Check for updates

Received: 3 February 2021 | Revised: 12 October 2021 | Accepted: 7 January 2022

DOI: 10.1002/ecy.3683

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

# Accounting for temporal change in multiple biodiversity patterns improves the inference of metacommunity processes

**Laura Melissa Guzman**[1,2,3] | **Patrick L. Thompson**[2] | **Duarte S. Viana**[4,5] | **Bram Vanschoenwinkel**[6,7] | **Zsófia Horváth**[8,9,10] | **Robert Ptacnik**[9] | **Alienor Jeliazkov**[4,11,12] | **Stéphanie Gascón**[13] | **Pieter Lemmens**[8] | **Maria Anton-Pardo**[13] | **Silke Langenheder**[14] | **Luc De Meester**[8,15,16,17] | **Jonathan M. Chase**[4,11]

[1]Marine and Environmental Biology Section at the Department of Biological Sciences, University of Southern California, Los Angeles, California, USA

[2]Department of Zoology & Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

[3]Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

[4]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany

[5]Leipzig University, Leipzig, Germany

[6]Department of Biology, Vrije Universiteit Brussel, Brussel, Belgium

[7]Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa

[8]Laboratory of Aquatic Ecology, Evolution and Conservation, KU Leuven, Leuven, Belgium

[9]WasserCluster Lunz - Biologische Station, Lunz am See, Austria

[10]Institute of Aquatic Ecology, Centre for Ecological Research, Budapest, Hungary

[11]Department of Computer Sciences, Martin Luther University, Halle-Wittenberg, Germany

[12]University of Paris-Saclay, INRAE, HYCAR, Antony, France

[13]University of Girona, GRECO, Institute of Aquatic Ecology, Girona, Spain

[14]Department of Ecology and Genetics/Limnology, Uppsala University, Uppsala, Sweden

[15]Leibniz Institut für Gewässerökologie und Binnenfischerei (IGB), Berlin, Germany

[16]Institute of Biology, Freie Universität Berlin, Berlin, Germany

[17]Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany

**Correspondence**
Jonathan M. Chase
Email: jonathan.chase@idiv.de

**Funding information**
Alexander von Humboldt-Stiftung; Deutsche Forschungsgemeinschaft, Grant/Award Number: FZT 118; Fonds Wetenschappelijk Onderzoek, Grant/Award Number: G0B9818; GINOP, Grant/Award Number: 2.3.2.-15-2016-00057;

## Abstract

In metacommunity ecology, a major focus has been on combining observational and analytical approaches to identify the role of critical assembly processes, such as dispersal limitation and environmental filtering, but this work has largely ignored temporal community dynamics. Here, we develop a "virtual ecologist" approach to evaluate assembly processes by simulating metacommunities varying in three main processes: density-independent responses

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Ecology* published by Wiley Periodicals LLC on behalf of The Ecological Society of America.

Interreg V-A Austria-Hungary, Grant/Award Number: Vogelwarte - Madárvárta 2; Killam Trusts; KU Leuven, Grant/Award Number: C16/2017/002; Magyar Tudományos Akadémia, Grant/Award Number: János Bolyai Research Scholarship; Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: CGS-D PDF; Nemzeti Kutatási Fejlesztési és Innovációs Hivatal, Grant/Award Number: FK-132095; Österreichische Forschungsgemeinschaft, Grant/Award Number: 06/15539; sDiv, Grant/Award Number: sTURN working group; Simon Fraser University; University of British Columbia, Grant/Award Number: Four year fellowship

**Handling Editor:** Kathryn L. Cottingham

to abiotic conditions, density-dependent biotic interactions, and dispersal. We then calculate a number of commonly used summary statistics of community structure in space and time and use random forests to evaluate their utility for inferring the strength of these three processes. We find that (i) both spatial and temporal data are necessary to disentangle metacommunity processes based on the summary statistics we test, and including statistics that are measured through time increases the explanatory power of random forests by up to 59% compared to cases where only spatial variation is considered; (ii) the three studied processes can be distinguished with different descriptors; and (iii) each summary statistic is differently sensitive to temporal and spatial sampling effort. Including repeated observations of metacommunities over time was essential for inferring the metacommunity processes, particularly dispersal. Some of the most useful statistics include the coefficient of variation of species abundances through time and metrics that incorporate variation in the relative abundances (evenness) of species. We conclude that a combination of methods and summary statistics is probably necessary to understand the processes that underlie metacommunity assembly through space and time, but we recognize that these results will be modified when other processes or summary statistics are used.

**KEYWORDS**
metacommunity ecology, random forests, simulation study, spatiotemporal dynamics, summary statistics, variation partitioning

## INTRODUCTION

Metacommunity ecology lies at the interface of local coexistence, which results from species traits and interactions, and spatial processes such as habitat heterogeneity and dispersal. As such, metacommunity ecology provides a powerful framework from which to understand the composition, diversity, and abundances of species and how they vary in space and time (Leibold & Chase, 2017; Thompson et al., 2020). While the theory of metacommunities is thriving, empirical work has lagged. Most notably, empirical tests of two alternative processes have been emphasized in the literature. First, species interactions and environmental filtering have dominated "niche-based" thinking (e.g., Chase & Leibold, 2003; Tilman, 1982), where, for example, if patterns of species composition are well matched to environmental variation in the landscape, niche-based processes are more likely to have underlain those patterns. Second, aspects of stochasticity and dispersal limitation are more associated with "neutral-based" perspectives (e.g., Hubbell, 2001; O'Dwyer & Cornell, 2018).

A large body of empirical work in metacommunity ecology has focused on using statistical analyses of natural patterns of species diversity and composition to identify signatures of the underlying processes (e.g., niche vs. neutral) in structuring community assembly (Cottenie, 2005; Ovaskainen et al., 2019; Soininen, 2014, 2016). Unfortunately, it is now clear that inference-based analyses can suffer from a number of statistical biases (e.g., Clappe et al., 2018; Gilbert & Bennett, 2010) and that analyses of metacommunities at a single point in time (snapshot) are often insufficient to differentiate among multiple ecological processes (Jabot et al., 2020). These findings have called into question the results of a number of empirical metacommunity studies and syntheses thereof. Instead, new tools, concepts, and approaches are likely necessary to make progress in understanding the underlying processes that influence metacommunity assembly.

Although metacommunity ecologists have continued to develop more refined analytical tools (Clappe et al., 2018; Münkemüller et al., 2012; Ovaskainen et al., 2017, 2019; Peres-Neto & Legendre, 2010; Viana et al., 2021), most approaches rely on (i) analyses of one or only a few metrics describing community structure, such as the shape of the species abundance distribution or the proportion of variation explained by spatial or environmental factors; (ii) static (snapshot) patterns of these metrics over a relatively short time period. We address both of

these topics by explicitly considering multiple patterns simultaneously (Brown et al., 2016; May et al., 2015; Ovaskainen et al., 2019) and by emphasizing the importance of temporal dynamics.

Here, we used a process-based approach to explore the ability of multiple metrics and analyses to distinguish among metacommunity processes when both temporal and spatial patterns of species abundances are available (Figure 1). To do so, we started by taking a "virtual ecologist" approach (sensu Zurell et al., 2010) by simulating metacommunity dynamics and varying the strengths of three key processes emphasized in Thompson et al. (2020): (1) density-independent responses to abiotic conditions, modeled as the range of environmental conditions that allow for positive growth of organisms (i.e., the fundamental abiotic niche); (2) density-dependent biotic interactions, modeled as the per-capita effect of one species on another, usually competition for resources; and (3) dispersal, modeled via the emigration and immigration of individuals from one patch to another. All processes were implemented simultaneously, and each process had a stochastic component (Shoemaker et al., 2020) (Figure 1—arrow i, Box 1).

To capture the dynamic nature of metacommunity patterns, we estimated spatial and temporal variation in patterns of relative abundance, diversity, and composition using summary statistics, including patterns and model-based statistics, calculated from the simulated metacommunities (Figure 1—arrow ii). We then used a suite of these summary statistics in random forest models to identify those most informative for distinguishing variation in metacommunity processes (Figure 1—arrow iii). As a final step, we examined the influence of sampling effort in time and space.

