ECOGRAPHY

Research

What can observational data reveal about metacommunity processes?

Otso Ovaskainen, Joel Rybicki and Nerea Abrego

O. Ovaskainen (https://orcid.org/0000-0001-9750-4421) ✉ (otso.ovaskainen@helsinki.fi) and J. Rybicki, Organismal and Evolutionary Biology Research Programme, Univ. of Helsinki, Finland. OO also at: Centre for Biodiversity Dynamics, Dept of Biology, Norwegian Univ. of Science and Technology, Trondheim, Norway. JR also at: Inst. of Science and Technology Austria, Klosterneuburg, Austria. – N. Abrego (https://orcid.org/0000-0001-6347-6127), Dept of Agricultural Sciences, Univ. of Helsinki, Finland.

A key challenge for community ecology is to understand to what extent observational data can be used to infer the underlying community assembly processes. As different processes can lead to similar or even identical patterns, statistical analyses of non-manipulative observational data never yield undisputable causal inference on the underlying processes. Still, most empirical studies in community ecology are based on observational data, and hence understanding under which circumstances such data can shed light on assembly processes is a central concern for community ecologists. We simulated a spatial agent-based model that generates variation in metacommunity dynamics across multiple axes, including the four classic metacommunity paradigms as special cases. We further simulated a virtual ecologist who analysed snapshot data sampled from the simulations using eighteen output metrics derived from beta-diversity and habitat variation indices, variation partitioning and joint species distribution modelling. Our results indicated two main axes of variation in the output metrics. The first axis of variation described whether the landscape has patchy or continuous variation, and thus was essentially independent of the properties of the species community. The second axis of variation related to the level of predictability of the metacommunity. The most predictable communities were niche-based metacommunities inhabiting static landscapes with marked environmental heterogeneity, such as metacommunities following the species sorting paradigm or the mass effects paradigm. The most unpredictable communities were neutral-based metacommunities inhabiting dynamics landscapes with little spatial heterogeneity, such as metacommunities following the neutral or patch sorting paradigms. The output metrics from joint species distribution modelling yielded generally the highest resolution to disentangle among the simulated scenarios. Yet, the different types of statistical approaches utilized in this study carried complementary information, and thus our results suggest that the most comprehensive evaluation of metacommunity structure can be obtained by combining them.

Keywords: assembly process, beta diversity, community structure, joint species distribution model, metacommunity paradigm, variation partitioning

1

Introduction

Community ecologists apply theoretical, experimental and observational approaches to determine which processes structure ecological communities (Morin 2011). Experimental approaches provide the most direct way of testing the effects of specific processes in assembling particular communities, yet they hardly resemble the complexity of natural systems (Carpenter 1996, Schindler 1998). Likewise, while mathematical models can be used to clarify the links from underlying mechanisms to the dynamics of ecological communities, they typically rely on highly simplified assumptions and thus can be difficult to relate to empirical data. Data acquired by non-manipulative observational approaches are shaped by the full complexity of assembly processes, but as these processes can seldom be observed directly, the challenge with this approach is to infer the assembly processes from the observed patterns. As different underlying processes can lead to similar or even identical patterns, statistical analyses of non-manipulative observational data never yield undisputable causal inference on the underlying processes (Cale et al. 1989). Still, most empirical studies in community ecology are based on observational data, and hence understanding if and how such data can shed light on assembly processes is a central concern for community ecologists (Logue et al. 2011).

Metacommunity theory explains how networks of local species communities (metacommunities) result from the interplay of various stochastic (e.g. related to dispersal and extinction) and deterministic processes (e.g. niche differences and variation in competitive abilities), which are collectively called assembly processes (Lortie et al. 2004, Leibold and McPeek 2006, Vellend 2010, Götzenberger et al. 2012). Classic metacommunity theory synthesizes four perspectives, each arising from a different – but not mutually exclusive – conceptual framework: neutral, patch dynamics, species sorting and mass effects paradigms (Leibold et al. 2004, Holyoak et al. 2005). In the neutral paradigm, all individuals are considered to be equal in competitive capabilities as well as niche preferences irrespective to which species they belong to, so that the differences in species composition emerge solely from stochastic ecological drift. The patch dynamics perspective assumes that species track ephemeral habitat patches through colonization–extinction dynamics, in which case species coexistence is facilitated by e.g. colonization–competition trade-off. Species sorting focuses on the differences in the species' niche preferences along environmental gradients. The mass effects perspective differs from species sorting by assuming a much greater dispersal rate between the metacommunities which may influence variation in community composition e.g. through source–sink dynamics.

While metacommunity theory describes how different ecological processes influence community structure, a major challenge for community ecologists is to relate their data to theory, that is, to disentangle the relative importance of the various assembly processes in structuring the focal community (Cottenie et al. 2003, Tuomisto et al. 2003, Cottenie

2005, Vanschoenwinkel et al. 2007, Legendre et al. 2009, Kuglerová et al. 2015). For example, if observing that the community varies only little over space, this may be either because of high dispersal rates or because of low level of environmental variation or lack of niche differentiation (Chave and Leigh 2002, Condit et al. 2002, Chase 2003). Conversely, if the community varies much over space, this may indicate either the importance of stochastic process, or a high level of environmental variation combined with niche differences (Chase and Myers 2011, Weinstein et al. 2014). Conclusive disentangling among the alternative explanations is not possible solely based on observational data, but requires other kinds of information or biological knowledge on the system. For instance, if the relevant aspects of environmental variation are unknown or unmeasured, the researcher may falsely conclude that the community is structured by stochastic processes rather than niche differentiation (Chang et al. 2013).

