e^{-\theta} \prod_{\ell(\neq i)}^S \tau_\ell(\theta) + \sum_{i=1}^S u_{i,\circ} e^{-\theta} \sum_{\substack{k=1; \\ k \neq i}}^S u_k e^{-\theta} \prod_{\substack{\ell=1; \\ \ell \neq i, k}}^S \tau_\ell(\theta) \right) d\theta. \quad (\text{D66})$$

969 We can evaluate this integral as did we before. However, the resulting equation is heavily complicated
970 (involving, e.g., the harmonic numbers) and computationally expensive.

971 **Approximating variance using Hubbard-Stratonovich transformation**

972 Here, we take a different approach to evaluate the variance. We use the identity:

$$\frac{v_1^2}{v_2^2} \equiv -\frac{\partial}{\partial \xi} \int_0^\infty e^{-(v_2 + v_1^2 \xi)\theta} d\theta \Big|_{\xi=0} \quad (\text{D67})$$

973 for $0 \leq v_1 \leq v_2$, as well as the Hubbard-Stratonovich transformation (Hubbard 1959):

$$e^{-\xi\theta\lambda^2} \equiv \frac{1}{\sqrt{4\pi\xi\theta}} \int_{-\infty}^\infty e^{-\frac{\zeta^2}{4\xi\theta} - i\lambda\zeta} d\zeta \quad (\text{D68})$$

974 where i represents the imaginary unit. Combining the identities gives:

$$\begin{aligned} (\beta^J)^2 &= \left(\frac{\sum_{i=1}^S x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}}{\sum_{i=1}^S x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}} \right)^2 \\ &\stackrel{\text{Eqn (D67)}}{=} \int_0^\infty \exp \left(- \left(\sum_{i=1}^S (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) + \xi \left(\sum_{i=1}^S x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2} \right)^2 \right) \theta \right) d\theta \\ &= - \int_0^\infty d\theta \int_{-\infty}^\infty d\zeta \exp \left(-\theta \left(\sum_{i=1}^S (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) \right) \right) \frac{1}{\sqrt{4\pi\xi\theta}} \exp \left(-\frac{\zeta^2}{4\xi\theta} - i\zeta \sum_{i=1}^S (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) \right) \end{aligned} \quad (\text{D69})$$

975 Let us evaluate the Boolean variable:

$$-\theta (x_{i,1} + x_{i,2} - x_{i,1}x_{i,2}) - i\zeta (x_{i,1} + x_{i,2} - 2x_{i,1}x_{i,2}) = \begin{cases} 0 & x_{i,1} = x_{i,2} = 0; \\ -\theta - i\zeta & x_{i,1} + x_{i,2} = 1; \\ -\theta & x_{i,1} = x_{i,2} = 1; \end{cases} \quad (\text{D70})$$

976 then we get:

$$\mathbf{E}[(\beta^J)^2] = -\frac{\partial}{\partial \xi} \int_0^\infty d\theta \int_{-\infty}^\infty d\zeta e^{-\frac{\zeta^2}{4\xi\theta}} \frac{1}{\sqrt{4\pi\xi\theta}} \prod_{i=1}^S \left(d_{i,\circ} + (1 - d_{i,\circ} - b_{i,\circ}) e^{-\theta - i\zeta} + b_{i,\circ} e^{-\theta} \right). \quad (\text{D71})$$

977 If we approximate the product as:

$$\prod_{i=1}^S \left(d_{i,\circ} + (1 - d_{i,\circ} - b_{i,\circ}) e^{-\theta - i\zeta} + b_{i,\circ} e^{-\theta} \right) \approx \left(d_{*,\circ} + u_{*,\circ} e^{-\theta - i\zeta} + b_{*,\circ} e^{-\theta} \right)^S, \quad (\text{D72})$$

978 where the *-subscripted quantities are the arithmetic means, over $i \in \{1, \dots, S\}$, of the corresponding
979 quantities, i.e. $d_{*,\circ} := \frac{1}{S} \sum_{i=1}^S d_{i,\circ}$, $b_{*,\circ} := \frac{1}{S} \sum_{i=1}^S b_{i,\circ}$, and $u_{*,\circ} := 1 - d_{*,\circ} - b_{*,\circ}$, then the expected value is

980 approximated by

$$\mathbf{E}\left[(\beta^J)^2\right] \approx -\frac{\partial}{\partial \xi} \int_0^\infty d\theta \int_{-\infty}^\infty d\zeta e^{-\frac{\zeta^2}{4\xi\theta}} \frac{1}{\sqrt{4\pi\xi\theta}} \left(d_{*,o} + u_{*,o}e^{-\theta-i\zeta} + b_{*,o}e^{-\theta}\right)^S \quad (\text{D73})$$

981 evaluated at $\xi = 0$.