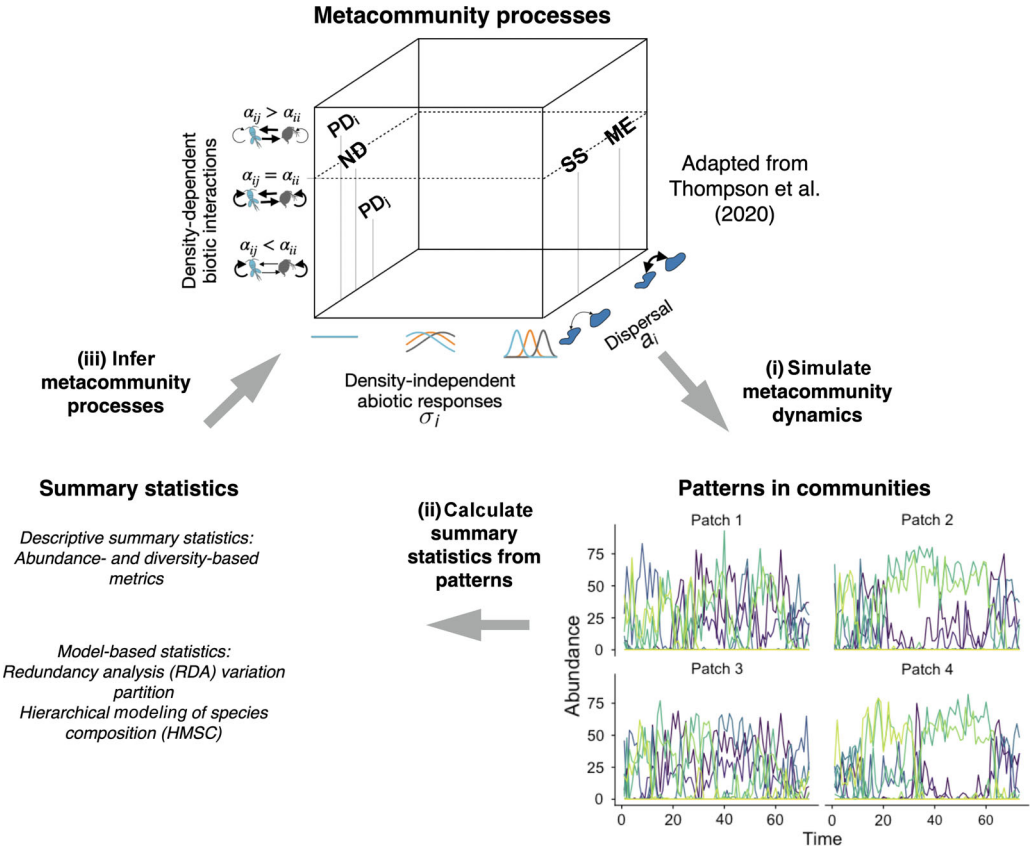


**FIGURE 1** Workflow for finding summary statistics that distinguish patterns resulting from different metacommunity processes. (i) Simulate metacommunity dynamics: Different combinations of metacommunity processes can be achieved by changing the values of three key continuous model parameters—abiotic niche breadth ($\sigma_i$), competition strengths ($\alpha_{ij}$), and probability of dispersal ($a_i$). Each of the three processes is included simultaneously, but the values of each parameter vary. Stochasticity is incorporated into each process through probability distributions. Different combinations of these parameters will result in different time series of species abundances. (ii) Calculate summary statistics from pattern: For each time series we calculate summary statistics that summarize metacommunity dynamics (for the full list see Table 1). (iii) Infer metacommunity processes: Across all simulations and parameter combinations, use random forests to identify the best summary statistics for distinguishing the model parameters. ME, mass effects; ND, neutral dynamics; PD, patch dynamics ($PD_i$ for species $i$ and $PD_j$ for species $j$); SS, species sorting

**BOX 1 Model**

The dynamics of the metacommunity are governed by a Beverton-Holt model (Beverton & Holt, 1957) growth dynamics with generalized Lotka-Volterra competition:

$$N_{ix}(t+1) = \hat{N}_{ix}(t+1) - E_{ix}(t) + I_{ix}(t), \tag{1}$$

where $\hat{N}_{ix}(t+1)$ is the population size at time $t$ before dispersal.

$$\hat{N}_{ix}(t+1) = Poisson\left(max\left\{N_{ix}(t)\frac{r_{ix}(t)}{1+\sum_{j=1}^{S}\alpha_{ij}N_{jx}(t)}, 0\right\}\right), \tag{2}$$

where $N_{ix}(t)$ is the population size of species $i$ in patch $x$ at time $t$, $\alpha_{ij}$ is the per capita interaction effect of species $j$ on species $i$, and $S$ is the total number of species. Stochasticity in local demographic outcomes is incorporated through the Poisson draw. $r_{ix}(t)$ is the density-independent growth rate of species $i$ in patch $x$ at time $t$ which is given by

$$r_{ix}(t) = e^{-\left(\frac{z_i - env_x(t)}{2\sigma_i}\right)^2}, \tag{3}$$

where $z_i$ is the environmental optimum of species $i$, $env_x(t)$ is the environmental condition in patch $x$ at time $t$, and $\sigma_i$ is the abiotic niche breadth of species $i$. The number of emigrants $E_{ix}(t)$ was determined as the outcome of $N_{ix}(t)$ draws of a binomial distribution, each with a probability equal to $a_i$. $\hat{I}_{ix}(t)$ is the probability that a dispersing individual of species $i$ immigrates to patch $x$ at time $t$ and it is given by

$$\hat{I}_{ix}(t) = \frac{\sum_{y\neq x}^{M} E_{iy}(t)e^{-L_i d_{xy}}}{\sum_{x=1}^{M} E_{ix}(t)}, \tag{4}$$

where $M$ is the total number of patches, $E_{iy}(t)$ is the number of immigrants of species $i$ from another patch $y$, $L_i$ is the strength of exponential decrease of dispersal with distance and $d_{xy}$ is the distance between patches $x$ and $y$.

We found that none of the summary statistics examined was able to fully distinguish the different simulated metacommunity processes or serve as a reliable indicator of metacommunity dynamics. Instead, typically several metrics and analytical approaches were necessary to evaluate the importance of each process. Further, we showed that most inferences were strongly influenced by the number of time points and patches used in analyses, and we concluded that for the breadth of metacommunity dynamics reproduced by our model, insufficient sampling effort in time and space may hinder our ability to infer metacommunity processes in empirical studies.

# METHODS

## Simulations

Our goal with the simulations was to produce time series of metacommunity dynamics from which we could apply various metrics and analytical approaches intended to retrieve the underlying processes. Our simulation varied the three core processes—abiotic niche breadth, density-dependent biotic interaction structure, and dispersal—developed in Thompson et al. (2020) (overviewed in Box 1). While simulation models are necessarily limited, this metacommunity model allowed us to explore a wider

array of dynamics compared to earlier models. Simulated metacommunities were composed of 100 patches with coordinates randomly drawn from a uniform distribution and a starting regional richness of 50 species. Species initial abundances were randomly drawn from a Poisson distribution with $\lambda = 0.5$, which resulted in spatial variation in occupancy and initial abundances. Populations were reseeded every 10 time steps over the first 100 time steps to ensure that all species were introduced to each patch at least once, but with variation in their order of arrival. We ran each simulation for 2200 time steps, including an initialization where the environment was held constant (200 time steps) and a burn-in period (800 time steps); these were discarded. Following the burn-in period, we retained every 20th time step, leaving a total of 60 retained time steps per simulation. The timescale in the simulations was somewhat arbitrary and its relation to empirical communities would depend on the generation time of the organisms in question.

To change the density-independent responses to abiotic heterogeneity, we adjusted the abiotic niche breadth parameter ($\sigma_i$) from weak (10) to strong (0.01) (values were evenly distributed on a natural log scale). To change the density-dependent biotic responses, we adjusted the inter- and intracompetition strengths in four different scenarios (for details see Thompson et al., 2020): equal competition ($\alpha_{ij} = \alpha_{ii}$), stabilizing competition ($\alpha_{ij} < \alpha_{ii}$), mixed competition ($\alpha_{ij}$ can be less than or greater than $\alpha_{ii}$), and competition–colonization trade-off (30% of species are competitively dominant [$\alpha_{ij} > \alpha_{ii}$], but their dispersal rates are an order of magnitude lower, and 70% of species are subdominant and have stabilizing competitive interactions). Finally, to change dispersal rates, we adjusted the probability of dispersal parameter ($a_i$), ranging from metacommunities that were effectively disconnected (0.00001) to fully connected (0.464) (values were evenly distributed on a natural log scale). The abiotic environment in any given patch varied between 0 and 1 and was spatially and temporally autocorrelated.

We ran 660 combinations of dispersal rate, abiotic niche breadth, and competition scenarios for 15 randomly generated replicate landscapes that varied in the location of the patches and the environmental conditions. Landscapes were created by drawing *xy* coordinates in geographic space for each habitat patch from 1:100 and converting these into a torus to avoid edge effects. This resulted in 9900 simulation runs, but we only considered the 7880 runs where at least 10 out of 15 replicates for a parameter combination resulted in the persistence of at least 2 species. This ensured that poorly replicated parameter combinations did not drive the patterns.

## Summary statistics

We calculated 85 summary statistics from each simulation run. These included descriptive statistics (e.g., diversity metrics) as well as model-based statistics commonly used in metacommunity ecology (e.g., variation partitioning fractions) (see Table 1 for the main description of the statistics and Appendix S1: Table S1 for a full tally of all the statistics we used).

