ECOGRAPHY

Research article

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

Ryosuke Iritani[1,2], **Vicente J. Ontiveros**[3,4,5], **David Alonso**[4], **José A. Capitán**[5], **William Godsoe**[6] and **Shinichi Tatsumi**[7]

[1]RIKEN Center for Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS), RIKEN, Wako, Saitama, Japan
[2]Department of Biology, Faculty of Sciences, The University of Tokyo
[3]Institute of Aquatic Ecology, University of Girona, Girona, Spain
[4]Theoretical and Computational Ecology, Center for Advanced Studies of Blanes (CEAB-CSIC), Spanish Council for Scientific Research, Blanes, Spain
[5]Department of Applied Mathematics, Complex Systems Group, Universidad Politécnica de Madrid, Madrid, Spain
[6]BioProtection Research Centre, Lincoln University, Lincoln, New Zealand
[7]Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Correspondence: Ryosuke Iritani (lambtani@gmail.com)

Ecography
2025: e07737
doi: 10.1002/ecog.07737
Subject Editor: Andres Baselga
Editor-in-Chief: Miguel Araújo
Accepted 14 March 2025

A fundamental problem in ecology is understanding the changes in species composition among sites (i.e. beta-diversity). It is unclear how spatial heterogeneity in species occupancy across sites shapes patterns of beta-diversity. To address this question, we develop probabilistic models that consider two spatial or temporal sites, where presence probabilities vary both among species and between the sites. We derive analytical and approximate formulae for the expectation of pairwise beta-diversity. Using a graphical tool, stochastic incidence plots (SIPs), which depict the presence probabilities in two sites along species labels, we develop a means to conceptualize the heterogeneity in presence probabilities: the steepness or unevenness of SIPs reflects species-level heterogeneity, while the degree of overlap between SIPs indicates site-level heterogeneity. We find that when SIPs completely overlap (i.e. two sites have the same presence probability for each species), flat SIPs - with all species having the same presence probability - maximize the expected beta-diversity. We refer to this prediction as the 'transfer principle for beta'. Second, using SIPs and the probabilistic method in a two-species scenario, we demonstrate that beta-diversity is lower when SIPs are parallel compared to when they are anti-parallel. We also find that this prediction is consistent with the well-known checkerboard pattern in incidence matrices. Finally, we apply the method to the species distribution models for five woodpecker species in Switzerland, showing that their spatial distributions will change significantly. Overall, this work improves our understanding of how pairwise beta-diversity responds to occupancy heterogeneity.

Keywords: beta-diversity, heterogeneity, Jaccard index, null models, similarity coefficient, species distribution models, stochasticity

NORDIC SOCIETY OIKOS
www.ecography.org

Introduction

Beta-diversity represents the spatial or temporal variation in species compositions (Whittaker 1972, Anderson et al. 2010, Chase et al. 2019). Beta-diversity links diversity across scales, and varies with fundamental processes such as dispersal, environmental filtering, and species interactions (Anderson et al. 2010, Vellend 2010, Socolar et al. 2016, Maynard et al. 2017, Legendre 2019, Thompson et al. 2020). The decline in beta-diversity, biotic homogenization, has raised significant concerns, as it may reduce ecosystem functioning across the globe (Olden and Poff 2003, Olden and Rooney 2006, Hautier et al. 2017, Mori et al. 2018, Olden et al. 2018, Albrecht et al. 2021, Wang et al. 2021, Rolls et al. 2023). Clarifying the patterns of changes in beta-diversity ('beta-diversity patterns') in response to variations in biotic and abiotic conditions should lead to an improvement in our ability in biodiversity management, conservation, and urban planning in our modern society, from a spatially global perspective (Crowther et al. 2015).

Estimates for beta-diversity include pairwise indices derived from empirical presence–absence data (incidence data; Colwell and Coddington 1994, Koleff et al. 2003). Incidence-based beta-diversity leverages simplicity by using binary data. Despite its simplicity, patterns of variability in incidence-based beta-diversity are not well conceptualized, with previous studies yielding mixed results. For example, some theoretical work shows that dispersal reduces beta-diversity (Loreau 2000, Thompson et al. 2020), whereas other work shows that the impact of dispersal depends on metacommunity model-structure (Lu et al. 2019, Lu 2021). Meanwhile, experimental work suggests that dispersal may promote beta-diversity (Vannette and Fukami 2017). One of the challenging goals in the field is to make better sense of such complex beta-diversity patterns.

Organisms live in highly heterogeneous environments (Abson 2017), with such heterogeneity rendering the problem of beta-diversity patterns even more challenging. That is, patterns of species occupancy (which we refer to as 'chance of occurrence' in this article) are highly heterogeneous and may shape complex beta-diversity patterns. In the simplest scenario, the chance of occurrence is equal among all species across all sites (Fig. 1A). This premise does not always hold, however. In general, species have different chances of occurrence in space (for instance, due to differences in niches), exhibiting geographically interspersed distributions (Fig. 1B; 'checkerboard pattern', Diamond 1975, Connor and Simberloff 1979). To provide another example, one site may have a higher chance of occurrence for all species than other sites, even if all species have an equal chance of occurrence (Fig. 1C). In such cases, the incidence at one site is nested within that of another, thus leading to large beta-diversity (Darlington 1957, Patterson and Atmar 1986, Patterson 1987, Ulrich and Gotelli 2007). In general, the chance of occurrence differs both among species and sites (Fig. 1D), i.e. the chance of occurrence is heterogeneous among species and across sites ('occupancy heterogeneity', MacKenzie et al. 2018, Chapter 4); in such cases, interpreting beta-diversity patterns, if any, is highly challenging. The problem thus is how we can better conceptualize beta-diversity patterns associated with occupancy heterogeneity.

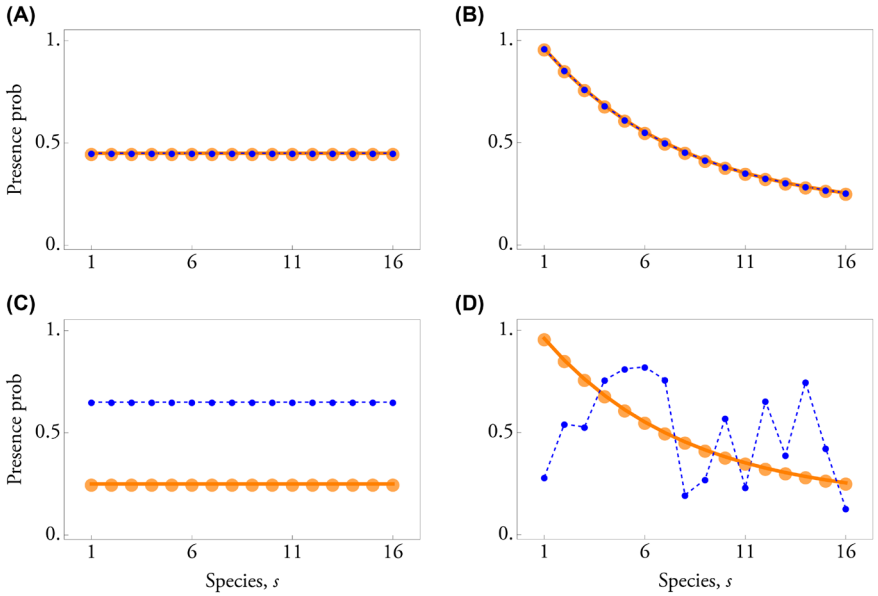

Figure 1. Stochastic incidence plots (SIPs). Presence probabilities in site 1 (orange) and site 2 (blue) are depicted along species labels. When two SIPs completely overlap, the two sites are homogeneous (A)–(B). When two SIPs are both flat, all species have the same chance of occurrence (A), (C). In general, two SIPs are different and neither of them is flat (D), with numerous patterns of SIPs possible. SIPs therefore capture the intricate heterogeneity in the presence probabilities of species across space.

Numerous theoretical studies have attempted to predict beta-diversity patterns. Specifically, recent work has developed a general, metacommunity dynamical model on the effects of density-independent abiotic factors, density-dependent species interactions, and dispersal among patches, on beta-diversity (Whittaker 1972, Thompson et al. 2020). However, Thompson et al. (2020) do not explicitly consider how beta-diversity co-varies with occupancy heterogeneity per se. We take a complementary, static approach: we develop a probabilistic framework and model occupancy heterogeneity as the differences in the probabilities of species occurrences, without explicitly specifying mechanisms and processes of community assembly. Applying the stochastic theory has several advantages. First, it allows us to study summary statistics, e.g. expectation and variance, of beta-diversity and to examine how they vary with the variations in the presence probabilities among species and among sites. It thus enables us to gain information on the null distribution of beta-diversity against which to generate and test hypotheses (Gotelli and Graves 1996, Leibold et al. 2004, Gotelli and Ulrich 2011, Hui and McGeoch 2014, Chung et al. 2019, Thompson et al. 2020). Second, it allows us to consider various patterns of occupancy heterogeneity (Fig. 1A–D) and therefore enables an exploration of their co-variability with beta-diversity. Third, the recent theoretical development of occupancy estimation is based on the stochasticity of species occurrences (MacKenzie et al. 2018).