982 Interchanging the order of the derivative and the double integral, we get

$$\mathbf{E}\left[(\beta^J)^2\right] \approx -\int_0^\infty d\theta \int_{-\infty}^\infty d\zeta \frac{\partial}{\partial \xi} \left(\frac{1}{\sqrt{4\pi\xi\theta}} e^{-\frac{\zeta^2}{4\xi\theta}} \right) e^{S \log(d_{*,o} + u_{*,o}e^{-\theta-i\zeta} + b_{*,o}e^{-\theta})}. \quad (\text{D74})$$

983 In the limit $\xi \rightarrow 0$, the function $\frac{\partial}{\partial \xi} \left(\frac{1}{\sqrt{4\pi\xi\theta}} e^{-\frac{\zeta^2}{4\xi\theta}} \right) = \frac{\zeta^2 - 2\xi\theta}{8\sqrt{\pi}(\xi\theta)^{5/2}} e^{-\frac{\zeta^2}{4\xi\theta}}$ is very peaked about $\zeta = 0$. Therefore,
 984 we expect the integrand to be nicely approximated if we substitute the logarithm by its series expansion
 985 about $\zeta = 0$,

$$\log\left(\kappa_1 + \kappa_2 e^{-i\zeta}\right) \approx \log(\kappa_1 + \kappa_2) - i \frac{b\zeta}{\kappa_1 + \kappa_2} - \frac{\kappa_1 \kappa_2 \zeta^2}{2(\kappa_1 + \kappa_2)^2} + \mathcal{O}(\zeta^3), \quad (\text{D75})$$

986 with $\kappa_1 := d_{*,o} + b_{*,o}e^{-\theta}$ and $\kappa_2 := u_{*,o}e^{-\theta}$. Inserting this second approximation into Eqn (D74) we get

$$\mathbf{E}\left[(\beta^J)^2\right] \approx -\int_0^\infty d\theta \left(d_{*,o} + (1 - d_{*,o})e^{-\theta}\right)^S \frac{\partial}{\partial \xi} \frac{1}{\sqrt{4\pi\xi\theta}} \int_{-\infty}^\infty d\zeta e^{-\frac{\zeta^2}{4\xi\theta} - i \frac{S\kappa_2\zeta}{\kappa_1 + \kappa_2} - \frac{S\kappa_1\kappa_2\zeta^2}{2(\kappa_1 + \kappa_2)^2}}, \quad (\text{D76})$$

987 which, again, has to be evaluated at $\xi = 0$. The integral over ζ can be evaluated as

$$\frac{1}{\sqrt{4\pi\xi\theta}} \int_{-\infty}^\infty d\zeta e^{-\frac{\zeta^2}{4\xi\theta} - i \frac{S\kappa_2\zeta}{\kappa_1 + \kappa_2} - \frac{S\kappa_1\kappa_2\zeta^2}{2(\kappa_1 + \kappa_2)^2}} = \frac{\kappa_1 + \kappa_2}{\sqrt{(\kappa_1 + \kappa_2)^2 + 2S\kappa_1\kappa_2\xi\theta}} e^{-\frac{\xi\theta(\kappa_2 S)^2}{(\kappa_1 + \kappa_2)^2 + 2S\kappa_1\kappa_2\xi\theta}}. \quad (\text{D77})$$

988 Now, we can take the derivative with respect to ξ and evaluate it at $\xi = 0$ to get

$$\frac{\partial}{\partial \xi} \frac{1}{\sqrt{4\pi\xi\theta}} \int_{-\infty}^\infty d\zeta e^{-\frac{\zeta^2}{4\xi\theta} - i \frac{S\kappa_2\zeta}{\kappa_1 + \kappa_2} - \frac{S\kappa_1\kappa_2\zeta^2}{2(\kappa_1 + \kappa_2)^2}} \Bigg|_{\xi=0} = -\frac{S\kappa_2\theta(\kappa_1 + \kappa_2 S)}{(\kappa_1 + \kappa_2)^2}. \quad (\text{D78})$$

989 Therefore, inserting this expression into Eqn (D76) and replacing κ_1 and κ_2 by their expressions in terms
 990 of $d_{*,o}$, $b_{*,o}$, $u_{*,o}$, and θ , we obtain