First, we calculated statistics that describe patterns of species' relative abundances (i.e., relative species commonness and rarity) based on Hill numbers (Chao et al., 2014; Hill, 1973). ${}^{2}D$ or Hill number 2 (hereafter referred to as Hill 2) is the inverse Simpson's diversity, ${}^{1}D$ (Hill 1) is the exponent of Shannon's diversity, and ${}^{0}D$ (Hill 0) is species richness. We used two main ratios, Hill 1/Hill 0 and Hill 2/Hill 0, to capture aspects of evenness in the metacommunity. When a community is exactly even, Hill 2 or Hill 1 approximates Hill 0, and therefore the ratio of these values is close to 1; their ratios will be smaller when the community is uneven. Second, we calculated the coefficient of variation (CV) of local and regional abundance across all species through time as a standardized index of variation (Loreau et al., 2003). Third, we calculated both spatial and temporal beta diversity of the Hill numbers as the ratio of the Hill number at the smaller, alpha scale (a given spatial or temporal sample) to the larger, gamma scale (the sum of all spatial or temporal samples). We also decomposed these beta measures into richness and replacement components (Podani & Schmera, 2011). Species replacement indicates the turnover of species among samples, for example, due to environmental filtering or competition (Legendre, 2014), whereas richness differences may reflect differential coexistence in different locations or dispersal limitation independent of species replacement (Schmera et al., 2020). Finally, we calculated the proportion of patches occupied for each species and then calculated the species' mean, minimum, and maximum occupancy. Low occupancy can indicate dispersal limitation or the strength of competition, whereas high occupancy could point to mass effects (Ehrlén & Eriksson, 2000).

For the model-based statistics, we first used redundancy analysis (RDA)-based variation partitioning (Borcard et al., 1992) to quantify compositional variation explained by (i) the environment and space using the final time point and (ii) environment, space, and time across the entire time series. We modeled the spatial component using distance-based Moran's eigenvector maps (MEMs) calculated across all patches using the *dbmem* function from the *adespatial* package in the R programming language (Dray et al., 2019). We used all MEMs that capture positive spatial autocorrelation in the

**TABLE 1** Main types of summary statistics used as descriptors of metacommunity dynamics

| Name | Equation and parameters | Variants | Predictors and outputs | Reference |
|---|---|---|---|---|
| **A—Descriptive statistics** | | | | |
| Abundance and occupancy ratios—local scale (α) metacommunity scale (γ) | Hill numbers given by ${}^qD = \left(\sum_{i=1}^{S} p_{ix}^q\right)^{\frac{1}{1-q}}$, and ratios given by ${}^1D/{}^0D$ (Hill 1/Hill 0) and ${}^2D/{}^0D$ (Hill 2/Hill 0). $p_{ix}$ is the relative abundance of species $i$ in patch $x$. $S$ is the total no. species. $q$ is sensitivity to relative frequencies. | 1. Abundance ratios—local scale.<br>2. Occupancy ratios—local scale where $p_i$ is relative occupancy.<br>3. Abundance ratios—metacommunity scale. | ... | (Hill, 1973) |
| Coefficient of variation in community abundance | $\mathrm{CV} = \mathrm{CV}_t\left(\sum_{i=1}^{S} N_i\right)$ $N_i$ is abundance of species $i$. $S$ is total no. species. | 1. Local scale.<br>2. Metacommunity scale. | ... | (Tilman, 1995) |
| Beta diversity | ${}^qD_\beta = \frac{{}^qD_\gamma}{{}^qD_\alpha}$ ${}^qD_\gamma$ is diversity at the metacommunity scale and ${}^qD_\alpha$ is at local scale. $q$ was 0, 1 or 2. | 1. Spatial beta diversity.<br>2. Temporal beta diversity. | ... | (Tuomisto, 2010) |
| Beta diversity decomposition using Jaccard-based indexes | $\mathrm{BD}_{\mathrm{Total}} = \sum_{h=1}^{n-1}\sum_{i=h+1}^{n} D_{hi}/[n(n-1)]$ $\mathrm{BD}_{\mathrm{Total}} = \mathrm{Repl}_{\mathrm{Total}} + \mathrm{RichDiff}_{\mathrm{Total}}$ $D_{hi}$ is dissimilarity between sites $h$ and $i$. | 1. Beta diversity decompositions in space.<br>2. Beta diversity decompositions in time. | ... | (Legendre, 2014) |
| Proportion of patches occupied across species | $p_i = s_i/s_{\mathrm{Total}}$ $s_i$ is no. patches occupied by species $i$ and $s_{\mathrm{Total}}$ is total no. patches sampled. | 1. Mean proportion of patches occupied across species.<br>2. Minimum<br>3. Maximum | ... | (Ehrlén & Eriksson, 2000) |
| **B—Model-based statistics** | | | | |
| Variation partitioning only in space | ... | ... | Predictors: Moran's eigenvector maps (MEMs), environment<br>Output: Variation explained by space (S) and environment (E) (and shared (S∩E)) for final time step | (Borcard et al., 1992) |
| Variation partitioning in time and space | ... | ... | Predictors: MEMs, Environmental variables, asymmetric eigenvector maps<br>Output: Variation explained by space (S), time (T) and environment (E), and shared (S∩E, S∩T, T∩E, S∩E∩T) | (Blanchet et al., 2008) |
| Hierarchical modeling of species composition | ... | ... | Predictors: Fixed effects: Environment (second-degree polynomial) Random effects: Space × time, Space, Time<br>Output: Variation explained by time, environment, space, space-time, and species associations. | (Ovaskainen et al., 2017) |

analysis, because selecting specific MEMs does not fully account for spatial autocorrelation in the residuals (Peres-Neto & Legendre, 2010). While using a selection procedure of MEMs can yield better results for individual simulation runs, we kept the number of MEMs consistent across simulation runs for comparability. We modeled the temporal component using asymmetric eigenvector maps (AEMs) using the *aem.time* function from the *adespatial* package (Blanchet et al., 2008; Dray et al., 2019). We used the percentage of variation explained by each component in the random forest, regardless of its statistical significance.

Second, we used hierarchical modeling of species communities (HMSC) (Ovaskainen et al., 2017), which is a hierarchical Bayesian joint species distribution model that uses fixed environmental predictors and spatiotemporal random effects to make community-level inferences. We used species abundance with an assumed Poisson distribution as the response variable. The environmental variable was included as a fixed linear and quadratic effect to fit the Gaussian-shaped response of species to the abiotic environment. The spatial and temporal structures of the data were modeled through autocorrelated spatial, temporal, and spatiotemporal random factors. The spatial random effect was modeled using the $x$ and $y$ coordinates. The temporal random effect was modeled using time step as a random temporal coordinate. The spatiotemporal random effect was modeled using the $x$, $y$, and time coordinates (for further details see Appendix S1). The summary statistics we used in the random forest consisted of partitioned explained variation according to all fixed and random effects: environment, space, time, space–time. We used both the raw variation fractions and fractions standardized by total amount of variation explained to account for differences in the amount of residual variation across simulations. We also used the estimates of species associations aggregated into the proportion of positive or negative associations (Ovaskainen et al., 2019). HMSC was run across 4 Markov chain Monte Carlo (MCMC) chains, each with 1000 samples and a transient period of 5000 steps. While this length of MCMC sampling does not lead to full convergence, it was sufficient to provide estimates of the summary statistics that we used in our analysis (Appendix S1: Figure S10). HMSC was implemented using the HMSC-R package in R (Tikhonov et al., 2020).

### Random forests

We used random forest modeling (Breiman, 2001) to assess the performance of the different summary statistics to distinguish the simulated metacommunity processes. Random forests have a number of useful characteristics for our purpose: (i) they can be used for both regression and classification problems (note that dispersal and abiotic niche breadth parameters are continuous, while the competition scenario is categorical); (ii) they are non-parametric, allowing any relationship between predictor and response; (iii) they can be used to rank the importance of the predictors; and (iv) they can deal well with correlated predictors (Breiman, 2001).

To determine which set of summary statistics were better at predicting each of the three metacommunity processes, we divided the summary statistics into seven different groups according to their characteristics (e.g., descriptive vs. model-based) (Table 2) and used them as predictors of each metacommunity process, resulting in a total of 21 random forest models (with 500 trees each) (Table 2). The variable importance is determined by observing how much the prediction error increases when each variable is permuted, while the other variables remain the same (Breiman, 2001; Liaw & Wiener, 2002).

Because some simulations were discarded (see preceding discussion), we analyzed 6596 simulations (Table 2) for the simple metacommunity descriptors. In addition, some of the model-based statistics could not be calculated for some communities, for example, we could not calculate covariances of transient species that occurred in only one time step. As a result, the numbers of simulations used in the random forest for the HMSC (5728) and all summary statistics (5185) were lower.