A key question for empirical researchers is whether and how the broad array of available analysis methods allows one to relate observational data to metacommunity theory. One common approach for assessing how communities are structured along environmental gradients, or more generally across space, is to examine patterns in species turnover, i.e. beta-diversity, which can be measured by various indices (Baselga 2010, Chase and Bengtsson 2010, Chase and Myers 2011). Another related approach is that of variation partitioning, where variation in community structure is decomposed into environmental and spatial components (McArdle and Anderson 2001, Legendre et al. 2005, Smith and Lundholm 2010). These approaches can be combined with data on species traits or phylogenetic relationships, e.g. to assess patterns of trait beta-diversity or phylogenetic beta-diversity (Weinstein et al. 2014). Alternatively, one may start from the species-specific perspective, assessing how the occurrences of individual species vary over space or environmental gradients, as done with species distribution modeling (D'Amen et al. 2017). Recently emerged joint species distribution models (JSDM) consider multiple interlinked species simultaneously, making it possible to derive both species- and community-level inferences (Warton et al. 2015, Clark et al. 2017, Ovaskainen et al. 2017).

Several studies have examined the ability of diversity indices and variation partitioning approaches to disentangle among neutral- and niche-based metacommunity processes using data simulated either by mechanistic metacommunity models based on birth–death processes (Smith and Lundholm 2010, Münkemüller et al. 2012, Tucker et al. 2016, Clappe et al. 2018) or phenomenological metacommunity models that generate directly patterns of species occurrence (Miller et al. 2017). These studies have shown that diversity indices and variation partitioning can under certain conditions separate communities simulated with neutral dynamics, environmental filtering and competitive interactions, especially when utilized along with phylogenetic and functional information. Further, Zurell et al. (2018) showed

that residual species-to-species correlations estimated by JSDMs coincide with co-occurrence indices in homogeneous landscapes, but that inferring the causal species interactions behind co-occurrence is difficult, as e.g. predator–prey interactions can lead to either positive or negative pattern of co-occurrence.

In this paper, we use simulated metacommunities to ask to what extent assembly processes can be inferred from non-manipulative observational data. Compared to the previous studies discussed above, our aims differ in three important ways. First, we provide a comparison between the widely used index-based and variation partitioning approaches to the more recently proposed JSDM approaches, which comparison has thus far been lacking. Second, we fill in the knowledge gap identified by Zurell et al. (2018), which is to test the ability of JSDMs to infer metacommunity processes using mechanistic and spatially explicit population models. Third, while the earlier studies have been chiefly concerned in disentangling among the classical metacommunity paradigms (Smith and Lundholm 2010, Münkemüller et al. 2012, Tucker et al. 2016, Miller et al. 2017, Clappe et al. 2018), we focus on simultaneous variation in multiple assembly processes underlying these paradigms rather on the paradigms themselves. To do so, we devise a general agent-based metacommunity model to simulate many kinds of metacommunities, including as special cases scenarios that mimic the four central metacommunity paradigms, but more importantly also the parameter space between these distinct paradigms. Our approach is conceptually depicted in Fig. 1. In addition to the metacommunity dynamics, we simulate a 'virtual ecologist' (Zurell et al. 2010) who gathers data from the simulated metacommunities by sampling both environmental covariates and species occurrences over a set of spatial locations, and then statistically analyses the data. Our key aim is to identify how output metrics that can be derived from snapshot data co-vary with the community assembly processes, and conclude how effective the output metrics are in disentangling the assembly processes that generated the observed data.

Material and methods

The simulation model

We simulated metacommunity dynamics using an agent-based resource–consumer model, the details of which are given in the Supplementary material Appendix 1. To generate scenarios that span the range of the classic metacommunity paradigms, we consider a number of parameterizations of the individual-based model, presented with the help of the nine choices (C1–C9) that are described conceptually below and in Fig. 2, and in full detail in Supplementary material Appendix 1. To generate variation in resource types and their availability, we assume that there are four types of habitat patches, each of which produce six types of resource particles, so that the total type of resource types is 24. The landscape either has or has not large-scale gradients in resource availability (Choice C1). The landscape is further either patchy so that it consists of distinct patches, or that the patches overlap to the extent that spatial variation in the habitat availability is continuous (Choice C2). We further assume that either all patches are of equal quality, or that there is quality variation, some patches being sources and others sinks (Choice C3).
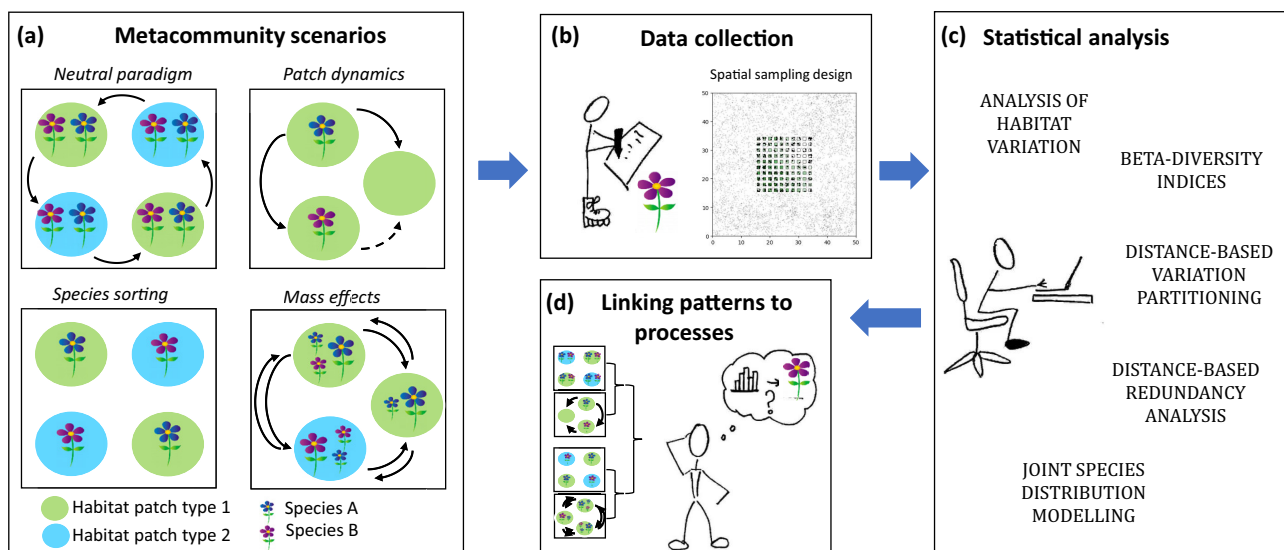