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} \int_0^\infty d\theta e^{-\theta} \left(d_{*,o} + b_{*,o}e^{-\theta} + Su_{*,o}e^{-\theta}\right) \left(d_{*,o} + (1 - d_{*,o})e^{-\theta}\right)^{S-2} \theta. \quad (\text{D79})$$

991 Changing to the variable $z = e^{-\theta}$ yields

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} \int_0^1 dz (-\log z) \left(d_{*,o} + b_{*,o}z + Su_{*,o}z\right) \left(d_{*,o} + (1 - d_{*,o})z\right)^{S-2}. \quad (\text{D80})$$

992 We now use the binomial expansion $\left(d_{*,o} + (1 - d_{*,o})z\right)^{S-2} = d_{*,o}^{S-2} \sum_{k=0}^{S-2} \binom{S-2}{k} \left(\frac{1-d_{*,o}}{d_{*,o}}\right)^k z^k$ to get

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} d_{*,o}^{S-2} \sum_{k=0}^{S-2} \binom{S-2}{k} \left(\frac{1-d_{*,o}}{d_{*,o}}\right)^k \int_0^1 dz (-\log z) \left(d_{*,o} + b_{*,o}z + Su_{*,o}z\right) z^k, \quad (\text{D81})$$

993 which, upon evaluation of the integral, yields

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} d_{*,o}^{S-2} \sum_{k=0}^{S-2} \binom{S-2}{k} \left(\frac{1-d_{*,o}}{d_{*,o}}\right)^k \left(\frac{d_{*,o}}{(k+1)^2} + \frac{b_{*,o} + Su_{*,o}}{(k+2)^2} \right). \quad (\text{D82})$$

994 The sum above can be expressed in terms of generalized hypergeometric functions ${}_pF_q(\{A\}, \{B\}; Z)$ as

$$\begin{aligned} \mathbf{E}\left[(\beta^J)^2\right] &\approx \frac{u_{*,o}(b_{*,o} + Su_{*,o})(1 - d_{*,o}^S)}{(S-1)(1 - d_{*,o})^2} - \frac{Su_{*,o}d_{*,o}^{S-1}(b_{*,o} + Su_{*,o})}{(S-1)(1 - d_{*,o})} {}_3F_2\left(\{1, 1, 1 - S\}, \{2, 2\}; 1 - \frac{1}{d_{*,o}}\right) \\ &+ Su_{*,o}d_{*,o}^{S-1} {}_3F_2\left(\{1, 1, 2 - S\}, \{2, 2\}; 1 - \frac{1}{d_{*,o}}\right). \end{aligned} \quad (\text{D83})$$

995 As a consequence, we find the following approximation for the variance,

$$\begin{aligned} \mathbf{V}[\beta^J] &\approx \frac{u_{*,o}(b_{*,o} + Su_{*,o})(1 - d_{*,o}^S)}{(S-1)(1 - d_{*,o})^2} - \frac{Su_{*,o}d_{*,o}^{S-1}(b_{*,o} + Su_{*,o})}{(S-1)(1 - d_{*,o})} {}_3F_2\left(\{1, 1, 1 - S\}, \{2, 2\}; 1 - \frac{1}{d_{*,o}}\right) \\ &+ Su_{*,o}d_{*,o}^{S-1} {}_3F_2\left(\{1, 1, 2 - S\}, \{2, 2\}; 1 - \frac{1}{d_{*,o}}\right) - \left(\frac{u_{*,o}}{1 - d_{*,o}}\right)^2, \end{aligned} \quad (\text{D84})$$

996 where we have approximated the expectation $\mathbf{E}[(\beta^J)^2]$ with the square of our heuristic approximation,

$$\mathbf{E}[(\beta^J)^2] \approx (\beta_{\text{heur}}^J)^2 = \left(\frac{\sum_{i=1}^S (1 - b_{i,o} - d_{i,o})}{\sum_{i=1}^S (1 - d_{i,o})}\right)^2 = \left(\frac{u_{*,o}}{1 - d_{*,o}}\right)^2. \quad (\text{D85})$$

997 The analytical approximation obtained in Eqn (D84) yields always averaged standard deviation relative
998 errors less than 10%. In most of the cases relative errors for the standard deviation, averaged over realizations
999 of incidence vectors, are only about 2%.