The minimal model was selected using recursive feature elimination and a 10-fold cross-validation procedure that partitioned the data into test and training, where the partition was random across the entire set of simulations. The model was fit with all of the predictors to the training data and tested with the held-back samples, where each predictor received an importance value. The algorithm then kept the "n" most important variables, refit the model, and tested it again with the held-back samples. This procedure was repeated for 10 to 50 "n" predictors (Appendix S1: Figure S1). The algorithm determined the best number of predictors and the best predictors based on the prediction accuracy (with the held-back samples). The whole procedure was repeated 10 times (Ambroise & McLachlan, 2002; Svetnik et al., 2004).

## Sensitivity analysis to spatial and temporal sampling effort

We calculated the summary statistics on the "full" time series, except for the "spatial-only" variation partitioning, which was calculated only on the last time point. The full time series was obtained after burn-in and the

**TABLE 2** Different sets of summary statistics were more or less informative at explaining the variance in each of the three metacommunity processes. We used random forests to compare the performance (classification success/$R^2$) of sets of summary statistics. The higher the classification success or the $R^2$, the better are the sets of summary statistics at explaining variation in each of the three metacommunity processes: Density-independent responses to abiotic conditions (abiotic responses), density-dependent biotic interactions (biotic interactions), and dispersal. We ran 7 different random forests (each with a different subset of summary statistics) for each of the 3 processes for a total of 21 random forests. The minimal model kept the most important summary statistics for each metacommunity process through a recursive backwards elimination algorithm. The minimal model for each metacommunity process can contain different summary statistics

| Explanatory model | Summary statistics used as predictors | No. simulations | Biotic interactions | Dispersal | Abiotic responses |
|---|---|---|---|---|---|
| 1—Spatial descriptive statistics | Spatial statistics in Table 1A | 6596 | 69.78% | 57.51% | 32.48% |
| 2—Temporal and spatial descriptive statistics | All statistics in Table 1A | 6596 | 83.29% | 85% | 48.39% |
| 3—Variation partitioning (VP) in space | Table 1B—VP in space | 5613 | 32.99% | 0% | 35.6% |
| 4—Variation partitioning in space and time | Table 1B—VP in time and space | 6590 | 49.71% | 59.05% | 49.14% |
| 5—Hierarchical modeling of species composition (HMSC) | Table 1B—HMSC | 5728 | 78.6% | 63.85% | 51.33% |
| 6—All summary statistics | All summary statistics (Table 1) | 5185 | 90.24% | 85.88% | 61.78% |
| 7—Minimal model | Subsets of model 6 | 5185 | 90.55% | 86.77% | 62.49% |

subsampling regime, so the full time series was actually only 5% of the entire simulation.

Empirical data are inevitably limited in space and time. We therefore investigated the effect of sample limitation by repeating our analyses on a subset of patches and time points, as would be the case in most empirical studies. For this question, we used simulations that had stabilizing competition ($\alpha_{ij} < \alpha_{ii}$), intermediate levels of abiotic niche breadth (4.64159), and intermediate levels of dispersal (0.00215), which yielded a gamma diversity above one. Then we randomly sampled $m$ of 100 patches where $m = 4, 8, 12, 16, ..., 100$, repeated 1000 times. We also sampled $t$ of 60 time points, where $t = 4, 8, 12, 16, ..., 60$; this was not repeated because time points are sequential. We subsampled time and space factorially for every combination of number of patches and time points. To calculate the effect of missing time points or patches, we calculated the so-called error half-life. The error half-life is the minimum number of patches or sequential time points needed to reduce the "error" in the summary statistic by half.

# RESULTS

## Comparing explanatory random forest Models 1–7

For the simulation scenarios explored in this study, we found that both spatial and temporal dynamics were necessary to distinguish metacommunity dynamics employing the metrics used in this study. Including statistics that are measured through time increased the explanatory power of the random forests by up to 59% compared to cases where only spatial variation was considered, namely, when temporal variation was incorporated into the descriptive statistics (Table 2—Model 1 vs. Model 2) and in the RDA-based variation partitioning (Table 2—Model 3 vs. Model 4).

We also found that several summary statistics were complementary and captured different aspects of metacommunity dynamics. For instance, the most important summary statistics for inferring variation in dispersal rates were the ratio of occupancies at a local scale (Hill 1/Hill 0), the total richness difference in time, and the variation partitioning time ($T$) component. Of these, the ratio of occupancies had a unimodal relationship with dispersal, richness differences in time had a decreasing relationship with dispersal, and the variation explained by time ($T$) increased with dispersal (Appendix S1: Figures S5–S7). In all, we found that including descriptive and model-based summary statistics increased the explanatory power by up to 22% (Table 2—Models 6 and 7).

We found that descriptive statistics were more informative than model-based statistics for detecting variation in dispersal and density-dependent biotic interactions, even with fewer predictors (Table 2—Model 2 vs. Model 5). The descriptive statistics in Model 2, such as variation in community biomass through time and the ratios of occupancy (evenness), captured more variation in metacommunity

dynamics than RDA or HMSC (Figure 2). Measures of temporal variability and the relative abundances of species were also important. Furthermore, these statistics were more informative when combined with other statistics, such as beta diversity (Appendix S1: Table S1).

Model-based statistics tended to be more informative when including time or species interactions. RDA-based variation partitioning using only space and environment, but no time (Model 3), had the lowest performance to distinguish metacommunity dynamics (Table 2). Once we included temporal dynamics, the RDA-based variation partitioning increased the classification success for different types of biotic interactions to 50% and explained more variation in dispersal (59%) and responses to abiotic conditions (49%) (Table 2—Model 4). The amount of variation explained in dispersal, an explicitly spatiotemporal process, was particularly sensitive to the inclusion of time, increasing from 0% to 59%. Likewise, the random forests for biotic interactions and abiotic responses were more successful when we included time. HMSC, which included time and summary statistics of species interactions, had a greater predictive power of metacommunity processes than RDA with time (Table 2—Model 5 vs. Model 4). The best improvement was in the classification of biotic interactions (success rate 79% for Model 5 vs. 50% for Model 4).

Including all descriptive and model-based statistics together (Model 6) yielded better performance for all three simulated metacommunity processes than using one type of statistic. Several statistics provided complementary information that was useful for distinguishing metacommunity processes (Table 2). However, this random forest included up to 85 predictors (all predictors in Appendix S1: Table S1), and very little predictive ability was lost when we reduced it to a minimal model through backwards selection (Table 2). These minimal models included a different subset of statistics depending on the inferred process (Figure 2). The most informative statistics had the smallest variance at each parameter value but exhibited the most variation across parameter space (Figure 3a,e,i vs. Figure 3c,d,h).

### Density-dependent biotic interactions

The minimal model to distinguish different types of biotic interactions had a classification success of 91%. Stabilizing competition was easiest to distinguish with the available summary statistics (3% error rate), while the competition–colonization trade-off was the most challenging to separate from the other types (20% error rate). The most important summary statistic that helped to distinguish biotic interactions was the CV of abundance at the local scale (Figure 2 and Appendix S1: Figure S2). This was lowest under stabilizing competition, increased with equal competition and mixed competition, and was highest with competition–colonization trade-offs (Figure 3a). Under stabilizing competition, local diversity was higher, whereas the increased variability in the other scenarios was due to the stronger competitive effects (Chesson, 2000; Thompson et al., 2020).

### Dispersal

The minimal model explained 87% of the variation in dispersal in the simulated metacommunities (Table 2). Here, our measure of community evenness, the occupancy ratio, emerged as the most important summary statistic (Figure 2 and Appendix S1: Figure S3). This statistic had a U-shaped relationship with dispersal (Figure 3e). Low dispersal resulted in a relatively high evenness, suggesting that most species occupied a relatively similar number of patches, despite variation. At intermediate dispersal, species effectively tracked their niches, and their abundance varied depending on existing environmental conditions (i.e., strong species sorting). High dispersal again resulted in a relatively high evenness, suggesting that most species were distributed evenly across the landscape due to mass effects (Figure 3e).