Throughout the article, we assume that species incidences are independent of each other both within and between sites to nullify correlations among them, the so-called 'species-independence assumption' (Chung et al. 2019, Lu et al. 2019, Kalyuzhny et al. 2021, Lu 2021, Ontiveros et al. 2021). Biologically, the species-independence assumption implies that no biotic interaction occurs, and environmental effects are captured by the probability that species are present. We derive the exact formulae and approximation of the expectation and variance of a pairwise beta-diversity index, Jaccard dissimilarity (Jaccard 1908, 1912, Veech 2013, Arita 2017, Keil et al. 2021). We then derive several theoretical predictions of the consequences of occupancy heterogeneity for the expectation of beta-diversity. In addition, we propose using a diagram ('stochastic incidence plots'; SIPs) that depicts the presence probabilities along species labels for each site, thereby easing a visual interpretation of occupancy heterogeneity (Fig. 1). For example, non-flat (e.g. sloped) SIPs represent a scenario where presence probabilities differ among species; also, when two SIPs do not fully overlap (e.g. nested), two sites are differentiated in species presence probabilities. SIPs can therefore allow us to interpret the co-variability of beta-diversity and occupancy heterogeneity. We highlight the two-species case as the minimal model to fully analyze the relationship between beta-diversity and occupancy heterogeneity. We thereby link the SIPs and the probability distributions of incidence matrices. This analysis allows us to examine how a pair of species with different chances of occurrence influences beta-diversity.

Method and result

The notation is summarized in Table 1.

Model

Incidence matrix

We consider a landscape consisting of two spatial locations (or temporal points) and compare the difference in species compositions. We write S_{T} for the maximum number of species present in the landscape ('species pool size'). We write $X_{r,s}$ for a binary state of incidence of species s in site r ($X_{r,s}=1$ for presence and $X_{r,s}=0$ for absence, respectively), with $s=1,2,\ldots,S_{\mathrm{T}}$ and $r=1,2$. We generically write $\mathbf{X}$, whose (r,s)-element is $X_{r,s}$, for a 2-by-S_{T} incidence matrix.

Diversity index

Here we define diversity measures. The sum $\sum_{s=1}^{S_{\mathrm{T}}} X_{r,s}$ represents species richness in site r, and $\sum_{s=1}^{S_{\mathrm{T}}} X_{1,s}X_{2,s}$ represents the number of species shared by the two sites. Also, we define the following Boolean variables. If species s is present in at least one of the sites, then $X_{1,s}+X_{2,s}-X_{1,s}X_{2,s}=1$ (otherwise, 0). Therefore, gamma-diversity, defined as the number of species present in the landscape, is given by $\sum_{s=1}^{S_{\mathrm{T}}}(X_{1,s}+X_{2,s}-X_{1,s}X_{2,s})$. Similarly, if species s is present in only one of the sites (referred to as being 'unique' in this study, following Colwell and Coddington 1994), then $X_{1,s}+X_{2,s}-2X_{1,s}X_{2,s}=1$ (otherwise, 0). Therefore, the number of unique species is given by $\sum_{s=1}^{S_{\mathrm{T}}}(X_{1,s}+X_{2,s}-2X_{1,s}X_{2,s})$.

Using these Boolean variables, Jaccard dissimilarity (JD) is given by:

$$J_{\mathbf{X}}=\frac{\left(\#\text{ unique species}\right)}{\left(\text{gamma diversity}\right)}=\frac{\sum_{s=1}^{S_{\mathrm{T}}}X_{1,s}+X_{2,s}-2X_{1,s}X_{2,s}}{\sum_{s=1}^{S_{\mathrm{T}}}X_{1,s}+X_{2,s}-X_{1,s}X_{2,s}} \quad (1)$$

(Jaccard 1908, 1912; the Supporting informationexplains why JD is a suitable beta-diversity measure). Though JD has long been used as a measure of beta-diversity, its probabilistic properties are poorly understood (Real and Vegas 1996).

Stochastic analysis

Here we describe stochasticity in incidence matrices assuming that species incidence is independent of each other (species independence assumption). We write $p_{r,s}$ for the probability that species s is present in site r, i.e. probability that $X_{r,s}=1$ (MacArthur and Wilson 1963, Real et al. 2016, Carmona and Pärtel 2020). For notational simplicity, we write $a_{r,s}=1-p_{r,s}$ for the probability that species s is absent from site r. The species-independence assumption implies that the probability of the occurrence of incidence matrix $\mathbf{X}$ is given by:

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

Notation	Definition	Note
s, s'	Species label	$s=1,2,\ldots,S_T$
S_T	The maximum number of present species	'species pool size'
r	Site label, $r=1,2$	'sites' may be spatial or temporal
$X_{r,s}$	Incidence of species s in site r	0 (absence) or 1 (presence)
$\mathbf{X}_{(r,S_T)}$	Incidence matrix of size 2-by-S_T	abbreviated to $\mathbf{X}$
$:=$	Defining a quantity	
$p_{r,s}$	Prob species s is present in r	$a_{r,s}=1-p_{r,s}$ for the prob of absence
b_s	Prob species s is present in both sites	b for 'both'
d_s	Prob species s is absent from both sites	d for 'double absence'
u_s	Prob species s is present in either site 1 or 2	u for 'unique'
$P_{\mathbf{X}}$	Prob the incidence matrix $\mathbf{X}$ occurs	Eq. 2
$J_{\mathbf{X}}$	Jaccard dissimilarity for an incidence matrix $\mathbf{X}$	JD
$\gamma_{\mathbf{X}}$	The number of species present for matrix $\mathbf{X}$	'gamma-diversity'
E[J]	Expectation of Jaccard dissimilarity	cE[J] for conditional expectation
J_{heur}	Approximation of cE[J]	'heuristic approximation'
μ_r	Average presence probability in site r	
α	Alpha diversity (richness per site)	c.f. key result 2

$$P_{\mathbf{X}} = \prod_{s=1}^{S_T}\prod_{r=1}^{2} p_{r,s}^{X_{r,s}} a_{r,s}^{1-X_{r,s}} \tag{2}$$

Next, we define the following probabilities: species s is 1) present in both sites with probability $b_s := p_{1,s}\cdot p_{2,s}$; 2) unique with probability $u_s := a_{1,s}\cdot p_{2,s} + a_{2,s}\cdot p_{1,s}$; or 3) absent from both sites with probability $d_s := a_{1,s}\cdot a_{2,s}$, with $b_s+u_s+d_s=1$. Also, we write $\mu_r := \sum_{s=1}^{S_T} p_{r,s} / S_T$ for the average presence probability in site r.

It is worth examining how these probabilities, b_s, u_s, and $1-d_s(=b_s+u_s)$, change with presence probabilities ($p_{1,s}$, $p_{2,s}$), especially the effect of similarity between $p_{1,s}$ and $p_{2,s}$ (Supporting information). The probability of unique presence, u_s, is large when two sites have dissimilar presence probabilities (Supporting information). The probability of double-presence, b_s, is large when two sites have large and similar presence probabilities. Finally, the probability of presence in at least one site, $1-d_s$, is large when two sites have large and dissimilar presence probabilities.

Stochastic incidence plots (SIPs)

Given that beta-diversity varies with occupancy heterogeneity, how can we interpret such co-variability? We here introduce a graphical tool, stochastic incidence plots (SIPs). SIPs consist of a pair of plots each depicting the presence probabilities in site r, $(p_{r,1}, p_{r,2}, \ldots, p_{r,S_T})$, along species labels $s=1,2,\ldots S_T$, where we set $p_{1,1}$ as the maximum among the all $p_{r,s}$'s. Note that the idea of depicting occupancy probability along some variables (e.g. species, or environmental gradient) per se is not novel.

SIPs provide a useful visual interpretation of occupancy heterogeneity (Fig. 1). First, SIPs fully overlap each other when two sites have the same presence probabilities for each species (Fig. 1A, B). Biologically, this scenario is relevant when: 1) presumably the environmental conditions in two sites are similar, or 2) the actual environmental conditions are unknown and therefore one uses a common value for both sites; note that we can incorporate correlations between sites for the analyses of JD (Supporting information), but we focus on the case of two sites being independent.

Second, SIPs are both flat if and only if all species have the same presence probability in each site: $p_{r,1} = p_{r,2} = \ldots = p_{r,S_T}$ for each site, $r=1,2$, (Fig. 1A, C). Biologically, this occurs when all species are equiprobably present (i.e. all species have the equal chance of occurrence). Finally, SIPs do not fully overlap nor are flat (Fig. 1D), in which case there are diverse patterns of intersections of SIPs. SIPs can therefore allow us to qualitatively characterize the occupancy heterogeneity.