Figure 1. The general workflow of the study. Panel (a) illustrates the metacommunity scenarios corresponding to the four central metacommunity paradigms (adapted from Leibold et al. 2004). Panel (b) illustrates the 'virtual ecologist' collecting data on environmental covariates and community composition along a number of sampling locations. Panel (c) illustrates the ecologist applying statistical analyses to the data, resulting into the output metrics described in Table 1. Panel (d) symbolizes the ecologist trying to connect the statistical results to the underlying processes structuring the observed communities.
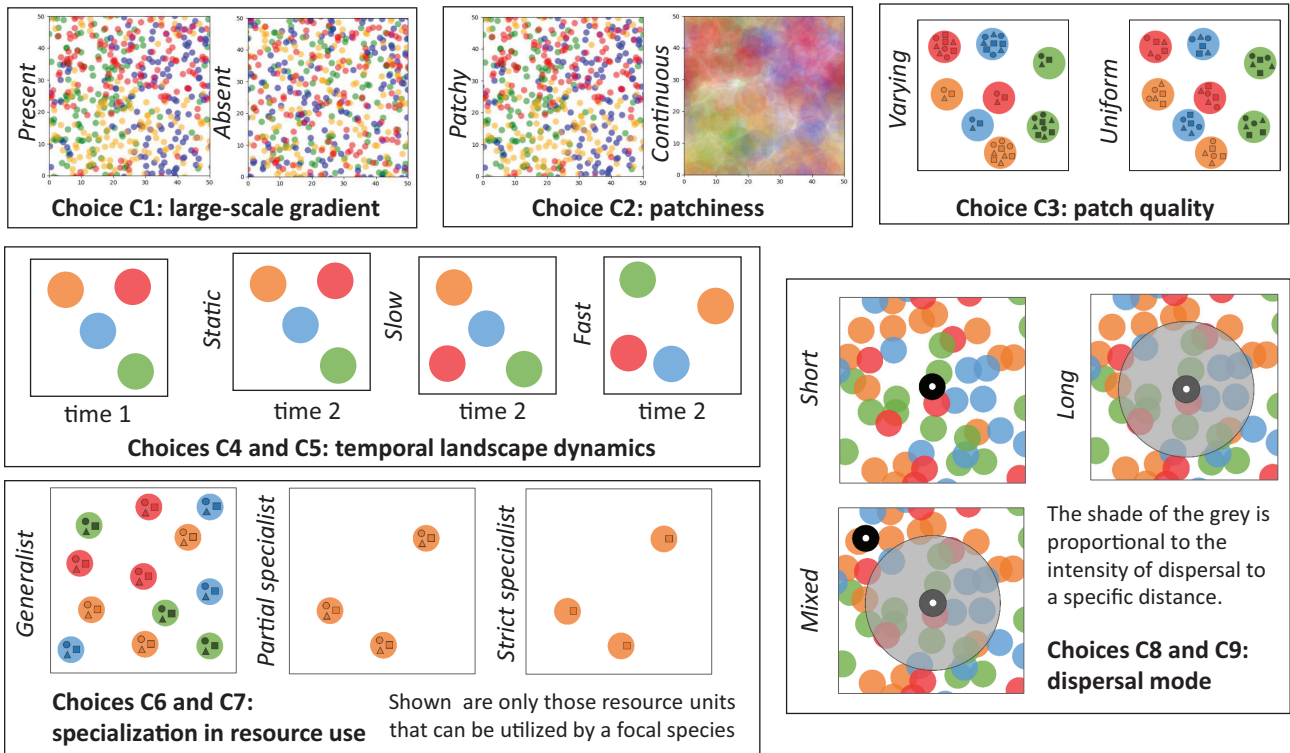
Figure 2. Illustration of the Choices C1–C9 behind the simulated scenarios. Habitat patch types are shown by different colours, and resource types are shown by symbols (square, triangle or circle).