1000 **Leading term in the limit of large S**

1001 In order to get more insight about the dependence with S in the limit $S \rightarrow \infty$, we have computed an
1002 asymptotic expansion of the variance to get the leading term in the series expansion on S . First let us write
1003 Eqn (D79) as

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} \int_0^\infty d\theta e^{-\theta} \theta \left(d_{*,o} + b_{*,o}e^{-\theta} + Su_{*,o}e^{-\theta}\right) e^{(S-2)\log(d_{*,o} + (1-d_{*,o})e^{-\theta})}. \quad (\text{D86})$$

1004 In the limit of large S , the exponential function will be very peaked at the maximum of the function
1005 $\log(d_{*,o} + (1 - d_{*,o})e^{-\theta})$. So we expect to have a good approximation in the limit $S \rightarrow \infty$ if we replace the
1006 logarithm by its series expansion,

$$\log(d_{*,o} + (1 - d_{*,o})e^{-\theta}) \approx -(1 - d_{*,o})\theta + \mathcal{O}(\theta^2), \quad (\text{D87})$$

1007 about the point at which the maximum is reached, i.e. $\theta = 0$. Then, for large S , Eqn (D79) will be nicely
1008 approximated by

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} \int_0^\infty d\theta e^{-\theta} \theta \left(d_{*,o} + b_{*,o}e^{-\theta} + Su_{*,o}e^{-\theta}\right) e^{-(S-2)(1-d)\theta}. \quad (\text{D88})$$

1009 This integral can be actually evaluated to give

$$\mathbf{E}\left[(\beta^J)^2\right] \approx Su_{*,o} \left(\frac{d_{*,o}}{(1 + (S-2)(1 - d_{*,o}))^2} + \frac{b_{*,o} + Su_{*,o}}{(2 + (S-2)(1 - d_{*,o}))^2} \right) \quad (\text{D89})$$

1010 plus subleading terms in S . Here we observe that our approximation for $\mathbf{E}\left[(\beta^J)^2\right]$ converges to the squared
1011 heuristic Jaccard measure approximation,

$$\lim_{S \rightarrow \infty} \mathbf{E}\left[(\beta^J)^2\right] \approx \left(\frac{u_{*,o}}{1 - d_{*,o}}\right)^2, \quad (\text{D90})$$

1012 so, in the limit of large S we find the following leading term for the variance approximation:

$$V[\beta^J] \approx \frac{u_{*,o} \left((1 - d_{*,o})(b_{*,o} + d_{*,o}) - 4d_{*,o}u_{*,o} \right)}{(1 - d_{*,o})^3 S}. \quad (D91)$$

1013 The variance decreases as S^{-1} in the case of large number of species. This explains why our heuristic
1014 approximation works very well in that limit.

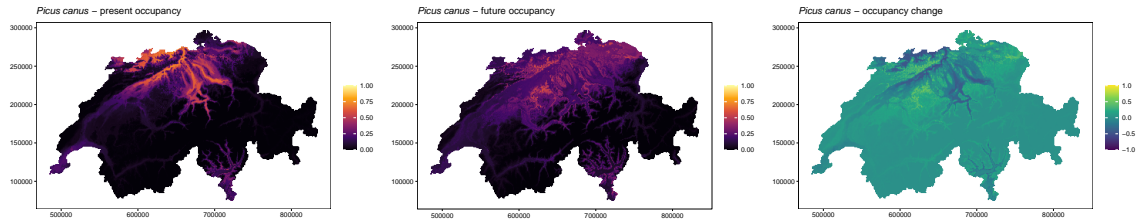
1015 Appendix E SDM

1016 Data was collected over a four-year period (1993-1996) in usually three visits per year (2 above the treeline)
1017 using a simplified territory mapping approach, and integrated in the Swiss breeding bird atlas at 1-by-1 km
1018 resolution (Schmid *et al.* 1998, 2018). The data source we used included environmental predictor variables
1019 corresponding to climate, topography and vegetation structure at the same spatial scale, but geographic
1020 coordinates were removed. Data can be found at Zurell *et al.* 2019b,a. We used the whole dataset to infer
1021 SDMs using only the climatic variables, as current and future values of these variables for Switzerland are
1022 available in worldclim (www.worldclim.org; Hijmans *et al.* 2005). We downloaded current climatic data
1023 using function `getData` from the R package `raster` using argument `name = 'worldclim'`, and future climate
1024 with the same function call with arguments `name = 'CMIP5'`, `rcp = 45`, `year = 50`, and `model = 'NO'` For each
1025 species, we used an ensemble approach of, initially, four different algorithms: generalized linear models
1026 (GLMs), generalized additive models (GAMs), random forests (RFs) and boosted regression trees (BRTs).
1027 However, GLMs and GAMs produced unreliable projections and were subsequently excluded from our
1028 analyses. We then projected current and future incidences for each species in our ensemble approach.
1029 With those incidences, we calculated the expected dissimilarity provided by Eqn (2) at each location of
1030 Switzerland for the subcommunity of woodpeckers.