Analytical result: expectation of JD and its approximation

We denote the expectation of JD by $\mathrm{E}[J] := \sum_{\mathbf{X}} P_{\mathbf{X}} J_{\mathbf{X}}$ where the sum is taken over all possible incidence matrices sized 2-by-S_T. Using the probability-generating polynomials (Cressie et al. 1981, Lange 2010), we get:

$$\mathrm{E}[J] = \sum_{s=1}^{S_T} u_s \int_0^1 \left(\prod_{k(\neq s)}^{S_T} \left(d_k + (1-d_k) z \right) \right) \mathrm{d}z \tag{3}$$

(See the Supporting information for derivation; note that we can calculate analytically the integral of the product by using the elementary symmetric polynomials and the beta function). Equation 3 reads as the sum, over species label s, of the probability that species s is present in only one of the sites (u_s) times the integral of the polynomial product over an interval from 0 to 1 (with the polynomial referred to as the probability-generating polynomial). In the Supporting information, we prove that the integral for each species s represents the conditional expectation of the reciprocal of gamma-diversity (including species s). We also search for analytical formulae of

other beta-diversity measures, namely the indices of Sørensen (1948) and Whittaker (1972), by using the same method (Supporting information).

The conditional expected value of JD, given that there is at least one species present, is given by $\mathrm{cE}[J] := \mathrm{E}[J] / \left(1 - \prod_{s=1}^{S_T} d_s\right)$, which can be accurately approximated by:

$$J_{\mathrm{heur}} := \frac{\sum_{s=1}^{S_T} u_s}{\sum_{s=1}^{S_T} (1 - d_s)} \tag{4}$$

(Chung et al. 2019, Lu et al. 2019, Kalyuzhny et al. 2021, Lu 2021, Ontiveros et al. 2021, Supporting information). We refer to Eq. 4 as 'heuristic approximation' in this literature. The heuristic approximation reads as the proportion of the expected number of unique species to the expected number of present species. We find that the heuristic approximation provides a near-identical result with the conditional expectation (Fig. 2) and converges to the exact value as the species pool size increases, $S_T \to +\infty$ (Supporting information; note that this limit leads to $\mathrm{cE}[J] \approx \mathrm{E}[J]$).

We also derive the exact and approximation formulae of the variance of JD using the generalized hypergeometric functions. We find that for flat SIPs, the variance tends to change only moderately with varying presence probabilities, and decreases as the species pool size increases (Supporting information). We do not provide more results of the variance, as the technical details are complex.

Benchmark result: species-equivalence (flat SIPs)

We here assume that all species have an even chance of occurrence, so that SIPs are both flat (though not necessarily fully overlapping). In this case, we find $\mathrm{cE}[J] = J_{\mathrm{heur}}$; that is, if $p_{r,s}$ is common for all species for each site $r = 1,2$, then the heuristic approximation equates to the exact conditional expectation (Supporting information).

This result generalizes the results of the previous literature. Chung et al. (2019) and Ontiveros et al. (2021) have used the heuristic approximation for the special case where the presence probabilities are all equal (flat and overlapping; Fig. 1A), and Lu et al. (2019) and Lu (2021) have generalized the result to the case where SIPs are flat but are not fully overlapping (Fig. 1C). Therefore, the present results of the heuristic approximation are consistent with the previous studies that use the heuristic approximation as an alternative to the exact expected value of JD assuming species-equivalence in presence probability.

Key result 1: site-homogeneity (overlapping SIPs)

We next consider fully overlapping SIPs while keeping the total average, $\mu_1 + \mu_2$, fixed. We show that, as the chance of occurrence gets less even (i.e. more uneven), the expectation of beta-diversity gets smaller (Supporting information). Specifically, the expectation is maximized when all species have the same chance of occurrence (Fig. 3). We refer to this result as 'transfer-principle for beta' (Dalton 1920, Jost 2010, Chao and Ricotta 2019). The transfer-principle for beta can be interpreted in three different ways. Mathematically, the expectation of JD is a Schur-concave function (Marshall et al. 1979, Tuomisto 2012, Chao and Ricotta 2019); graphically, flat SIPs (thus with an equal height for all species) have the maximum beta-diversity (Fig. 3; see Supporting information for more details); and biologically, if the occurrence chance is dominated by a subset of species equally in both sites, beta-diversity is likely low, because such species are likely found in both sites but the others are likely absent from both sites.

Key result 2: two species-case under heterogeneity (general SIPs)

In the general case where species differ and sites are heterogeneous, it is hard to interpret co-variability of beta-diversity with the heterogeneity. To facilitate intuitive understanding, we focus on a case with two species, s and s', using alphabetical labels to avoid confusion with site indices 1 and 2. In this scenario, there are in total $2^{2\cdot 2} = 16$ incidence matrices with different values of JD. Specifically, a pair of species is picked from the species pool, and we ask whether they contribute to higher or lower beta-diversity. We do so by analyzing two-by-two incidence matrices and their relationship with SIPs.

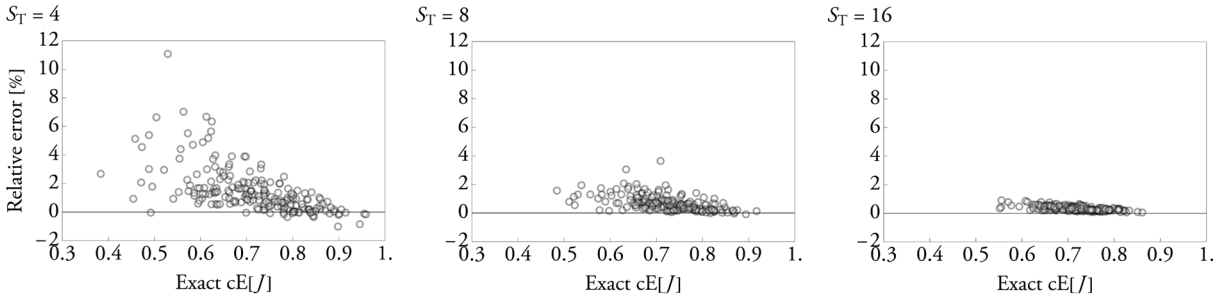

Figure 2. The relative error of the heuristic approximation (defined as $(J_{\mathrm{heur}} - \mathrm{cE}[J])/\mathrm{cE}[J]$), for different species pool sizes $S_T = 4,8,16$. We perform Monte Carlo simulations 200 times by drawing $p_{r,s}$ from the beta distribution (with parameters 0.95 and 1.15) and compute resulting relative errors. The heuristic approximation provides the near-identical approximation of the exact, conditional expectation.

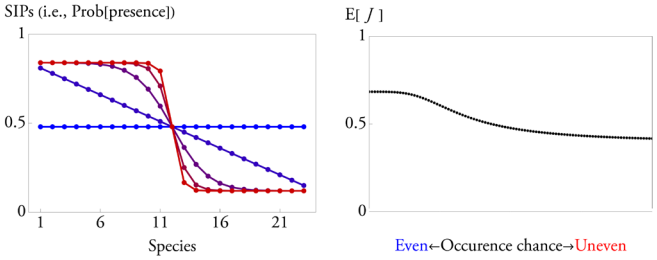

Figure 3. Key result 1, the transfer-principle for beta. The presence probability is identical for both sites for each of the species with a species pool size $S_T = 23$. As the chance of occurrence becomes uneven (i.e. stochastic incidence plots (SIPs) become more sloped, from blue to red; left panel), the expectation of JD decreases (right panel). We use a Hill function that ranges from 0.12 to 0.84, with the average presence probability $(\mu_1 + \mu_2)/2$ kept fixed. An unevenness parameter is varied from 0 (completely even; blue), 1, 4, 9 to 16 (= 4^2; red). See the Supporting information for more details.

Our procedure is summarized as follows (Fig. 4). Step 1), we classify two-by-two incidence matrices (in total 16) into seven classes, based on their effects on raising versus reducing JD. Step 2), we permute four arbitrary values of presence probabilities to generate six SIPs, which we refer to as 'fundamental SIPs', based on their shapes. Step 3), we examine how likely it is that each of the six fundamental SIPs produces each of the incidence matrices of the seven classes.

Step 1): classifying incidence matrices

We first classify the 16 incidence matrices into seven types, based on how JD changes with appending each matrix to an arbitrary incidence matrix. For now, we write $\left((x_{1,1}, x_{1,2}, \ldots, x_{1,S_T}), (x_{2,1}, x_{2,2}, \ldots, x_{2,S_T})\right)$ for a two-by-S_T incidence matrix (to make in-line equation notation concise). With this notation, we suppose one has a hypothetical incidence matrix, say $\mathbf{X}_{(2,3)} = ((1,0,1), (1,1,0))$ at hand, and we append it a two-by-to incidence matrix, say $\mathbf{X}_{(2,2)} = ((1,0), (0,1))$, to get a new incidence matrix $\mathbf{X}_{(2,5)} = ((1,0,1,1,0), (1,1,0,0,1))$. Since $J_{\mathbf{X}_{(2,3)}} = 2/3 < J_{\mathbf{X}_{(2,3)}} = 4/5$, appending the two-by-two $\mathbf{X}_{(2,2)}$ to the hypothetical $\mathbf{X}_{(2,3)}$ increases beta-diversity in this example.