## Density-independent responses to abiotic conditions

The minimal model to explain variation in density-independent environmental filtering captured only 62% of the variation. The most important summary statistic for inferring the responses to abiotic conditions was the environmental (E) component of variation partitioning through time and space (Figure 2 and Appendix S1: Figure S4). This component had a hump-shaped relationship with the strength of abiotic conditions (Figure 3i). When species responses to abiotic variation were strong (i.e., $\sigma_i$—niche breath—is small), temporal variation in the environment reduced abundances and species richness, allowing stochasticity to play a large role such that the environmental (E) component explained relatively little variation in community composition. When responses to the abiotic conditions were weaker (i.e., intermediate $\sigma_i$), the environmental (E) component explained the most variation in community composition because species could respond to environmental variation by moving to more suitable patches. Finally, when the responses to abiotic conditions were very weak (i.e., $\sigma_i$ is large), the

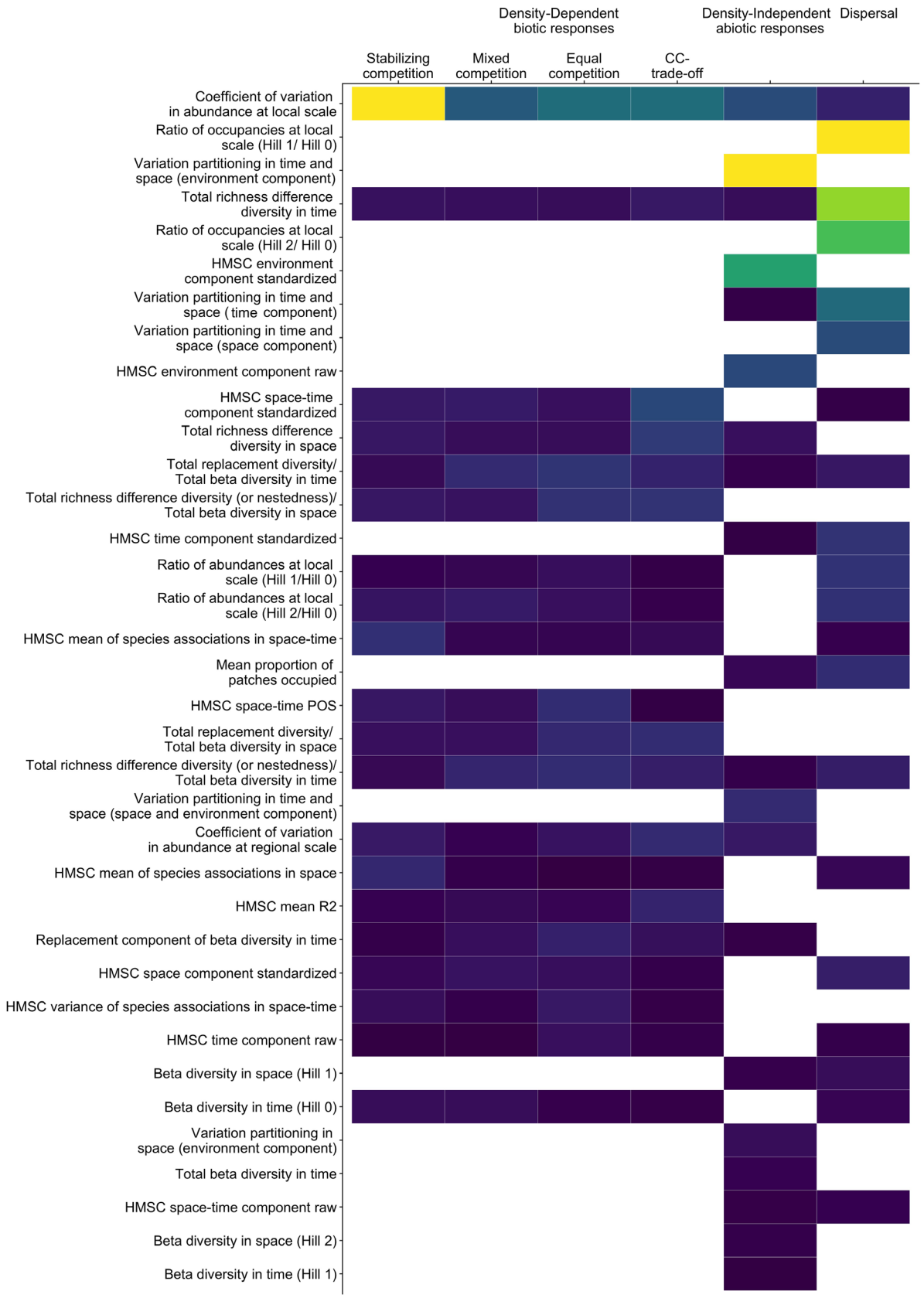


**FIGURE 2** Legend on next page.

amount of variation explained by the environmental (E) component was again low, because variation in environmental conditions did not lead to changes in community composition (Figure 3i).

## Sensitivity analysis

When we reduced sample completeness to be more in line with what is more typical in empirical studies, the summary statistics were not equally affected (Figure 4). We evaluated the error half-life of the summary statistics when time was fully sampled but patches were not, and vice versa. Some summary statistics were more sensitive to the loss of sampled patches, while some were more sensitive to reduced coverage in time. When time was fully sampled (which will be somewhat subjective in empirical studies), total beta diversity in time, beta diversity in space (Hill 1 and 2), compositional replacement in time, and richness differences in time, as well as all of the environmental (E) and shared components of spatiotemporal variation partitioning, were robust to a reduced number of sampled patches. These statistics needed less than 8% of patches sampled to reduce the error rate by half. On the other hand, the spatial (S) component of variation partitioning was very sensitive to the loss of patches and reached half the error at 72% of patches remaining (Appendix S1: Table S2, Figure S8). When space was fully sampled, minimum and maximum proportion of patches occupied, the space–time component (S∩T) of variation partitioning, beta diversity in space and time (Hill 1 and 2), and the CV of abundance at the local scale were robust to the loss of time points (i.e., only 8% of time points are needed to reduce the error rate by half). On the other hand, the space–environment shared component (S∩E) of variation partitioning and replacement and richness differences through time needed more than 80% of time points to reduce the error rate in half (Appendix S1: Table S3, Figure S9).

Some statistics were sensitive to the loss of both time points and patches, as measured by the difference between values calculated over the entire space or time (i.e., true values) and values calculated from undersampled patches or time points (i.e., sample values). For example, the CV of abundance at a local scale, which is critical for detecting biotic interactions, was not very sensitive to the number of time points, but the difference between true and sampled values decreased as more patches were sampled (Figure 4a). Richness differences through time were highly sensitive to the number of time points sampled, becoming more accurate as more time points were sampled, but less sensitive to the number of patches sampled. However, the estimates became more precise as more patches were sampled (Figure 4b). Finally, the mean proportion of patches used by species was more sensitive to the loss of sampled patches than time points. When the entire 100-patch metacommunity was sampled, 30% of the time points were sufficient to produce reliable estimates, whereas when few patches were sampled, even complete temporal sampling led to large discrepancies between estimated and true values (Figure 4c).

# DISCUSSION

Overall, we found that accounting for temporal dynamics was necessary to distinguish simulated metacommunity processes with the summary statistics included in this study. This suggests that neither a snapshot of communities in a single time point nor substituting space for time in observational studies is sufficient to quantify the underlying processes in metacommunities. While this may not seem particularly surprising since metacommunities are clearly dynamic in time, the vast majority of empirical metacommunity studies largely ignore the role of time. Empirical findings also support the need to account for temporal dynamics in metacommunity ecology (Tatsumi et al., 2020).

Indeed, we found that RDA-based variation partitioning, one of the most popular methods used to distinguish metacommunity-level processes, had generally poor performance at distinguishing metacommunity processes. This is likely because variation partitioning does

**FIGURE 2** The minimal model kept the most important summary statistics for each metacommunity process through a recursive backwards elimination algorithm. In random forests, variable importance is determined by observing how much the prediction error increases when each variable is permuted. The best number of predictors was 50 for the density-dependent biotic interactions, 32 for dispersal, and 29 for density-independent responses to abiotic conditions, but here we show the 20 best performing summary statistics for each process. The minimal model for each metacommunity process contains different sets of summary statistics (summary statistics not shared between processes are blank). The 20 best-performing summary statistics for each process are ordered from top (yellow—most important) to bottom (purple—less important) from each minimal model (Model 7). CC trade-off refers to the scenario of competition–colonization in the strength of biotic responses. For the definition of all summary statistics see Table 1 and Appendix S1: Table S1. HMSC refers to hierarchical modeling of species communities