Concerning the temporal structure, we assume that the landscape is either static or dynamic (Choice C4), and that in the dynamic case the patches are either short- or long-lived (Choice C5). In a snapshot, the dynamic and static landscapes are indistinguishable in their habitat structure. Concerning variation in resource use, we assume that either all species are generalists utilizing equally well all resource types (in which case they all compete for the same resources), or that the species are specialist for particular resource types (Choice C6). In the latter case, we further assume that the level of specialization is either strict (one-to-one correspondence between species and resource types, in which case the species do not show resource competition) or partial (species are specialized to patches rather than resource types, in which case those species that are specialized to the same patch compete with each other) (Choice C7). Concerning variation in dispersal capacity, we assume that the species are either restricted to short-distance dispersal or also capable of long-distance dispersal. The first choice (Choice C8) is whether all species follow the same dispersal strategy, or if half of the species follow short-distance dispersal and the remaining half long-distance dispersal. The latter case generates a colonization–competition trade-off: the short-distance dispersers leave more propagules locally, and thus, they are superior in within-patch competition, whereas the long-distance dispersers are superior in colonizing empty patches located far away from the source patches. If all species follow the same dispersal strategy, they follow either short or long-dispersal

(Choice C9). Combinations of the nine Choices C1–C9 yield a total number 216 different metacommunity scenarios (Supplementary material Appendix 1).

Some particular parameterizations of the general agent-based model conceptually match the four classic metacommunity paradigms. We first note that the metacommunity concept refers most characteristically to situations where it is possible to define well-lineated local communities, which is here achieved by patchy variation in habitat quality (Choice C2). The neutral perspective is represented by generalist resource use by the species, and the absence of competition–colonization trade-off: with these assumptions, species' identity does not influence birth and death rates, and only total abundance of species is regulated, as is characteristic of the neutral paradigm. The species sorting perspective corresponds most characteristically to a static landscape that involves large-scale gradient in habitat availability, uniform patch quality, strict specialization in resource use, and species that follow the short dispersal strategy. In this case, the species have time to colonize the parts of the landscape that are most suitable for them (due to the static landscape), different parts of the landscape support different subsets of species (due to the presence of large-scale gradient), all patches are of high enough quality for supporting local populations (due to uniform patch quality), and the dynamics within local communities are not much influenced by other local communities (due to short dispersal). The mass effect perspective is similar to the species sorting paradigm,

but without large-scale variation in resource availability and with varying patch quality, and with all species following the long dispersal strategy. In this case, the dynamics within local communities are substantially influenced by other local communities (due to long dispersal range and lack of large-scale gradient separating different types of local communities), and some patches are not of sufficient quality for supporting local populations (due to varying patch quality), which creates source–sink dynamics. The patch dynamics perspective is based on a dynamic landscape (which creates patch turnover), absence of large-scale gradients in resource availability and uniform patch qualities (which creates identical patches and homogeneous connectivities among patches), generalist resource use (which makes the patches identical from the species point of view), and that some species follow the short and others the long dispersal scenario (which creates a competition–colonization trade-off).

Data sampling

We assume that a virtual ecologist samples three kinds of data from the simulations: data on species occurrences on a set of sites ($\mathbf{Y}$), data on environmental conditions on those sites ($\mathbf{X}$) and the spatial coordinates of the sites ($\mathbf{xy}$). While trait data and phylogenetic data can be highly useful for inferring metacommunity processes, for simplicity we do not assume the availability of such data in the current study, but instead refer to earlier work (Münkemüller et al. 2012, Weinstein et al. 2014, Miller et al. 2017) that has examined the ability of trait-based and phylogenetic-based divergence indices to separate metacommunity paradigms. We assumed that a virtual ecologist acquired data from the final state of the simulation, placing 100 study plots into a regular 10×10 grid (Fig. 1b). The researcher scored the presence–absence of each species in each study plot, resulting in the matrix $\mathbf{Y}$, where the columns correspond to species and rows to sampling locations. The virtual ecologist collected data also on habitat quality according to two sampling scenarios: either the virtual ecologist acquired covariate data on all four habitat types included in our simulations (Supplementary material Appendix 1), or only on three out of the four habitat types, the fourth one thus being missing covariate (Choice C10). These data form the matrix $\mathbf{X}$, where the columns correspond to habitat types and the rows to sampling locations, and the values give the habitat qualities. As the number of metacommunity scenarios is 216 (Choices C1–C9) and the number of sampling scenarios is two (Choice 10), the total number of the types of data sets that we generated was 432.

Statistical methods

The virtual ecologist computed in total eighteen output metrics characterizing the data (Table 1). These output metrics include two indices of habitat variation (based on $\mathbf{X}$ and $\mathbf{xy}$), three beta-diversity indices (Baselga 2010) (based on $\mathbf{Y}$ only), three metrics derived from distance-based variation partitioning (Legendre et al. 2005, Smith and Lundholm 2010) (based on $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{xy}$), three metrics derived from distance-based redundancy analysis (McArdle and Anderson 2001) (based on $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{xy}$), and seven output metrics derived from the joint species distribution model HMSC (Ovaskainen et al. 2017) (based on $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{xy}$). These output metrics are listed in Table 1, and details on how they were calculated as well as hypotheses about how they might relate to assembly processes are given in the Supplementary material Appendix 1.

The analyses described above produced eighteen output metrics for each of the 432 scenarios. To synthesize these results, we performed three kinds of analyses. First, to

Table 1. Output metrics included in the analyses. The eighteen output metrics are derived from five different analysis types: analysis of habitat variation (HAB), beta-diversity indices (BETA), distance-based variation partitioning (db-VP), distance-based redundancy analysis (db-RDA) and joint species distribution modelling (JSDM).