We generalize this procedure. We append each of the 16 two-by-two incidence matrices to an arbitrary incidence matrix (like $\mathbf{X}_{(2,3)}$ above) and assess whether this appendix increases or decreases JD. As is summarized in Table 2, it

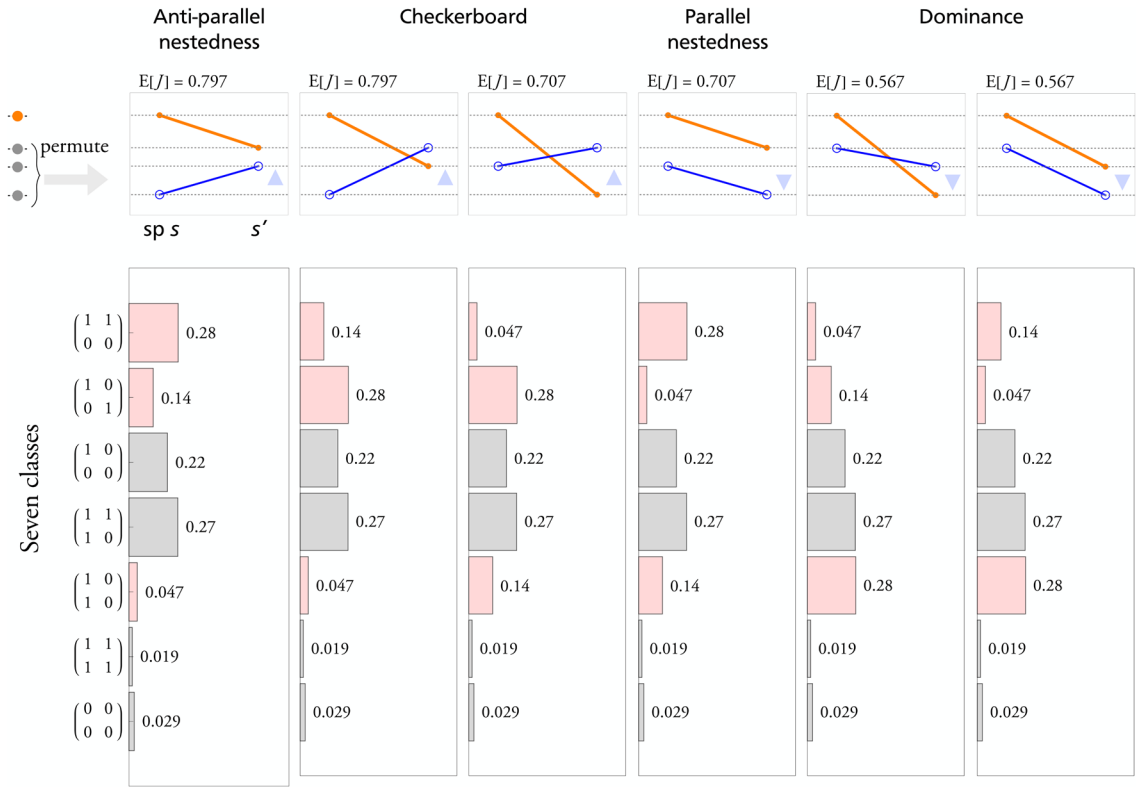

Figure 4. The permutation approach to produce the six fundamental stochastic incidence plots (SIPs) (top six panels), with four values (0.88, 0.56, 0.38, and 0.10) assigned to the presence probabilities. Blue triangle indicates a slope of SIP in site 2 (upward or downward). The histograms of the seven classes of incidence matrices, in which gray bars are common across all the fundamental SIPs, whereas the pink bars are not. Anti-parallel nestedness SIPs most likely produce the nestedness class (top bar in red). The checkerboard SIPs most likely produce the checkerboard class (second top bar in red). Similarly, the parallel nestedness SIPs most likely produce the nestedness class, but also likely produce the dominance class (fifth top bar in red) that reduces Jaccard dissimilarity (JD). The dominance SIPs most likely produce the dominance class (fifth top bar in red).

Table 2. Classification of incidence matrices, with seven classes characterizing the two-by-two matrices. The incidence matrices that have the average number of species (alpha-diversity, α) per site being 1 are further subclassified according to nestedness, checkerboard, and dominance patterns. The name 'dominance' accounts for the fact that, in both sites, one species has higher chance of occurrence than does the other. Otherwise if α = 0, 0.5, 1.5, or 2, then each of them consists of a single class as indicated by '*'. A symbol '++' indicates it has larger effect than does '+', and '—' indicates it has a negatively larger effect than does '–'. The plus-minus symbol ('±') indicates a mixed effect (either increases or decreases Jaccard dissimilarity (JD)).

Incidence class, C	Richness per site (α)	Name	Effect on JD
$\begin{pmatrix}11\\00\end{pmatrix},\begin{pmatrix}00\\11\end{pmatrix}$	1	Nestedness	++
$\begin{pmatrix}10\\01\end{pmatrix},\begin{pmatrix}01\\10\end{pmatrix}$	1	Checkerboard	++
$\begin{pmatrix}10\\00\end{pmatrix},\begin{pmatrix}01\\00\end{pmatrix},\begin{pmatrix}00\\10\end{pmatrix},\begin{pmatrix}00\\01\end{pmatrix}$	0.5*	Single-one	+
$\begin{pmatrix}01\\11\end{pmatrix},\begin{pmatrix}10\\11\end{pmatrix},\begin{pmatrix}11\\01\end{pmatrix},\begin{pmatrix}11\\10\end{pmatrix}$	1.5*	Single-zero	±
$\begin{pmatrix}10\\10\end{pmatrix},\begin{pmatrix}01\\01\end{pmatrix}$	1	Dominance	–
$\begin{pmatrix}11\\11\end{pmatrix}$	2*	All-one	—
$\begin{pmatrix}00\\00\end{pmatrix}$	0*	All-zero	0

follows that: three classes of matrices always increase JD; two of such three classes (named 'nestedness' and 'checkerboard' classes, respectively) increase JD more than does the other; one class of matrices either increases or decreases JD, depending on the JD of the original, prepended matrix to be larger or small than 1/2; there are two classes of matrices that always reduce JD, one of which (named 'dominance' class) reduces JD more than does the other; JD is invariant with appending the all-zero matrix; and the incidence matrices other than nestedness, checkerboard, and dominance can be completely classified only by the average richness (marked '*' in Table 2). We name such seven classes in Table 2.

Step 2): producing SIPs by permutation – fundamental SIPs

To consider numerous shapes of SIPs, we use a permutation approach. We pick four arbitrary values between 0 and 1 (they can be generally distinct), with their maximum assigned to $p_{1,s}$, the presence probability of species s in site 1 (without loss of generality). We then assign the remaining three to the other presence probabilities $p_{1,s'}, p_{2,s}, p_{2,s'}$, obtaining $3! = 6$ pairs of SIPs (Fig. 4). We refer to these six pairs of SIPs as 'fundamental SIPs'.

The advantage of the permutation approach is twofold. First, by construction, the fundamental SIPs exhaustively cover possible shapes (e.g. uneven SIPs), including crossed, parallel, and anti-parallel ones. Second, the fundamental SIPs have the same, total average of presence probability $(\mu_1 + \mu_2)/2$, i.e. we can fix the effect of the average presence probability on the expectation of JD while varying the unevenness.

We then calculate the expected value of JD for each of the fundamental SIPs. Among the six fundamental SIPs, we obtain three distinct types of expected JD values. This halving occurs because JD is invariant under permutations of site labels within each species (i.e. swapping $p_{1,s}$ with $p_{2,s}$ does not change the expectation of JD).

We find that three pairs of fundamental SIPs are crossed each with different expectations of JD, and the others are uncrossed (Fig. 4; Supporting information). Among the uncrossed pairs of fundamental SIPs, two of them are 'parallel' (both SIPs cline downward) and the remaining are 'anti-parallel' (one of the SIPs upward while the other downward). In addition, we find that the anti-parallel pair tends to have higher beta-diversity than does the parallel pair.

We further categorize the six types of fundamental SIPs into four types (Fig. 4): 1) anti-parallel nestedness, where SIPs are anti-parallel but uncrossed; 2) checkerboard, where both SIPs are parallel and crossed; 3) parallel nestedness, where SIPs are parallel and uncrossed; and 4) dominance, where one species' minimum presence probability exceeds the other's maximum. Further details are in the Supporting information.

Step 3): examining the relation between the seven classes and six SIPs

We examine which of the fundamental SIPs contributes to higher or lower expected values of JD, by asking how likely it is that each of the six fundamental SIPs produces which class of matrices (Fig. 4). Towards this end, we evaluate a histogram (probability distribution) of the seven classes generated by each of the fundamental SIPs, using the following formula:

$$\text{Prob}\left[\text{a matrix } \mathbf{X} \text{ of incidence class } C \text{ occurs}\right] = \text{Prob}\left[\mathbf{X} \in C\right] = \sum_{\mathbf{X}\in C}\prod_{s=1}^{2}\prod_{r=1}^{2} P_{r,s}^{X_{r,s}} a_{r,s}^{1-X_{r,s}} \quad (5)$$

where the summation is taken over the 2-by-2 matrices belonging to class C.