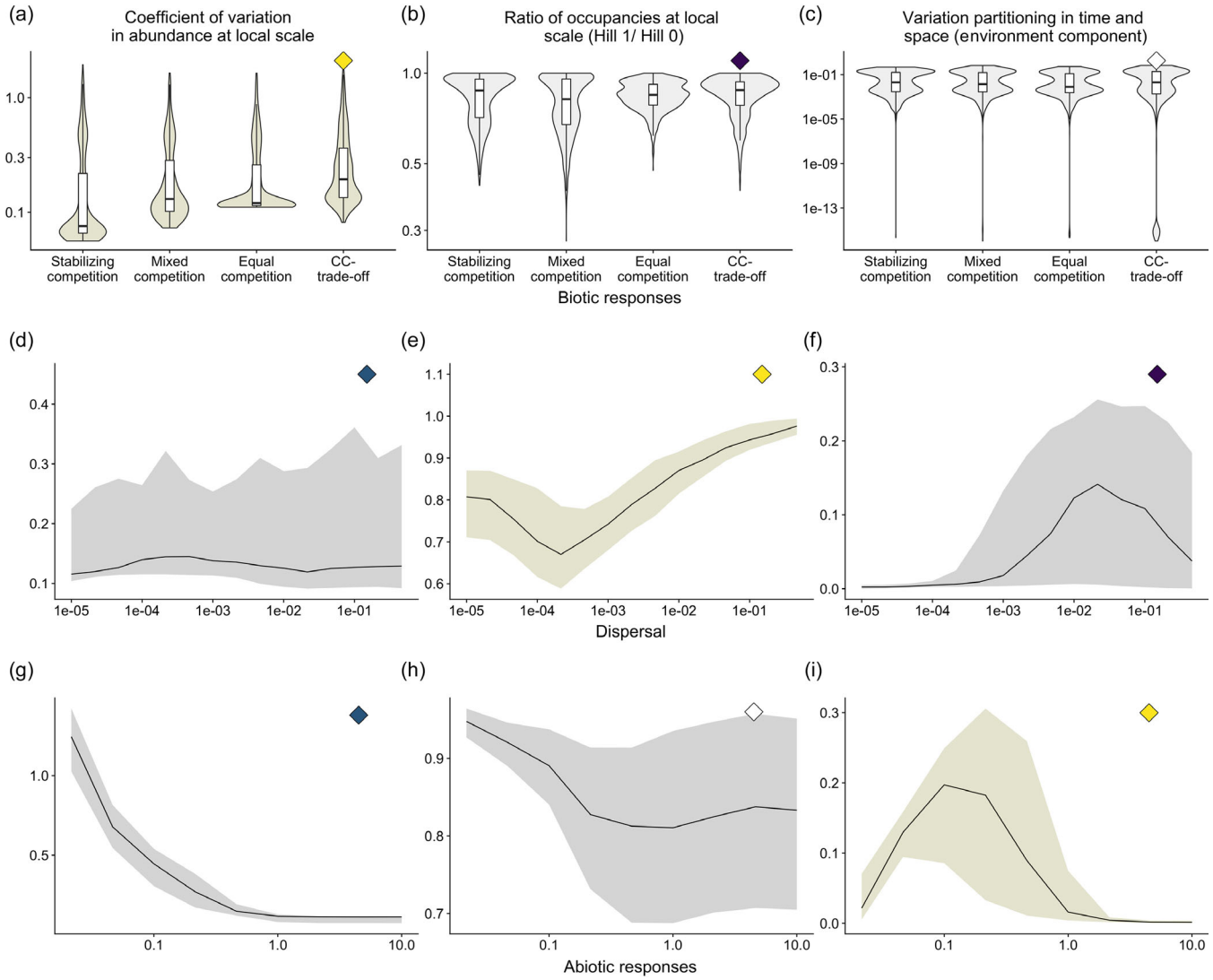


**FIGURE 3** The three most important summary statistics for predicting the underlying metacommunity processes (rows) as determined by the minimal random forest model (Model 7): (1) coefficient of local-scale variation in community abundance, (2) ratio of occurrence at local scale (Hill 1/Hill 0, this ratio is a measure of relative species rarity, which will be smaller when the community is uneven), and (3) environmental (E) component of variation partition through time and space. For the definition of all summary statistics see Table 1 and Appendix S1: Table S1. The color of the diamonds in the top right of each panel corresponds to the color in Figure 2 and shows the relative importance of each summary statistic. In panels (d–i) the ribbons represent the 1st and 3rd quartiles, while the middle lines represent the median values, mirroring the values used in the boxplots of panels (a–c). CC trade-off refers to the scenario of competition–colonization in the strength of biotic responses. The values on the *x*-axis for dispersal and abiotic responses were the values of the parameters' abiotic niche breadth ($\sigma_i$) and the probability of dispersal ($a_i$) used for the simulation. For each process we highlighted with yellow ribbons the most important summary statistic (equivalent to the yellow diamond) to add contrast

not take into account species interactions and temporal dynamics. In addition, Sokol et al. (2017) found that unimodal response curves hindered the ability of variation partitioning as a reliable metacommunity diagnostic tool, probably because of the difficulty in modeling bell-shaped or more complex responses with the linear models underlying RDA (Viana et al., 2021). Using metrics that describe community dynamics, rather than model fits, was ultimately more informative and useful for inferring metacommunity processes.

Modeling approaches that jointly accounted for abiotic, biotic, spatial, and temporal responses, such as HMSC, provided more power to distinguish metacommunity processes than did RDA-based approaches. Particularly relevant for the inference of metacommunity processes is that HMSC explicitly estimates positive or negative interspecific associations (Ovaskainen et al., 2017). Although the utility of using species associations for inferring the importance of species interactions is still debated (Blanchet et al., 2020), our results nevertheless suggest that these model-based statistics

increase our ability to distinguish among different types of biotic interactions.

Finally, we found that including a combination of descriptive statistics as well as model-based statistics provided the best results for determining metacommunity processes. Model-based statistics alone neither represented the most important summary statistics for all minimal models nor provided the highest explanatory power for the random forests. Unfortunately, many studies used model-based statistics alone (Cottenie, 2005; Ovaskainen et al., 2019; Soininen, 2014, 2016). These results suggest that using a combination of descriptive and model-based statistics will increase our ability to infer metacommunity processes.

## Sensitivity analysis

Limited sampling in space or time affects summary statistics in different ways. For example, some summary statistics, like the mean proportion of patches used, were more sensitive to the loss of patches, while the CV of abundance and richness differences were more sensitive to the loss of time points. As a result, we suggest that using a variety of summary statistics improves inference and can help reduce the sensitivity of an entire analysis to undersampling. A next step is to consider how temporal scale (duration of temporal sampling) can interact with temporal undersampling (frequency of sampling) to affect inferences about metacommunity structure (Castillo-Escrivà et al., 2020).

## General discussion and caveats

The relative importance of the summary statistics depends to some extent on the assumptions made in the simulation model. For example, we assumed that all the species had the same dispersal rate, but variation in dispersal rates would likely reduce the degree to which observed patterns can be used to assess dispersal processes (e.g., Monteiro et al., 2017). We made a similar assumption about niche breadth, where all the species had the same breadth but different optima and all patches were the same size. Differences among species in their niche breadth would again influence our ability to infer these processes from observed patterns. By modifying these and other assumptions, future studies can

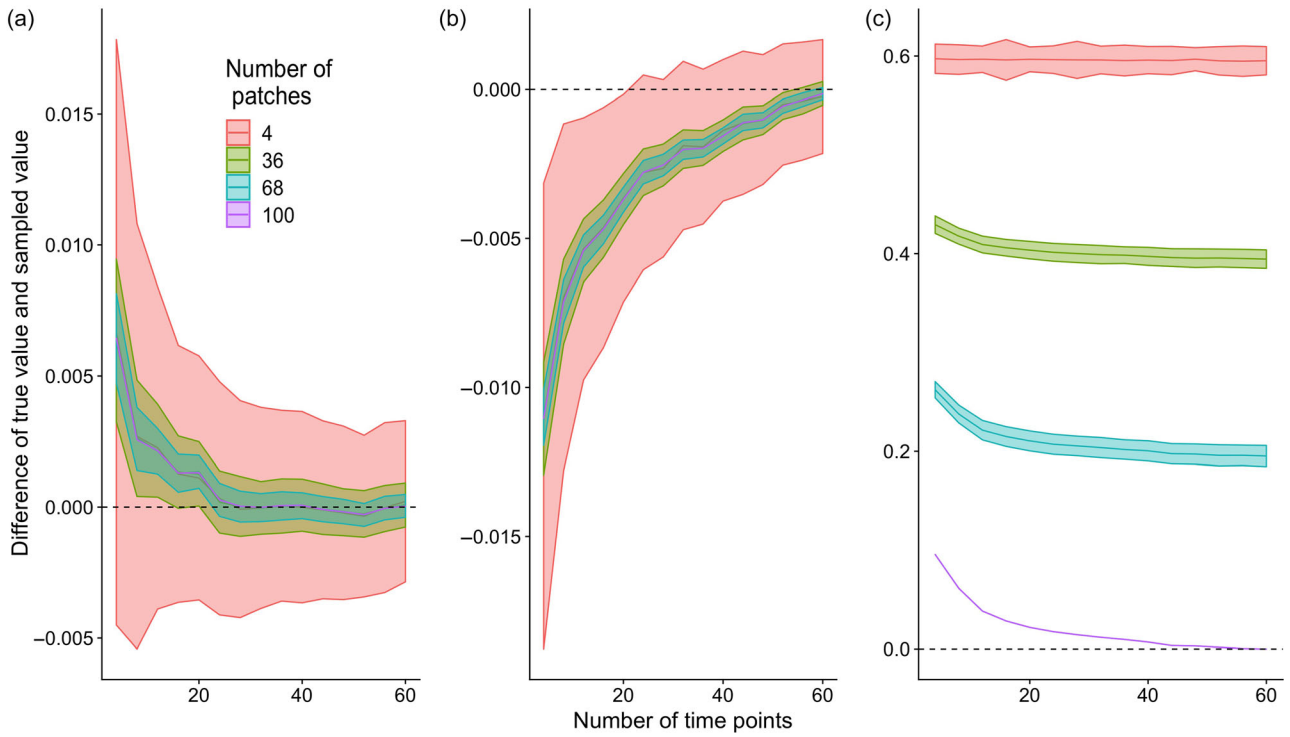


**FIGURE 4** Sampling fewer patches or time points increases the difference between an estimated summary statistic and its true value (i.e., when the metacommunity is fully sampled in space and time). The coefficient of variation in abundance at a local scale (a), the richness differences through time (b), and the mean proportion of patches occupied (c) vary in their sensitivity to loss of patches or time points. The middle lines represent the mean difference between the true value and the sampled value, while the ribbons represent one standard deviation

identify the robustness of the relationship between the processes and summary statistics.