No.	Analysis type	Metric	Description
1	HAB	V_{HAB}	Variance in habitat quality
2	HAB	D_{HAB}	Distance decay in habitat similarity
3	BETA	β_{SOR}	Sørensen-based multiple-site dissimilarity
4	BETA	β_{SIM}	Simpson-based multiple-site dissimilarity
5	BETA	β_{NES}	Nestedness-resultant multiple-site dissimilarity
6	db-VP	V_{ABC}	Total explained variance in dp-VP
7	db-VP	V_{AB}/V_{ABC}	Environmental proportion in dp-VP
8	db-VP	V_{BC}/V_{ABC}	Spatial proportion in dp-VP
9	db-RDA	R^2.adj	Total explained variance in dp-RDA
10	db-RDA	$X1|X2/R^2$.adj	Environmental proportion in dp-RDA
11	db-RDA	$X2|X1/R^2$.adj	Spatial proportion in dp-RDA
12	JSDM	AUC	Predictive power
13	JSDM	V_{RAND}	Variance attributed to random effects
14	JSDM	RUS	Evidence for resource use specialization
15	JSDM	POS	Proportion of species pairs with positive association
16	JSDM	NEG	Proportion of species pairs with negative association
17	JSDM	$E[\alpha]$	Posterior mean of spatial scale of residual variation
18	JSDM	$Pr[\alpha > 0]$	Posterior support for spatially structured residual variation

summarize how the output metrics vary generally among the scenarios, we applied redundancy analysis (RDA, Legendre and Anderson 1999), where we used the output metrics as the response variables and the choices C1–C10 as explanatory variables. Second, we explored to what extent the distributions of the individual output metrics were able to separate among each of the Choices C1–C10. To do so, we examined the distributions of the output metrics for the sets of scenarios that differed in each of the focal choices. We considered that an output metric provided substantial evidence for identifying a particular choice if its value showed a consistent difference in at least 80% of the comparisons that differed in the focal choice (say, in case of Choice C2, the output metric was greater in patchy than in continuous landscapes in at least 80% of comparisons). Third, to ask to which extent they can identify the scenario that generated the data, we applied multivariate probit-regression in HMSC, where the response variable was the Choices C1–C10 (a vector of zeros or ones or NA in case of the choice having no effect, e.g. slow or fast dynamics for static landscapes) and the explanatory variables were the output metrics. We recorded the predictive power for disentangling among simulated scenarios with AUC value based on 10-fold cross-validation. To examine which groups

of output metrics carried the most relevant information, we performed this analysis in six different ways: by including the output metrics either only from one of the five analysis types (Table 1), or by including all eighteen output metrics.

Results

Figure 3 shows the results of the RDA analysis synthesizing how the output metrics varied among the scenarios. We interpret the first axis of variation (RDA1) to describe whether the landscape was patchy or homogeneous. This result shows that the output metrics are generally sensitive to landscape structure, which in our study design was independent of the properties of the species that inhabit the landscape. We interpret the second axis of variation (RDA2) to relate to the predictability of the metacommunity. Overall, the most predictable metacommunities generally clustered to the upper and the most unpredictable metacommunities to the lower part of the RDA space, as indicated by the output metrics O2 (AUC), O6 (V_{ABC}) and O9 (R^2.adj) being located in the upper part of the space. Species belonging to predictable metacommunities were characterized by niche
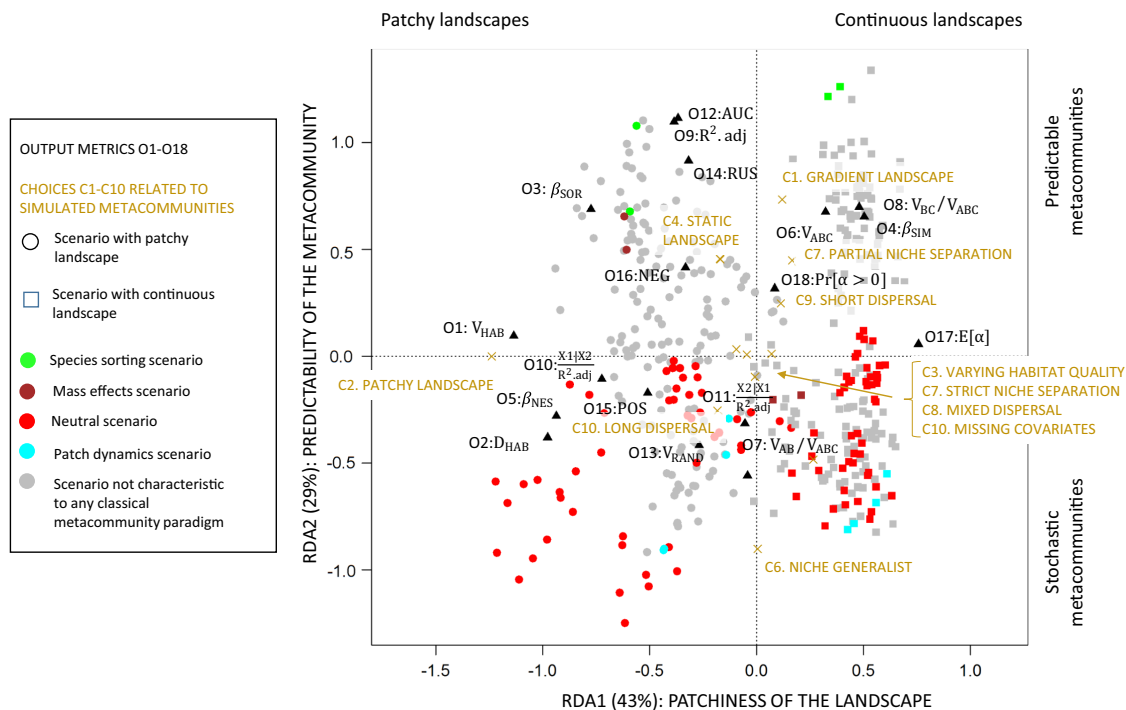