We find that, with the same probability, the fundamental SIPs produce each of the single-one, single-zero, all-one, and all-zero matrices (Fig. 4 gray bars, Table 2; Supporting information). Therefore, with different probabilities, the other three classes (nestedness, checkerboard, and dominance classes) are generated by the fundamental SIPs (Fig. 4 red bars), implying that the difference in the expected values of JD for the fundamental SIPs is due to the probabilities of these three classes.

We thus clarify the difference among the nestedness, checkerboard, and dominance incidence matrices by generating the probability distributions of these classes. The histogram, as well as mathematical arguments (Supporting

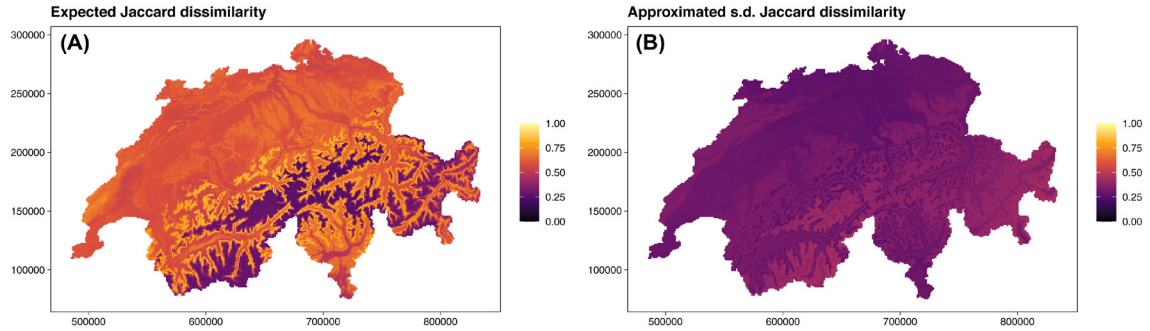

Figure 5. The change in the distributions of woodpeckers. We have quantified the expected, compositional dissimilarity of five woodpecker species at two time points, current and future, over the region of Switzerland, using occupancy estimations for current and future climatic conditions. (A) Expectation. Compositional changes are expected to be high in the upper limit of the current distribution and lowlands. (B) SD (approximated, using the Supporting information). The SD is overall small.

information), shows that: 1) the anti-parallel nestedness SIPs most likely produce the nestedness incidence matrices; 2) the checkerboard SIPs most likely produce the checkerboard incidence matrices; 3) the parallel nestedness SIPs most likely (as likely as do the anti-parallel nestedness SIPs) produce the checkerboard incidence matrices; and 4) the dominance SIPs most likely produce the dominance incidence matrices. We can generalize this result for any four distinct values of presence probabilities $\left(p_{1,s}, p_{1,s'}, p_{2,s}, p_{2,s'}\right)$ used for the fundamental SIPs (Supporting information). Hence, this analysis fully characterizes the relationship between the fundamental SIPs and the incidence classes.

To gain a better insight of beta-diversity patterns, we compare the difference in the expectations of JD for the fundamental SIPs. Specifically, we examine the expectation of JD for the parallel versus anti-parallel pairs of fundamental SIPs. We find that anti-parallel pairs have higher JD than do the parallel pairs. This prediction can be understood considering the association between the SIPs and incidence matrices: the parallel pair corresponds to dominance incidence matrices $\left((1,0),(1,0)\right)$ and $\left((0,1),(0,1)\right)$ in which one species dominates over the other in the occupancy in both sites, thus reducing JD. For the same reason, the anti-parallel pair more likely generates checkerboard patterns than does the parallel pair, resulting in higher JD. Taken together, these findings provide clear interpretations of the relation among SIPs, incidence matrices, heterogeneity, and beta-diversity; that is, the beta-diversity patterns associated with occupancy heterogeneity.

Application to species distribution model

Applications of the present method encompass species distribution models (SDMs; Elith and Leathwick 2009, Guisan et al. 2017, Zurell et al. 2020b). Generally, SDMs seek to estimate the presence probability of each species at a spatial site by using environmental information (e.g. temperature). As SDMs assume that the presence probability of each species is independent of each other, SDMs are well suited to the assumption of the present method. We apply the present method to SDMs of five woodpecker species (*Picus viridis*, *P. canus*, *Dendrocopos major*, *D. minor*, and *Dryocopus martius*; Supporting information, see https://datadryad.org/dataset/doi:10.5061/dryad.k88v330 for data of the present SDMs), using the temporal analogue of beta-diversity (temporal beta-diversity) for an estimate of the difference in the present and future spatial distributions of these species (Legendre 2019, Magurran et al. 2019). We find that the expectations of temporal JD are high across the landscape (Fig. 5), suggesting that the distribution of the woodpeckers will change significantly in the future. We also find that this result is primarily explained by species dynamics in lowland sites where some species thrived and others failed: *P. canus*, which will decrease its occupancy rate near the rivers and will increase in surrounding areas (Fig. 5), and *Dendrocopos minor*, whose occupancy is expected to increase in lowlands and valleys (Fig. 5). JD in hillsides is expected to be moderate due to a general increment in richness (Fig. 5). These results are consistent with a general trend of Switzerland forest birds moving to higher grounds as a response to environmental change (Maggini et al. 2014). See the Supporting information for the full details.

Discussion

Understanding the patterns of variability in beta-diversity estimates in response to heterogeneity is of crucial importance in ecology. We studied the effect of heterogeneity in species' presence probabilities (occupancy heterogeneity) on beta- diversity (Fig. 1). We derived the analytical and approximated formula of the expectation of Jaccard dissimilarity index (JD) assuming that presence probabilities of species across sites are generally different (Fig. 2). We first found that, when all species have the same chance of occurrence in each site, the formula of the expectation of JD used in previous studies provides the analytical expression of the conditional expectation of JD provided that at least one species is present (Chung et al. 2019, Lu et al. 2019, Kalyuzhny et al. 2021, Lu 2021, Ontiveros et al. 2021). Using the formula and numerical approaches, we found that when each species has the equal chance of occurrence in two sites, the evenness in

the chance of occurrence yields higher expectation of JD than does the uneven case (the transfer-principle for beta, Fig. 3, Supporting information). Also, using the graphical approach, which depicts the presence probabilities along species for each site (stochastic incidence plots, SIPs), we classified the patterns of SIPs associated with the incidence matrices (e.g. checkerboard and nestedness; Table 2). We identified two types of SIPs that produce the large expectation of JD: 1) the anti-paralleled nestedness SIPs, in which two SIPs are anti-parallel and one site is nested by the other, and 2) the checkerboard SIPs, in which two SIPs have opposite slopes (upward and downward) and crossed. In contrast, we also found that if there is a dominance relationship between species, in which a portion of species have higher chances of occurrence than do others, then the expectation of JD tends to be low (Fig. 4).

We first focused on the case where each species has the same chance of occurrence in two sites (i.e. two sites are homogeneous). In this case, if the chances of occurrence are more even among species, then the expectation of JD is higher (transfer-principle for beta). This prediction suggests that if certain species are dominant in the chance of occurrence over others, then the beta-diversity tends to be small. Consistently, the two-species analysis showed that the anti-parallel nestedness SIPs have a larger effect on JD than do the parallel nestedness SIPs; in the anti-parallel SIPs, the species' chances of occurrence are balanced between sites, whereas in the parallel SIPs certain species have a predominantly higher chance of occurrence over the other. This consistency suggests that chance of occurrence is balanced among species when the beta-diversity is large. Kvålseth (2015) and Chao and Ricotta (2019) argued that the evenness must satisfy the transfer-principle (or Schur-concavity; c.f. Hill 1973, Marshall et al. 1979, Routledge 1983). The transfer-principle for beta can be thus regarded as an analogous result of the original transfer-principle for local species diversity. From a more practical perspective, the transfer-principle for beta implies that the differences in species' chance of occurrence alone would yield lower spatial beta-diversity when the spatial environment is homogeneous, suggesting that preserving spatial heterogeneity is needed to retain large beta-diversity. It would be also interesting to examine how the results of species distribution models (SDMs; Fig. 5) align with the transfer-principle for beta. However, because the distributions of the species are predicted to change dramatically, direct comparison is likely difficult. Yet, the result might provide useful insight into the temporal changes in species distributions for data with temporal windows being fine-grained (thus those omitting the possibility of large changes). Taken together, the present result has both theoretical and practical implications for beta-diversity patterns associated with occupancy heterogeneity.

We carried out a detailed analysis of the relationship between occupancy heterogeneity and the expectation of JD for the simplest, two-species scenario. We identified the patterns of SIPs associated with the probability of certain incidence matrices to occur. We categorized the SIPs, which we generated by a permutation of four distinct probability values, into four types: anti-parallel nestedness, checkerboard, parallel nestedness, and dominance. These categories have directly to do with empirically suggested patterns of incidence matrices. The anti-parallel nestedness SIPs most likely produce nested patterns of incidence matrices, whereas the checkerboard SIPs most likely produce the checkerboard patterns of incidence matrices. The parallel nestedness SIPs also most likely produce the nested patterns of incidence matrices, but have a lower expectation of JD because these SIPs are also likely to produce dominance patterns of incidence matrices. Intuitively, in the anti-parallel nestedness and checkerboard SIPs, one site may be suitable for one species while the other site for other species, thus increasing beta-diversity. The dominance SIPs likely yield a low expectation of JD. This is because when one species is more likely to be present in both sites than is the other, it is likely that the incidence matrices have mixed or negative effects on JD. From this categorization, we can gain better insights on the estimates of beta-diversity, especially in relation to occupancy heterogeneity.