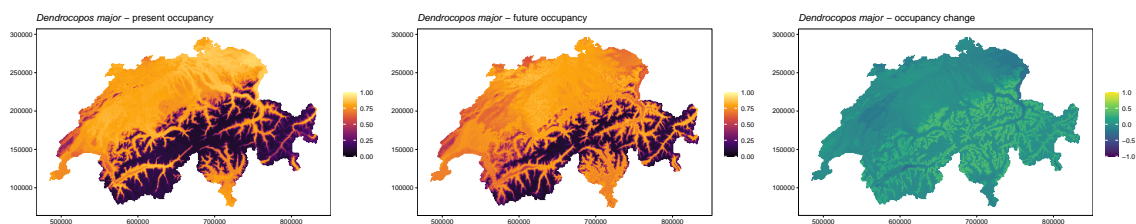
1031 Literatures cited in Appendices

- 1032 Arnold, B. C. (2012). *Majorization and the Lorenz order: A brief introduction*. **43**. Springer Science & Business
1033 Media (cit. on p. 33).
- 1034 Chao, A. & Ricotta, C. (2019). Quantifying evenness and linking it to diversity, beta diversity, and similarity.
1035 *Ecology*, **100**.12. DOI: [10.1002/ecy.2852](https://doi.org/10.1002/ecy.2852) (cit. on p. 33).
- 1036 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated
1037 climate surfaces for global land areas. *International Journal of Climatology*, **25**.15, pp. 1965–1978. DOI:
1038 [10.1002/joc.1276](https://doi.org/10.1002/joc.1276) (cit. on p. 37).
- 1039 Hoover, E. M. (1936). The measurement of industrial localization. *The Review of Economics and Statistics*,
1040 **18**.4, p. 162. DOI: [10.2307/1927875](https://doi.org/10.2307/1927875) (cit. on p. 31).
- 1041 Hubbard, J. (1959). Calculation of partition functions. *Physical Review Letters*, **3**.2, pp. 77–78. DOI: [10.1103/](https://doi.org/10.1103/physrevlett.3.77)
1042 [physrevlett.3.77](https://doi.org/10.1103/physrevlett.3.77) (cit. on p. 35).
- 1043 Jost, L. (2006). Entropy and diversity. *Oikos*, **113**.2, pp. 363–375. DOI: [10.1111/j.2006.0030-1299.14714.x](https://doi.org/10.1111/j.2006.0030-1299.14714.x)
1044 (cit. on p. 33).
- 1045 — (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, **88**.10, pp. 2427–
1046 2439. DOI: [10.1890/06-1736.1](https://doi.org/10.1890/06-1736.1) (cit. on p. 33).
- 1047 Leinster, T. (2021). *Entropy and Diversity: The Axiomatic Approach*. Cambridge University Press (cit. on p. 33).
- 1048 Lu, M., Vasseur, D., & Jetz, W. (2019). Beta diversity patterns derived from island biogeography theory. *The*
1049 *American Naturalist*, **194**.3, E52–E65. DOI: [10.1086/704181](https://doi.org/10.1086/704181) (cit. on pp. 2, 6, 7, 12, 25).
- 1050 Marshall, A. W., Olkin, I., & Arnold, B. C. (1979). *Inequalities: theory of majorization and its applications*. **143**.
1051 Springer (cit. on p. 33).
- 1052 McVinish, R. & Lester, R. J. G. (2020). Measuring aggregation in parasite populations. *Journal of The Royal*
1053 *Society Interface*, **17**.165, p. 20190886. DOI: [10.1098/rsif.2019.0886](https://doi.org/10.1098/rsif.2019.0886) (cit. on p. 33).