Figure 3. An RDA analysis synthesizing variation in the output metrics O1–O18 of the statistical analyses (Table 1) over the 432 simulated scenarios that vary among the choices C1–C10 related to their underlying processes. The dots (respectively, squares) depict scenarios simulated on patchy (respectively, continuous) landscapes. The colours indicate whether the simulated scenario follows one of the classical metacommunity paradigms (green, brown, read and cyan) or falls in between them (grey). The output metrics O1–O18 (Table 1) are shown by the black triangles, and the Choices C1–C10 related to the simulated scenarios are shown by the brown crosses. The locations of both the of output metrics O1–O18 as well as the Choices C1–C10 in the RDA space have been multiplied by a constant to spread them more evenly over the ordination space and thus to avoid extensive overlap of symbols in the figure.

separation and short-distance dispersal, and they inhabit static landscapes with large-scale gradient in habitat availability. Examples of predictable metacommunities are given by those following the species sorting and the mass effects scenarios, whereas examples of unpredictable metacommunities are given by those following the neutral and patch dynamics scenarios (Fig. 3). The neutral scenarios spread over a large part of the RDA space, and the scenarios that followed the classical paradigms (red, green, brown and cyan dots in Fig. 3) are mixed with those that do not follow any classical paradigm (grey dots in Fig. 3). That is, the classical metacommunity paradigms reflected only part of the community variation.

Figure 3 suggests that patchy landscapes are characterized by high values of the output metrics O1 (V_{HAB}), O2 (D_{HAB}), O3 (β_{SOR}), O5 (β_{NES}), O10 (X1|X2/R^2.adj) and O15 (POS) and low value of O17 (E[α]). These results obtained strong statistical support by the evaluation of the distributions of the individual output metrics. Figure 3 further suggests that predictable metacommunities are characterized by high values of the output metrics O2 (AUC), O9 (R^2.adj) and O14 (RUS), but these results did not gain strong support based on the distributions of the individual output metrics (Supplementary material Appendix 1). In general, the individual output metrics separated the simulated scenarios only in few cases on top for separating whether the landscape was patchy or continuous. As some exceptions, the measure and O14 (RUS) was informative for separating specialists from generalists, and the measure O3 (β_{SIM}) was informative from separating long dispersal from short dispersal (Supplementary material Appendix 1).

While the individual output metrics provided only limited information to disentangle among assembly processes, their combinations were more informative. This was especially the case with the combination of output metrics derived from the JSDM model, which were informative in separating more aspects of the simulated scenarios than other types of output metrics (Table 2). All types of output metrics were successful in identifying if the simulated scenario involved a large-scale gradient in the distributions of habitat types, and if the landscape was patchy or continuous, but none of them

were able to identify whether patch quality was uniform or heterogeneous (Table 2). The JSDM approach was the only approach capable of identifying whether metacommunity dynamics were simulated on static or dynamic landscape, and it further identified whether landscape dynamics were slow or fast. Both the beta-diversity indices as well as the JSDM approach were able to separate whether the species were generalists or specialists. The beta-diversity indices, distance-based redundancy analysis, and the JSDM approach were able to identify whether the species followed short or long dispersal, but none of the approaches was able to separate whether the species varied in their dispersal distances. The JSDM approach was the only approach that was able to identify that the analyses missed some of the relevant environmental covariates (Table 2). Importantly, using combined information from all approaches yielded the highest predictive power generally in all cases, suggesting the different analysis types carry complementary information (Table 2). In particular, only the combined approach was able to separate whether resource use specialisation was partial or full.

Discussion

In this work, we have applied statistical analyses to data on simulated species communities to examine what kind of signatures community assembly processes leave on non-manipulative observational data. Beyond the much studied neutral–niche axis, we followed Brown et al. (2017) and considered variation among communities in a multidimensional space consisting of the properties of the involved taxa as well as their abiotic environment. As a result, our metacommunity scenarios included not only the classical paradigms, but also the parameter space between them. In concordance with earlier simulation studies (Smith and Lundholm 2010, Münkemüller et al. 2012, Tucker et al. 2016, Miller et al. 2017, Clappe et al. 2018) we found that many kinds of output metrics can be used to separate niche-based and neutral communities. However, the classification of the scenarios into the classical paradigms reflects only part of the community variation that the assembly processes

Table 2. Predictive power to disentangle among simulated metacommunity scenarios. The values give AUC values for predicting each of the Choices C1–C10 based on 10-fold cross-validation, using as predictors the output metrics related to each of the five analyses types listed in Table 1. The column ALL shows the predictive power if using all output metrics of the five statistical approaches simultaneously. Values smaller than 75% are not shown to emphasize the cases where predictive power is substantial.