We also assumed that all species were governed by the same metacommunity dynamics. However, when different species, or different locations, within a metacommunity are governed by different dynamics, significant departures from these expectations can emerge (Leibold et al., 2022). Empirical metacommunities may not support this assumption and may in fact have different subassemblages of taxa governed by distinct metacommunity dynamics (Thompson et al., 2017). Furthermore, more complexity is inherent in metacommunities with trophic interactions, which can further complicate analyses (Guzman et al., 2019). Nevertheless, an approach and workflow like ours that considers a multitude of metrics aimed at capturing a diverse set of dynamic patterns should also be useful when applied to these sorts of complexities in metacommunities.

Our analysis also uncovered another critical source of uncertainty in empirical metacommunity analysis that likely contributes to its typically low explanatory power—undersampling in space and time. Our sensitivity analysis shows that while some of the metrics are robust to small sample size in space or time, very few are robust when both space and time are undersampled. If empirical metacommunity studies sample only a few of the relevant patches or time points, understanding the processes that structure metacommunities might not be possible. Fortunately, there has been continued growth in the monitoring and compilation of longer-term time series of communities that vary in space (i.e., metacommunity time series) (see e.g., Dornelas et al., 2018), making us hopeful that more empirical metacommunity analyses with sufficient temporal sampling could be forthcoming.

## CONCLUSIONS AND FUTURE DIRECTIONS

Although we recognize that our investigation leaves many questions to be answered, we believe it provides an important step forward to help improve the empirical investigation of metacommunities. In particular, we present a proof of concept that illustrates that most previous empirical approaches for analyzing metacommunities are limited by focusing on only a few metrics or analytical approaches, as well as by limited sampling, especially through time. However, analyses can be greatly improved by using multiple metrics and analyses that provide complementary information on the relative importance of the core metacommunity processes (dispersal, biotic interactions, abiotic responses), so long as metacommunities are also adequately sampled in time and space.

We can also use our approach to develop hypotheses for experimental interventions, by testing the relationships between the important summary statistics and the core processes (Figure 3). For example, based on our simulations, we found that community evenness had a U-shaped relationship with dispersal. One could explore whether these relationships are altered when dispersal is altered, in the context of habitat fragmentation or the addition of dispersal corridors.

In addition to applying and improving our minimal model for empirical studies, a methodological next step could be to use the summary statistics identified here in an Approximate Bayesian Computation (ABC) framework (Pontarp et al., 2019). Such a framework would use simulated data and summary statistics to determine the posterior distribution of parameters of interest (e.g., dispersal rate) based on the distance between empirical data and summary statistics. This method can be used with an absolute model fit to assess how well the simulation model explains empirical data (Pennell et al., 2015).

In all, we highlight two main take-home messages from our study. First, although the majority of studies of metacommunities focus on static patterns of species abundances and distributions, we found that temporal dynamics were key for distinguishing the processes driving metacommunity dynamics. Indeed, temporal dynamics improved our ability to explain variation in density-dependent, density-independent, and dispersal processes by up to 60%. Second, although there can never be a one-to-one matching of pattern to process, we showed that the use of multiple summary statistics simultaneously was essential for disentangling processes. We found that both model-based and descriptive statistics were needed for the highest performance in inferring metacommunity processes and may be crucial in reducing inference uncertainty and accuracy when metacommunities are undersampled (i.e., virtually always). Although further testing is needed, the minimal models we propose here represent a step toward inferring the relative strengths of the three key metacommunity processes—dispersal, biotic interactions, and abiotic responses—in any empirically measured metacommunity.

## ACKNOWLEDGMENTS

This paper resulted from the sTURN working group funded by sDiv, the Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118). Additional funding came from Österreichische Forschungsgemeinschaft (ÖFG; International Communication, project 06/15539). Patrick L. Thompson was supported by Killam and NSERC postdoctoral fellowships. Luc De Meester acknowledges KU Leuven

Research Fund project C16/2017/002 and FWO project G0B9818. Zsófia Horváth acknowledges support by GINOP 2.3.2.-15-2016-00057, NKFIH OTKA FK-132095, NKFIH-872, and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Laura Melissa Guzman was supported by NSERC CGS-D, UBC Four Year Fellowships, and SFU. Jonathan M. Chase, Duarte S. Viana, and Alienor Jeliazkov were also supported by the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118). Duarte S. Viana was also supported by the Alexander von Humboldt Foundation.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

Code (Guzman, 2022) is available on Zenodo at: https://doi.org/10.5281/zenodo.5842513.

## ORCID

*Laura Melissa Guzman* https://orcid.org/0000-0003-2198-9738
*Patrick L. Thompson* https://orcid.org/0000-0002-5278-9045
*Duarte S. Viana* https://orcid.org/0000-0002-7864-0871
*Bram Vanschoenwinkel* https://orcid.org/0000-0002-8973-6297
*Zsófia Horváth* https://orcid.org/0000-0003-2857-1094
*Alienor Jeliazkov* https://orcid.org/0000-0001-5765-3721
*Stéphanie Gascón* https://orcid.org/0000-0003-1951-9083
*Maria Anton-Pardo* https://orcid.org/0000-0003-1722-9867
*Jonathan M. Chase* https://orcid.org/0000-0001-5580-4303

## REFERENCES

Ambroise, C., and G. J. McLachlan. 2002. "Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data." *Proceedings of the National Academy of Sciences of the United States of America* 99(10): 6562–6.

Beverton, R. J. H., and Holt, S. 1957. *On the Dynamics of Exploited Fish Populations*, B.V. Fish and Fisheries Series 11. Berlin/Heidelberg, Germany: Springer Science and Business Media.

Blanchet, F. G., K. C. Blanchet, and D. Gravel. 2020. "Co-occurrence Is Not Evidence of Ecological Interactions." *Ecology Letters* 23(7): 1050–63.

Blanchet, F. G., P. Legendre, and D. Borcard. 2008. "Modelling Directional Spatial Processes in Ecological Data." *Ecological Modelling* 215(4): 325–36.

Borcard, D., P. Legendre, and P. Drapeau. 1992. "Partialling Out the Spatial Component of Ecological Variation." *Ecology* 73(3): 1045–55.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Brown, C., J. B. Illian, and D. F. R. P. Burslem. 2016. "Success of Spatial Statistics in Determining Underlying Process in Simulated Plant Communities." *The Journal of Ecology* 104(1): 160–72.

Castillo-Escrivà, A., F. Mesquita-Joanes, and J. Rueda. 2020. "Effects of the Temporal Scale of Observation on the Analysis of Aquatic Invertebrate Metacommunities." *Frontiers in Ecology and Evolution*. 8: 561838. https://doi.org/10.3389/fevo.2020.561838.

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. "Rarefaction and Extrapolation with Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies." *Ecological Monographs* 84(1): 45–67.

Chase, J. M., and M. A. Leibold. 2003. *Ecological Niches: Linking Classical and Contemporary Approaches*. Chicago and London: University of Chicago Press.

Chesson, P. 2000. "Mechanisms of Maintenance of Species Diversity." *Annual Review of Ecology and Systematics* 31(1): 343–66.

Clappe, S., S. Dray, and P. R. Peres-Neto. 2018. "Beyond Neutrality: Disentangling the Effects of Species Sorting and Spurious Correlations in Community Analysis." *Ecology* 99(8): 1737–47.

Cottenie, K. 2005. "Integrating Environmental and Spatial Processes in Ecological Community Dynamics." *Ecology Letters* 8(11): 1175–82.