We showed that the heuristic approximation of JD agrees well with the conditional expectation of JD, with its accuracy improving as the species pool size increases. The heuristic approximation saves much computational cost compared to calculating the exact in a brute force approach by avoiding integrating the polynomial products. We therefore suggest that when species pool size is large, using the heuristic approximation is suitable. If one wishes to evaluate the exact expectation (e.g. when species pool size is small), applying the Gauß fast Fourier transforms can help in saving the computational cost (Cooley and Tukey 1965, Heideman et al. 1984). We remark that the heuristic approximation uses the average values of uniqueness and double-presence probabilities, not the average values of the presence probabilities. Using the averaged presence probabilities corresponds to asking whether flat SIPs yield a good approximation of generally non-flat SIPs. We found that the relative error is large, although becomes small as the species pool size increases (Supporting information). These predictions provide useful guidance for researchers in deciding which of the statistics of JD to use.

We used a parsimonious approach, assuming the minimum species pool size $S_T = 2$, to examine the relationship between SIPs and incidence matrices (Fig. 4). This approach can be generalized to larger species pool sizes. For example, even if the species pool size is large like $S_T = 100$, we can extract a subset pair of species s and s' , and examine whether this pair form the nestedness, checkerboard, or dominance incidence-pattern. This approach has long been used in ecology, with a history dating back to the early studies of community assemblages and island biogeography (MacArthur and Wilson 1963, MacArthur and Wilson 1967, Diamond 1975, Connor and Simberloff 1979), nowadays known as 'null model analysis' (Gotelli and Graves 1996, Gotelli 2000, Peres-Neto et al. 2001, Leibold et al. 2004, Gotelli and Ulrich 2011, Hui and McGeoch 2014). Building up the theoretical framework from the simplest to more complex (and realistic) scenarios could be a next step to advance the field of quantitative ecology.

How does our framework align with traditional, randomization-based null model analyses? In general, null model analyses evaluate the null distribution of species co-occurrence to detect deviations of observed patterns. Within this framework, 'environmentally constrained null models' allow species to vary in presence probabilities (i.e. occupancy heterogeneity; Peres-Neto et al. 2001, D'Amen et al. 2017). Specifically, Peres-Neto et al. (2001) assume that the sum of presence probabilities for each species inhabiting numerous sites is fixed (to unity), analogous to the assumption of the fixed average sum $\mu_1 + \mu_2$ in deriving the transfer-principle for beta (Fig. 3). We can then formulate a corresponding prediction for co-occurrence analyses under environmentally constrained null models as follows: we swap the labels of species and sites (considering two species occupying numerous sites) and suppose measuring co-occurrence with Jaccard similarity ($1 - J$). Then, the transfer-principle for beta (Fig. 3) predicts that co-occurrence is more likely (i.e. JD is low) when sites exhibit uneven presence probabilities. Intuitively, if one species is present mainly in sites 1 and 2 with the other occupying mainly site 3, then co-occurrence is unlikely. This framework can be also be extended to evaluate whether other co-occurrence indices are Schur-convex (Stone and Roberts 1990), potentially clarifying the differences among the co-occurrence indices. Thus, our framework integrates with null model theory, providing a theoretical foundation for co-occurrence analyses while offering predictions that enrich our understanding of species distribution patterns.

Using the two-species case, we showed that the anti-parallel nestedness SIPs have higher beta-diversity than do the parallel SIPs. In the anti-parallel nestedness SIPs (one SIP upward and the other downward) that have a higher beta-diversity than do the parallel SIPs (downward SIPs), one species has a higher chance of occurrence in a site but a lower chance of occurrence in the other site, than does the other species. This means that these species tend to shape the checkerboard incidence patterns and thus have large beta-diversity. In contrast, in the parallel SIPs, one species dominates over the other in the chance of occurrence, likely shaping the dominance incidence patterns (one species is found commonly but the other is totally absent).

In addition, as the steepness of parallel nestedness SIPs gets larger, the likeliness of dominance matrices becomes higher and therefore the expectation of JD gets smaller. On the other hand, as the steepness of anti-parallel nestedness SIPs gets larger, the likeliness of checkerboard matrices gets larger and therefore so does the expectation of JD. As these incidence matrices have been the most prominent patterns in ecology assemblage theory (Diamond 1975, Connor and Simberloff 1979), the present framework suggests that it is promising to explicitly incorporate occupancy heterogeneity into existing theory to better understand beta-diversity patterns. Comparing SIPs with specific incidence matrices can offer insights into spatial incidence patterns, and using this framework may enhance our understanding of heterogeneity's impact on beta-diversity.

Numerous studies have sought to infer community assembly processes from incidence data, sparking debates over their validity (Diamond 1975, Connor and Simberloff 1979). However, Cazelles et al. (2015) and Blanchet et al. (2020) argued that the co-occurrence data provide merely a poor surrogate for community assembly process. Similarly, while we promote using SIPs to visualize beta-diversity patterns, we caution that linking SIPs to specific mechanisms of community assembly is also difficult. Our models are thus better suited to not inferring but incorporating mechanisms and processes of species occurrences (MacKenzie et al. 2018, Chapter 2). An intriguing link between mechanisms and occurrence is in the effect of dispersal. The prevailing view that dispersal induces biotic homogenization derived by mechanistic models (Loreau 2000, Mouquet and Loreau 2003) is recently found to be independent of niche-breadths (Thompson et al. 2020, but see Vannette and Fukami 2017, Lu 2021). If extremely narrow niche breadths result in the independence of incidence (Thompson et al. 2020), the mechanistic models and the present framework may provide a mutually complementary explanation for beta-diversity patterns, if colonization as a form of dispersal increases presence probabilities. Integrating these models with the recent theoretical framework that decomposes beta-diversity into colonization and extinction components (Tatsumi et al. 2021, 2022) could provide a more comprehensive understanding of beta-diversity patterns linking mechanisms and diversity measurements. In this context, the finding that the variance of JD tends to be small (and smaller as the species pool size increases) may suggest that such developed methodologies may be capable of accurately estimating JD.

Applying the method to species distribution models (SDMs) shows that woodpecker distributions in Switzerland will change significantly in the future. Although we assumed the species-independence in the model, joint SDMs (JSDMs) allow for incorporating environmentally mediated non-independence of species occurrences (Pollock et al. 2014). However, JSDMs do not control for the effects of biotic interactions on species' occurrences and are unable to tease apart environmental effects from species interactions (Poggiato et al. 2021). Also, JSDMs and SDMs tend to yield similar predictions of abiotic niches in some systems (Zurell et al. 2020a, Poggiato et al. 2021). Implementing JSDMs to the present application in woodpecker SDMs might therefore produce comparable results. Such results would yield a more convincing insight into conserving and managing the shifts in species distributions in response to global changes.

We adopted the species-independence assumption throughout our analysis (Simberloff 1969). Although this assumption is violated by various factors such as species interactions or environmental covariates, there may be cases where species incidences are independent. For example, the effect of species interactions on community assembly may be minor (Sandal et al. 2022, Halperin 2023). Similarly, disturbance may have dominating effects over competition

on community assembly (Connell and Slatyer 1977, Sousa 1979), in which case the correlation in species incidences may be weak. Examining the effect of species interactions on incidences and spatial biodiversity requires considering the conditional probabilities that some species is present given that the other is (or is not) or the probability of co-occurrence of two species (Supporting information). Our supplementary analyses showed that, for a two-species model, the transfer-principle for beta holds when the species' occurrences are negatively correlated; this prediction is intuitive, since negative association between species may likely generate checkerboard incidence matrices. For more than two species, however, one would have to consider all possible combinations of species incidences for conditional probabilities, as well as parametric constraints (Supporting information). There are various modeling approaches to examine the effect of species interactions, and resulting predictions may be comparable to the present findings for generating hypotheses.

Our model focused on incidence data and fixed a species pool size, but several promising directions exist for extending the framework: 1) abundance data and 2) detection uncertainty. Chao et al. (2004, 2006) have proposed an abundance data-analogue of JD, and combining our method with theirs could allow for addressing numerous questions of beta-diversity patterns: is the transfer-principle for beta applicable to abundance-based JD? What is the relationship between the expression of the derived expectation of JD and abundance-based JD? Can we partition the nestedness component of abundance-based JD further into subcomponents corresponding to anti-parallel and parallel nestedness, and if so, what do these subcomponents tell us (Baselga 2010, 2012)? Another is to consider detection errors (false positives and negatives; MacKenzie et al. 2018, Chapter 4) and undetected species in pursuit of rarefaction and extrapolation (dark diversity; Carmona and Pärtel 2020, Chao et al. 2023). Integrating the uncertainty into the present method would provide a unifying framework for the stochastic theory of biodiversity measurement.