Statistical approach	HAB	BETA	db-VP	db-RDA	JSDM	ALL
C1. presence versus absence of gradient	0.94		0.88	0.82	0.79	0.97
C2. patchy versus continuous landscape	1	0.99	0.78	0.97	0.92	1
C3. uniform versus varying patch quality						
C4. dynamic versus.static patches					0.78	0.82
C5. slow versus.fast patch turnover					0.8	0.93
C6. generalists versus specialists		0.76			0.9	0.96
C7. partial versus strict specialists						0.81
C8. variation versus no variation in dispersal strategy						
C9. short versus long dispersal		0.87		0.8	0.75	0.93
C10. missing versus not missing covariates					0.87	0.9

create. Thus, the classical paradigms should be viewed only as special cases in the continuum of metacommunity dynamics generated by the varying underlying processes. Importantly regarding which analyses are most informative to disentangle among metacommunity processes, our results demonstrate that the best practice is to apply different types of statistical approaches simultaneously to gain complementary information (Table 2). Thus, the most comprehensive evaluation of metacommunity structure can be obtained by combining different types of output metrics. This is in line with the JSDM approach being generally the most powerful approach for separating among the scenarios (Table 2), as it provides a wider set of output metrics than the other types of analyses considered here (Table 1).

Our results showed that snapshot data on species occurrences can to a limited extent be informative about dispersal capabilities of the species. Dispersal limitation can be expected to increase dissimilarity among local communities (Condit et al. 2002, Cottenie et al. 2003, Qian 2009), but thus far only few studies have rigorously tested this relationship (Cadotte 2006, Logue et al. 2011, Grainger and Gilbert 2016). Short dispersal is also expected to lead to steep distance decay in community similarity (Qian and Ricklefs 2007, Soininen et al. 2007). Our results were in line with these expectations in the sense that short distance dispersal was related to high Simpson and Sørensen dissimilarities, and high spatial proportion in db-RDA. However, even in the well-controlled simulations, it was generally difficult to disentangle the dispersal mode due to the presence of confounding factors, and thus identifying dispersal mode from empirical patterns of species occurrences includes a high amount of uncertainty. We further note that comparison of beta-diversity indices among studies is difficult also for the reasons that they depend strongly on the spatial scale considered (Loreau 2000, Soininen et al. 2007) and the size of the data (Baselga 2010).

One long-standing assumption in community ecology is that narrow environmental niches promote increased beta-diversity (Janzen 1967, Whittaker 1975). Several empirical studies have found that high resource diversity combined with resource use specialization is positively related to beta-diversity (Buckley and Jetz 2008, Jankowski et al. 2009). Our results were in line with this expectation, as we found communities consisting of specialists to be characterized by high overall beta diversity. Further, the observed level of habitat specialization that we derived from the model-based approach reflected well the assumed level of habitat specialization, meaning that snapshot data on species occurrence can carry a reliable signal of specialization level.

In our results, many of the output metrics strongly differed between communities inhabiting patchy and continuous landscapes. Similar results have been found in empirical studies showing e.g. that spatial isolation among habitat patches increases nestedness (Bender et al. 2017, Gianuca et al. 2017). The model-based results showed that linking snapshot data on species occurrences to prevailing habitat variation can inform about habitat turnover rate (Table 2), the signature of habitat turnover being given by low predictive power and high amount of variation attributed to random effects (Supplementary material Appendix 1). This is because in a dynamic landscape species occurrence does not depend on prevailing habitat conditions only but also on past ones, the latter representing missing covariates from the viewpoint of statistical modelling (Hodgson et al. 2009).

A strength of our study is that all scenarios were based on the same underlying agent-based model, and thus any differences among them is directly related to variation in the underlying assumptions. However, at the same time it is clear that our scenarios are simplified caricatures of reality, real communities being influenced by myriads of factors not considered here. Thus, the results provided in this study should be compared to results from real communities only qualitatively, not quantitatively. For example, while we found that the Simpson index β_{SIM} was higher for scenarios that involved short than long dispersal (Supplementary material Appendix 1), it is not possible to give a quantitative threshold value above which β_{SIM} would indicate 'short' and below which 'long' dispersal, or to link β_{SIM} to the actual length scale of the dispersal. This is because of β_{SIM}, as well as all the other output metrics, are simultaneously influenced by many factors. Yet, we consider our results highly useful because they show which output metrics are informative about each of the underlying assembly processes. For example, the researcher may compare β_{SIM} computed for the same empirical design for different subsets of species, to test whether these subsets of species are likely to vary in their dispersal distances. The interpretation of β_{SIM} as a measure of dispersal distance however assumes that 'everything else remains constant', e.g. that the spatial variation in resource availability is the same for both subsets of species, which may not be the case even if the species inhabit the same landscape.

A limitation of our study is that our analyses were restricted to spatial snapshot-data on species presence–absences derived from a stationary state of an agent-based simulation model. We hope future research to examine how much more information about metacommunity processes can be inferred if more rich data are available, such as abundance data, time-series data or the possibility to combine community data with trait-data and phylogenetic data. We further hope future research to test the robustness of our results with respect to other ecological situations, such as communities in a transient state.

References

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

Bender, M. G. et al. 2017. Isolation drives taxonomic and functional nestedness in tropical reef fish faunas. – Ecography 40: 425–435.