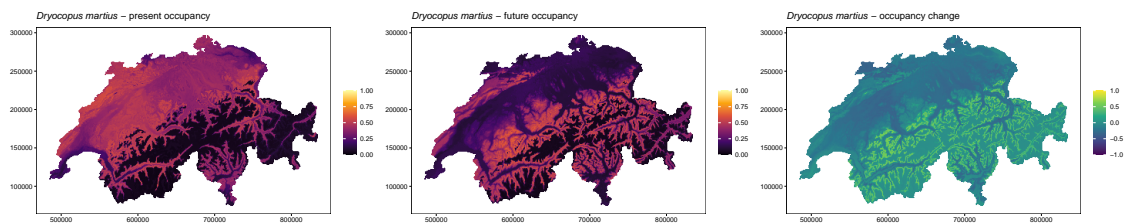
(A)



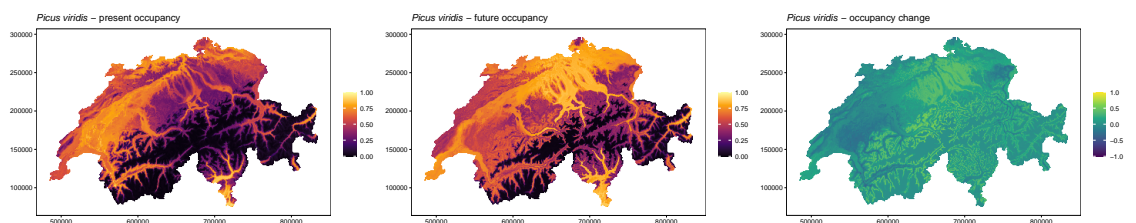
(B)



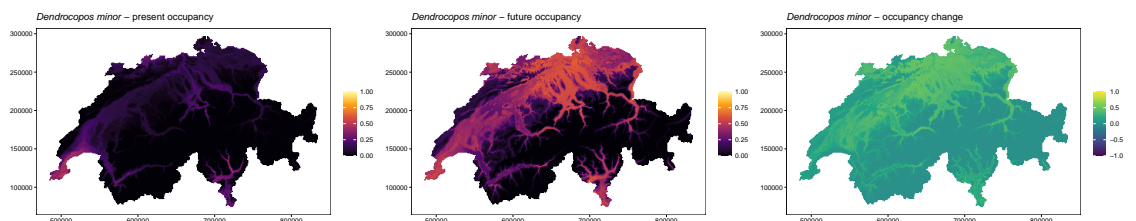
(C)



(D)



(E)



SI Figure 4: The presence probabilities: $p_{i,current}$, $p_{i,future}$, and $p_{i,future} - p_{i,current}$.

- 1054 Ontiveros, V. J., Capitán, J. A., Casamayor, E. O., & Alonso, D. (2021). The characteristic time of ecological
1055 communities. *Ecology*, **102**.2. DOI: [10.1002/ecy.3247](https://doi.org/10.1002/ecy.3247) (cit. on pp. 3, 7, 28).
- 1056 Routledge, R. D. (1983). Evenness indices: are any admissible? *Oikos*, **40**.1, p. 149. DOI: [10.2307/3544211](https://doi.org/10.2307/3544211)
1057 (cit. on p. 33).
- 1058 Schmid, H., M. Kestenholz, Knaus, P., L. Rey, & Sattler, T. (2018). The state of birds in Switzerland: special
1059 issue on the breeding bird atlas 2013–2016. en. DOI: [10.13140/RG.2.2.17872.15366](https://doi.org/10.13140/RG.2.2.17872.15366) (cit. on pp. 12, 37).
- 1060 Schmid, H., Posse, B., & Vogelwarte, S. S. (1998). *Schweizer Brutvogelatlas: Verbreitung der Brutvögel in der*
1061 *Schweiz und im Fürstentum Liechtenstein 1993-1996*. Schweizerische Vogelwarte (cit. on pp. 12, 37).
- 1062 Van Lint, J. H. & Wilson, R. M. (2001). *A course in combinatorics*. Cambridge University Press (cit. on pp. 6,
1063 26).
- 1064 Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon*, **21**.2-3, pp. 213–251. DOI:
1065 [10.2307/1218190](https://doi.org/10.2307/1218190) (cit. on pp. 2, 23, 24).
- 1066 Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T., & Wüest, R. O. (2019a). *Data from:*
1067 *Testing species assemblage predictions from stacked and joint species distribution models*. en. DOI: [10.5061/](https://doi.org/10.5061/DRYAD.K88V330)
1068 [DRYAD.K88V330](https://doi.org/10.5061/DRYAD.K88V330) (cit. on p. 37).
- 1069 — (2019b). Testing species assemblage predictions from stacked and joint species distribution models.
1070 *Journal of Biogeography*, **47**.1, pp. 101–113. DOI: [10.1111/jbi.13608](https://doi.org/10.1111/jbi.13608) (cit. on pp. 12, 37).