Our study has implications for conservation. Generally, beta-diversity is a key factor for ecosystem functioning from local to global scales (Socolar et al. 2016, Mori et al. 2018). Local ecosystem functioning may be driven by species' functional dissimilarity like niche-differentiations (Godoy et al. 2020). For example, Loiseau et al. (2016) pointed out that conservation management designed to protect taxonomic diversity cannot be fully reconciled with functional diversity management. This result has to do with our prediction that, when two sites are fully homogeneous, the uneven chance of occurrence reduces beta-diversity. Therefore, considering that occupancy heterogeneity is correlated with functional diversity (Palacio et al. 2022), where environmental similarity results in the similarity of presence probabilities of functionally similar species, the present prediction suggests that a conservation management aiming to maintain high beta-diversity be traded- off against the local, functional diversity. This trade-off becomes more complex when the two sites are heterogeneous in species' chances of occurrence. One promising approach to advance the field is thus to identify the condition for the chance of occurrence of species to be balanced (i.e. even). This then produces new interesting questions. Moving forward, open questions include: how does incidence-based beta-diversity respond to changes in functional diversity in colonization ability and extinction tolerance (the capabilities of persistence against disturbances)? How does functional diversity, in turn, respond against the reduction in compositional dissimilarity (biotic homogeneization)?

To conclude, we investigated how beta-diversity varies with the heterogeneity in the chances of occurrence of species in two sites. We analyzed the expectation and variance of the Jaccard dissimilarity index. Using a graphical representation, SIPs, we provided several key predictions for the effect of occupancy heterogeneity on beta-diversity. This work will help researchers to better understand the probabilistic, or stochastic, nature of Jaccard dissimilarity (Baroni-Urbani 1980, Real and Vargas 1996). Future studies may explore the effects of species associations on the probabilistic properties of Jaccard dissimilarity. It is also possible to carry out occupancy dynamics analyses, beyond pairwise dissimilarity analyses (MacKenzie et al. 2018), although careful attention on inferences of biological processes from incidence data is required (Blanchet et al. 2020). A promising approach includes a process-based approach (Thompson et al. 2020, Godsoe et al. 2022, Pilowsky et al. 2022), by which we can incorporate further complications that influence beta-diversity. Our method can incorporate additional realities to track and manage the changes in species distributions under global changes.

Acknowledgements– The authors thank Ryosuke Nakadai, Naoto Shinohara, Akira Terui, and anonymous reviewers for their helpful comments. RI was inspired by some of the questions and answers at Stack Exchange, in calculating the expectation of integer reciprocals. *Funding* – We thank JSPS-KAKENHI for funding (grant no. 19K22457, no. 19K23768, no. 20K15882, no. 24H02291, and no. 24H01528 to RI, and no. 21K14880 to ST). We also thank the Margarita Salas grant funded by the Spanish Ministry of Universities and the 'European Union - NextGenerationEU' to VJO, CRISIS (PGC2018-096577-B-I00) to DA and JAC, UNIQUE (PID2021-127202NB-C21) to DA, and PRIORITY (PID2021-127202NB-C22) to JAC, all funded by MCIN/AEI/10.13039/501100011033 and &ERDF A way of making Europe'. WG was supported by the Te Pūnaha Matatini Centre of Research Excellence.

Author contributions

Ryosuke Iritani: Conceptualization (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Visualization (lead); Writing – original draft (lead). **Vicente J. Ontiveros**: Conceptualization (supporting); Writing – review and editing (supporting). **David Alonso**: Conceptualization (supporting); Formal analysis (supporting); Methodology (supporting); Writing – review and editing (supporting). **José A. Capitán**: Formal analysis (supporting); Methodology

(supporting); Writing – review and editing (supporting). **William Godsoe**: Conceptualization (supporting); Writing – review and editing (supporting). **Shinichi Tatsumi**: Conceptualization (supporting); Investigation (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/ecog.07737.

Data availability statement

Data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.h18931zxv (Iritani et al. 2025).

Supporting information

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

References

Abson, D. 2017. Heterogeneity. – https://doi.org/10.1093/obo/9780199830060-0185.

Albrecht, J. et al. 2021. Species richness is more important for ecosystem functioning than species turnover along an elevational gradient. – Nat. Ecol. Evol. 5: 1582–1593.

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., Sanders, N. J., Cornell, H. V., Comita, L. S., Davies, K. F., Harrison, S. P., Kraft, N. J. B., Stegen, J. C. and Swenson, N. G. 2010. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. – Ecol. Lett. 14: 19–28.

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

Baroni-Urbani, C. 1980. A statistical table for the degree of coexistence between two species. – Oecologia 44: 287–289.

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

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

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

Carmona, C. P. and Pärtel, M. 2020. Estimating probabilistic site-specific species pools and dark diversity from co-occurrence data. – Global Ecol. Biogeogr. 30: 316–326.

Cazelles, K., Araújo, M. B., Mouquet, N. and Gravel, D. 2015. A theory for species co-occurrence in interaction networks. – Theor. Ecol. 9: 39–48.

Chao, A. and Ricotta, C. 2019. Quantifying evenness and linking it to diversity, beta diversity, and similarity. – Ecology 100: e02852.

Chao, A., Chazdon, R. L., Colwell, R. K. and Shen, T.-J. 2004. A new statistical approach for assessing similarity of species composition with incidence and abundance data. – Ecol. Lett. 8: 148–159.

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

Chao, A., Thorn, S., Chiu, C. H., Moyes, F., Hu, K. H., Chazdon, R. L., Wu, J., Magnago, L. F. S., Dornelas, M., Zelený, D., Colwell, R. K. and Magurran, A. E. 2023. Rarefaction and extrapolation with beta diversity under a framework of Hill numbers: the iNEXT.beta3D standardization. – Ecol. Monogr. 93: e1588.

Chase, J. M. et al. 2019. Species richness change across spatial scales. – Oikos 128: 1079–1091.

Chung, N. C., Miasojedow, B., Startek, M. and Gambin, A. 2019. Jaccard/Tanimoto similarity test and estimation methods for biological presence–absence data. – BMC Bioinform. 20: 644.

Colwell, R. K. and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. – Philos. Trans. R. Soc. B 345: 101–118.

Connell, J. H. and Slatyer, R. O. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. – Am. Nat. 111: 1119–1144.

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

Cooley, J. W. and Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. – Math. Comput. 19: 297–301.

Cressie, N., Davis, A. S. and Folks, J. L. 1981. The moment-generating function and negative integer moments. – Am. Stat. 35: 148–150.

Crowther, T. W. et al. 2015. Mapping tree density at a global scale. – Nature 525: 201–205.

Dalton, H. 1920. The measurement of the inequality of incomes. – Econ. J. 30: 348.

D'Amen, M., Mod, H. K., Gotelli, N. J. and Guisan, A. 2017. Disentangling biotic interactions, environmental filters, and dispersal limitation as drivers of species co-occurrence. – Ecography 41: 1233–1244.

Darlington Jr, P. J. 1957. Zoogeography: the geographical distribution of animals. – John Wiley & Sons, pp. 675–675.

Diamond, J. M. 1975. "Assembly of species communities". In: Diamond, J.M. and Cody, M.L., Eds., Ecology and Evolution of Communities. – Harvard Univ. Press, Boston, 342–344.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Godoy, O., Gómez-Aparicio, L., Matías, L., Pérez-Ramos, I. M. and Allan, E. 2020. An excess of niche differences maximizes ecosystem functioning. – Nat. Commun. 11: 4180.

Godsoe, W., Murray, R. and Iritani, R. 2022. Species interactions and diversity: a unified framework using Hill numbers. – Oikos 2023: e9282

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

Gotelli, N. J. and Graves, G. R. 1996. Null models in ecology. – Smithsonian Inst. Press.

Gotelli, N. J. and Ulrich, W. 2011. Statistical challenges in null model analysis. – Oikos 121: 171–180.

Guisan, A., Thuiller, W. and Zimmermann, N. E. 2017. Habitat suitability and distribution models: with applications in R. – Cambridge Univ. Press.

Halperin, T. 2023. Niches and ecological neutrality. – Synthese 202: 69.