Dornelas, M., L. H. Antão, F. Moyes, A. E. Bates, and A. E. Magurran. 2018. "BioTIME: A Database of Biodiversity Time Series for the Anthropocene." *Global Ecology and Biogeography: A Journal of Macroecology* 27(7): 760–86.

Dray, S., D. Bauman, G. Blanchet, D. Borcard, S. Clappe, G. Guenard, T. Jombart, et al. 2019. "Adespatial: Multivariate Multiscale Spatial Analysis (version 0.3-7)." R. https://CRAN.R-project.org/package=adespatial.

Ehrlén, J., and O. Eriksson. 2000. "Dispersal Limitation and Patchy Occupancy in Forest Herbs." *Ecology* 81(6): 1667–74.

Gilbert, B., and J. R. Bennett. 2010. "Partitioning Variation in Ecological Communities: Do the Numbers Add Up?" *Journal of Applied Ecology* 47(5): 1071–82.

Guzman, L. M. 2022. "lmguzman/disentangling_metacommunities: First Release (Version V1)." Zenodo. https://doi.org/10.5281/zenodo.5842513.

Guzman, L. M., R. M. Germain, C. Forbes, S. Straus, M. I. O'Connor, D. Gravel, D. S. Srivastava, and P. L. Thompson. 2019. "Towards a Multi-Trophic Extension of Metacommunity Ecology." *Ecology Letters* 22(1): 19–33.

Hill, M. O. 1973. "Diversity and Evenness: A Unifying Notation and its Consequences." *Ecology* 54(2): 427–32.

Hubbell, S. P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology, Vol 32. Princeton, NJ: Princeton University Press.

Jabot, F., F. Laroche, F. Massol, F. Arthaud, J. Crabot, M. Dubart, S. Blanchet, F. Munoz, P. David, and T. Datry. 2020. "Assessing Metacommunity Processes through Signatures in Spatiotemporal Turnover of Community Composition." *Ecology Letters* 23(9): 1330–9.

Legendre, P. 2014. "Interpreting the Replacement and Richness Difference Components of Beta Diversity." *Global Ecology and Biogeography* 23(11): 1324–34.

Leibold, M. A., and J. M. Chase. 2017. *Metacommunity Ecology*. Princeton and London: Princeton University Press.

Leibold, M. A., F. Javiera Rudolph, F. Guillaume Blanchet, L. De Meester, D. Gravel, F. Hartig, P. Peres-Neto, L. Shoemaker, and J. M. Chase. 2022. "The Internal Structure of Metacommunities" *Oikos* 2022(1), https://doi.org/10.1111/oik.08618

Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2(3): 18–22.

Loreau, M., N. Mouquet, and A. Gonzalez. 2003. "Biodiversity as Spatial Insurance in Heterogeneous Landscapes." *Proceedings of the National Academy of Sciences of the United States of America* 100(22): 12765–70.

May, F., A. Huth, and T. Wiegand. 2015. "Moving beyond Abundance Distributions: Neutral Theory and Spatial Patterns in a Tropical Forest." *Proceedings. Biological Sciences / The Royal Society* 282(1802): 20141657. https://doi.org/10.1098/rspb.2014.1657.

Monteiro, V. F., P. C. Paiva, and P. R. Peres-Neto. 2017. "A Quantitative Framework to Estimate the Relative Importance of Environment, Spatial Variation and Patch Connectivity in Driving Community Composition." *The Journal of Animal Ecology* 86(2): 316–26.

Münkemüller, T., F. de Bello, C. N. Meynard, D. Gravel, S. Lavergne, D. Mouillot, N. Mouquet, and W. Thuiller. 2012. "From Diversity Indices to Community Assembly Processes: A Test with Simulated Data." *Ecography* 35(5): 468–80.

O'Dwyer, J. P., and S. J. Cornell. 2018. "Cross-Scale Neutral Ecology and the Maintenance of Biodiversity." *Scientific Reports* 8(1): 10200.

Ovaskainen, O., J. Rybicki, and N. Abrego. 2019. "What Can Observational Data Reveal about Metacommunity Processes?" *Ecography* 42(11): 1877–86.

Ovaskainen, O., G. Tikhonov, F. Anna Norberg, G. Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego. 2017. "How to Make More out of Community Data? A Conceptual Framework and its Implementation as Models and Software." *Ecology Letters* 20(5): 561–76.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. "Model Adequacy and the Macroevolution of Angiosperm Functional Traits." *The American Naturalist* 186(2): E33–50.

Peres-Neto, P. R., and P. Legendre. 2010. "Estimating and Controlling for Spatial Structure in the Study of Ecological Communities." *Global Ecology* 19(2): 174–84.

Podani, J., and D. Schmera. 2011. "A New Conceptual and Methodological Framework for Exploring and Explaining Pattern in Presence - Absence Data." *Oikos* 120(11): 1625–38.

Pontarp, M., Å. Brännström, and O. L. Petchey. 2019. "Inferring Community Assembly Processes from Macroscopic Patterns Using Dynamic Eco-Evolutionary Models and Approximate Bayesian Computation (ABC)." *Methods in Ecology and Evolution* 10(4): 450–60.

Schmera, D., J. Podani, and P. Legendre. 2020. "What Do Beta Diversity Components Reveal from Presence-Absence Community Data? Let us Connect every Indicator to an Indicandum!" *Ecological Indicators* 117: 106540.

Shoemaker, L. G., L. L. Sullivan, I. Donohue, J. S. Cabral, R. J. Williams, M. M. Mayfield, J. M. Chase, et al. 2020. "Integrating the Underlying Structure of Stochasticity into Community Ecology." *Ecology* 101(2): e02922.

Soininen, J. 2014. "A Quantitative Analysis of Species Sorting across Organisms and Ecosystems." *Ecology* 95(12): 3284–92.

Soininen, J. 2016. "Spatial Structure in Ecological Communities - A Quantitative Analysis." *Oikos* 125(2): 160–6.

Sokol, E. R., B. L. Brown, and J. E. Barrett. 2017. "A Simulation-Based Approach to Understand how Metacommunity Characteristics Influence Emergent Biodiversity Patterns." *Oikos* 126 (5): 723–37. https://doi.org/10.1111/oik.03690.

Svetnik, V., A. Liaw, C. Tong, and T. Wang. 2004. "Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules." In *Multiple Classifier Systems*. Lecture Notes in Computer Science, Vol 3077, edited by F. Roli, J. Kittler, and T. Windeatt, 334–43. Berlin and Heidelberg: Springer.

Tatsumi, S., J. Strengbom, M. Čugunovs, and J. Kouki. 2020. "Partitioning the Colonization and Extinction Components of Beta Diversity across Disturbance Gradients." *Ecology* 101(12): e03183.

Thompson, P. L., L. M. Guzman, L. De Meester, Z. Horváth, R. Ptacnik, B. Vanschoenwinkel, D. S. Viana, and J. M. Chase. 2020. "A Process-Based Metacommunity Framework Linking Local and Regional Scale Community Ecology." *Ecology Letters* 23(9): 1314–29.

Thompson, P. L., B. Rayfield, and A. Gonzalez. 2017. "Loss of Habitat and Connectivity Erodes Species Diversity, Ecosystem Functioning, and Stability in Metacommunity Networks." *Ecography* 40(1): 98–108.

Tikhonov, G., O. Opedal, N. Abrego, A. Lehikoinen, and O. Ovaskainen. 2020. "Joint Species Distribution Modelling with HMSC-R." *Methods in Ecology and Evolution* 11(3): 442–7.

Tilman, D. 1982. "Resource Competition and Community Structure." *Monographs in Population Biology* 17: 1–296.

Tilman, D. 1995. "Biodiversity: Population Versus Ecosystem Stability." *Ecology* 77(2): 350–63.

Tuomisto, H. 2010. "A Diversity of Beta Diversities: Straightening up a Concept Gone Awry. Part 1. Defining Beta Diversity as a Function of Alpha and Gamma Diversity." *Ecography* 33(1): 2–22.

Viana, D. S., P. Keil, and A. Jeliazkov. In Press. "Disentangling Spatial and Environmental Effects: Flexible Methods for Community Ecology and Macroecology." *Ecosphere*. https://doi.org/10.1002/ecs2.4028.

Zurell, D., U. Berger, J. S. Cabral, F. Jeltsch, C. N. Meynard, T. Münkemüller, N. Nehrbass, et al. 2010. "The Virtual Ecologist Approach: Simulating Data and Observers." *Oikos* 119(4): 622–35.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Guzman, Laura Melissa, Patrick L. Thompson, Duarte S. Viana, Bram Vanschoenwinkel, Zsófia Horváth, Robert Ptacnik, Alienor Jeliazkov, et al. 2022. "Accounting for Temporal Change in Multiple Biodiversity Patterns Improves the Inference of Metacommunity Processes." *Ecology* 103(6): e3683. https://doi.org/10.1002/ecy.3683