Hautier, Y. et al. 2017. Local loss and spatial homogenization of plant diversity reduce ecosystem multifunctionality. – Nat. Ecol. Evol. 2: 50–56.
Heideman, M., Johnson, D. and Burrus, C. 1984. Gauss and the history of the fast Fourier transform. – Arch. Hist. Exact Sci. 34: 265–277.
Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. – Ecology 54: 427–432.
Hui, C. and McGeoch, M. A. 2014. Zeta diversity as a concept and metric that unifies incidence-based biodiversity patterns. – Am. Nat. 184: 684–694.
Iritani, R., Ontiveros, V. J., Alonso, D., Capitán, J. A., Godsoe, W. and Tatsumi, S. 2025. Data from: Jaccard dissimilarity in stochastic community models based on the species-independence assumption. – Dryad Digital Repository, https://doi.org/10.5061/dryad.h18931zxv.
Jaccard, P. 1908. Nouvelles recherches sur la distribution florale. – Bull. Soc. Vaud. Sci. Nat. 44: 223–270.
Jaccard, P. 1912. The distribution of the flora in the alpine zone. – New Phytol. 11: 37–50.
Jost, L. 2010. The relation between evenness and diversity. – Diversity 2: 207–232.
Kalyuzhny, M., Flather, C. H., Shnerb, N. M. and Kadmon, R. 2021. A framework for quantifying deviations from dynamic equilibrium theory. – Global Ecol. Biogeogr. 31: 183–195.
Keil, P., Wiegand, T., Tóth, A. B., McGlinn, D. J. and Chase, J. M. 2021. Measurement and analysis of interspecific spatial associations as a facet of biodiversity. – Ecol. Monogr. 91: e01452.
Koleff, P., Gaston, K. J. and Lennon, J. J. 2003. Measuring beta diversity for presence–absence data. – J. Anim. Ecol. 72: 367–382.
Kvålseth, T. O. 2015. Evenness indices once again: critical analysis of properties. – SpringerPlus 4: 232.
Lange, K. 2010. Applied probability, 2nd edn. – Springer.
Legendre, P. 2019. A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. – Ecol. Evol. 9: 3500–3514.
Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M. and Gonzalez, A. 2004. The metacommunity concept: a framework for multi-scale community ecology. – Ecol. Lett. 7: 601–613.
Loiseau, N., Legras, G., Kulbicki, M., Mérigot, B., Harmelin-Vivien, M. and Mazouni, N. 2016. Multi-component β diversity approach reveals conservation dilemma between species and functions of coral reef fishes. – J. Biogeogr. 44: 537–547.
Loreau, M. 2000. Are communities saturated? On the relationship between alpha, beta and gamma diversity. – Ecol. Lett. 3: 73–76.
Lu, M. 2021. Complex relationships between beta diversity and dispersal in meta-community models. – Ecography 44: 1769–1780.
Lu, M., Vasseur, D. and Jetz, W. 2019. Beta diversity patterns derived from island biogeography theory. – Am. Nat. 194: E52–E65.
MacArthur, R. H. and Wilson, E. O. 1963. An equilibrium theory of insular zoogeography. – Evolution 17: 373.
MacArthur, R. H. and Wilson, E. O. 1967. The theory of island biogeography. Vol. 1. – Princeton Univ. Press.
MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L. and Hines, J. E. 2018. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence, 2nd edn. – Elsevier.
Maggini, R., Lehmann, A., Zbinden, N., Zimmermann, N. E., Bolliger, J., Schröder, B., Foppen, R., Schmid, H., Beniston, M. and Jenni, L. 2014. Assessing species vulnerability to climate and land use change: the case of the Swiss breeding birds. – Divers. Distrib. 20: 708–719.
Magurran, A. E., Dornelas, M., Moyes, F. and Henderson, P. A. 2019. Temporal β diversity – a macroecological perspective. – Global Ecol. Biogeogr. 28: 1949–1960.
Marshall, A. W., Olkin, I. and Arnold, B. C. 1979. Inequalities: theory of majorization and its applications. Vol. 143. – Springer.
Maynard, D. S., Crowther, T. W. and Bradford, M. A. 2017. Competitive network determines the direction of the diversity–function relationship. – Proc. Natl Acad. Sci. USA 114: 11464–11469.
Mori, A. S., Isbell, F. and Seidl, R. 2018. β-diversity, community assembly, and ecosystem functioning. – Trends Ecol. Evol. 33: 549–564.
Mouquet, N. and Loreau, M. 2003. Community patterns in source-sink metacommunities. – Am. Nat. 162: 544–557.
Olden, J. D. and Poff, N. L. 2003. Toward a mechanistic understanding and prediction of biotic homogenization. – Am. Nat. 162: 442–460.
Olden, J. D. and Rooney, T. P. 2006. On defining and quantifying biotic homogenization. – Global Ecol. Biogeogr. 15: 113–120.
Olden, J. D., Comte, L. and Giam, X. 2018. The Homogocene: a research prospectus for the study of biotic homogenisation. – NeoBiota 37: 23–36.
Ontiveros, V. J., Capitán, J. A., Casamayor, E. O. and Alonso, D. 2021. The characteristic time of ecological communities. – Ecology 102: e03247.
Palacio, F. X., Callaghan, C. T., Cardoso, P., Hudgins, E. J., Jarzyna, M. A., Ottaviani, G., Riva, F., Graco-Roza, C., Shirey, V. and Mammola, S. 2022. A protocol for reproducible functional diversity analyses. – Ecography 2022: e06287.
Patterson, B. D. 1987. The principle of nested subsets and its implications for biological conservation. – Conserv. Biol. 1: 323–334.
Patterson, B. D. and Atmar, W. 1986. Nested subsets and the structure of insular mammalian faunas and archipelagos. – Biol. J. Linn. Soc. 28: 65–82.
Peres-Neto, P. R., Olden, J. D. and Jackson, D. A. 2001. Environmentally constrained null models: site suitability as occupancy criterion. – Oikos 93: 110–120.
Pilowsky, J. A., Colwell, R. K., Rahbek, C. and Fordham, D. A. 2022. Process-explicit models reveal the structure and dynamics of biodiversity patterns. – Sci. Adv. 8: eabj2271.
Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S. and Thuiller, W. 2021. On the interpretations of joint modeling in community ecology. – Trends Ecol. Evol. 36: 391–401.
Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A. and McCarthy, M. A. 2014. Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). – Methods Ecol. Evol. 5: 397–406.
Real, R. and Vargas, J. M. 1996. The probabilistic basis of Jaccard's index of similarity. – Syst. Biol. 45: 380–385.
Real, R., Barbosa, A. M. and Bull, J. W. 2016. Species distributions, quantum theory, and the enhancement of biodiversity measures. – Syst. Biol. 66: 453–462.
Rolls, R. J., Deane, D. C., Johnson, S. E., Heino, J., Anderson, M. J. and Ellingsen, K. E. 2023. Biotic homogenisation and differentiation as directional change in beta diversity: synthesising driver–response relationships to develop conceptual models across ecosystems. – Biol. Rev.. 98: 1388–1423.

Routledge, R. D. 1983. Evenness indices: are any admissible? – Oikos 40: 149.
Sandal, L., Grøtan, V., Saether, B.-E., Freckleton, R. P., Noble, D. G. and Ovaskainen, O. 2022. Effects of density, species interactions, and environmental stochasticity on the dynamics of British bird communities. – Ecology 103: e3731.
Simberloff, D. S. 1969. Experimental zoogeography of islands: a model for insular colonization. – Ecology 50: 296–314.
Socolar, J. B., Gilroy, J. J., Kunin, W. E. and Edwards, D. P. 2016. How should beta-diversity inform biodiversity conservation? – Trends Ecol. Evol. 31: 67–80.
Sørensen, T. A. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. – Biol. Skar 5: 1–34.
Sousa, W. P. 1979. Disturbance in marine intertidal boulder fields: the nonequilibrium maintenance of species diversity. – Ecology 60: 1225–1239.
Stone, L. and Roberts, A. 1990. The checkerboard score and species distributions. – Oecologia 85: 74–79.
Tatsumi, S., Iritani, R. and Cadotte, M. W. 2021. Temporal changes in spatial variation: partitioning the extinction and colonisation components of beta diversity. – Ecol. Lett. 24: 1063–1072.
Tatsumi, S., Iritani, R. and Cadotte, M. W. 2022. Partitioning the temporal changes in abundance-based beta diversity into loss and gain components. – Methods Ecol. Evol. 13: 2042–2048.
Thompson, P. L., Guzman, L. M., De Meester, L. D., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S. and Chase, J. M. 2020. A process-based metacommunity framework linking local and regional scale community ecology. – Ecol Lett. 23: 1314–1329.
Tuomisto, H. 2012. An updated consumer's guide to evenness and related indices. – Oikos 121: 1203–1218.
Ulrich, W. and Gotelli, N. J. 2007. Null model analysis of species nestedness patterns. – Ecology 88: 1824–1831.
Vannette, R. L. and Fukami, T. 2017. Dispersal enhances beta diversity in nectar microbes. – Ecol. Lett. 20: 901–910.
Veech, J. A. 2013. A probabilistic model for analysing species co-occurrence. – Global Ecol. Biogeogr. 22: 252–260.
Vellend, M. 2010. Conceptual synthesis in community ecology. – Q. Rev. Biol. 85: 183–206.
Wang, S. et al. 2021. Biotic homogenization destabilizes ecosystem functioning by decreasing spatial asynchrony. – Ecology 102: e03332.
Whittaker, R. H. 1972. Evolution and measurement of species diversity. – Taxon 21: 213–251.
Zurell, D., Zimmermann, N. E., Gross, H., Baltensweiler, A., Sattler, T. and Wüest, R. O. 2020a. Testing species assemblage predictions from stacked and joint species distribution models. – J. Biogeogr. 47: 101–113.
Zurell, D. et al. 2020b. A standard protocol for reporting species distribution models. – Ecography 43: 1261–1277.