Brown, B. L. et al. 2017. Making sense of metacommunities: dispelling the mythology of a metacommunity typology. – Oecologia 183: 643–652.

Buckley, L. B. and Jetz, W. 2008. Linking global turnover of species and environments. – Proc. Natl Acad. Sci. USA 105: 17836–17841.

Cadotte, M. 2006. Dispersal and species diversity: a meta-analysis. – Am. Nat. 167: 913–924.

Cale, W. G. et al. 1989. Inferring process from pattern in natural communities. – Bioscience 39: 600–605.

Carpenter, S. R. 1996. Microcosm experiments have limited relevance for community and ecosystem ecology. – Ecology 77: 677–680.

Chang, L. et al. 2013. Better environmental data may reverse conclusions about niche- and dispersal-based processes in community assembly. – Ecology 94: 2145–2151.

Chase, J. M. 2003. Community assembly: when should history matter? – Oecologia 136: 489–498.

Chase, J. M. and Bengtsson, J. 2010. Increasing spatio-temporal scales: metacommunity ecology. – In: Verhoef, H. A. and Morin, P. J. (eds), Community ecology: processes, models and applications. Oxford Univ. Press, pp. 57–68.

Chase, J. M. and Myers, J. A. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. – Phil. Trans. R. Soc. B 366: 2351–2363.

Chave, J. and Leigh, E. G. 2002. A spatially explicit neutral model of β-diversity in tropical forests. – Theor. Popul. Biol. 62: 153–168.

Clappe, S. et al. 2018. Beyond neutrality: disentangling the effects of species sorting and spurious correlations in community analysis. – Ecology 99: 1737–1747.

Clark, J. S. et al. 2017. Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. – Ecol. Monogr. 87: 34–56.

Condit, R. et al. 2002. Beta-diversity in tropical forest trees. – Science 295: 666–669.

Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. – Ecol. Lett. 8: 1175–1182.

Cottenie, K. et al. 2003. Zooplankton metacommunity structure: regional vs. local processes in highly interconnected ponds. – Ecology 84: 991–1000.

D'Amen, M. et al. 2017. Spatial predictions at the community level: from current approaches to future frameworks. – Biol. Rev. 92: 169–187.

Gianuca, A. T. et al. 2017. Effects of dispersal and environmental heterogeneity on the replacement and nestedness components of β-diversity. – Ecology 98: 525–533.

Götzenberger, L. et al. 2012. Ecological assembly rules in plant communities-approaches, patterns and prospects. – Biol. Rev. 87: 111–127.

Grainger, T. N. and Gilbert, B. 2016. Dispersal and diversity in experimental metacommunities: linking theory and practice. – Oikos 125: 1213–1223.

Hodgson, J. A. et al. 2009. Metapopulation responses to patch connectivity and quality are masked by successional habitat dynamics. – Ecology 90: 1608–1619.

Holyoak, M. et al. 2005. Metacommunities: spatial dynamics and ecological communities. – Univ. of Chicago Press.

Jankowski, J. E. et al. 2009. Beta diversity along environmental gradients: implications of habitat specialization in tropical montane landscapes. – J. Anim. Ecol. 78: 315–327.

Janzen, D. H. 1967. Why mountain passes are higher in the tropics. – Am. Nat. 101: 233–249.

Kuglerová, L. et al. 2015. Local and regional processes determine plant species richness in a river-network metacommunity. – Ecology 96: 381–391.

Legendre, P. and Anderson, M. J. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. – Ecol. Monogr. 69: 1–24.

Legendre, P. et al. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. – Ecol. Monogr. 75: 435–450.

Legendre, P. et al. 2009. Partitioning beta diversity in a subtropical broad-leaved forest of China. – Ecology 90: 663–674.

Leibold, M. A. and McPeek, M. A. 2006. Coexistence of the niche and neutral perspectives in community ecology. – Ecology 87: 1399–1410.

Leibold, M. A. et al. 2004. The metacommunity concept: a framework for multi-scale community ecology. – Ecol. Lett. 7: 601–613.

Logue, J. B. et al. 2011. Empirical approaches to metacommunities: a review and comparison with theory. – Trends Ecol. Evol. 26: 482–491.

Loreau, M. 2000. Are communities saturated? On the relationship between α, β and γ diversity. – Ecol. Lett. 3: 73–76.

Lortie, C. J. et al. 2004. Rethinking plant community theory. – Oikos 107: 433–438.

McArdle, B. H. and Anderson, M. J. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. – Ecology 82: 290–297.

Miller, E. T. et al. 2017. Phylogenetic community structure metrics and null models: a review with new methods and software. – Ecography 40: 461–477.

Morin, P. J. 2011. Community ecology, 2nd ed. – Wiley.

Münkemüller, T. et al. 2012. From diversity indices to community assembly processes: a test with simulated data. – Ecography 35: 468–480.

Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – Ecol. Lett. 20: 561–576.

Qian, H. 2009. Beta diversity in relation to dispersal ability for vascular plants in North America. – Global Ecol. Biogeogr. 18: 327–332.

Qian, H. and Ricklefs, R. E. 2007. A latitudinal gradient in large-scale beta diversity for vascular plants in North America. – Ecol. Lett. 10: 737–744.

Schindler, D. W. 1998. Whole-ecosystem experiments: replication versus realism: the need for ecosystem-scale experiments. – Ecosystems 1: 323–334.

Smith, T. W. and Lundholm, J. T. 2010. Variation partitioning as a tool to distinguish between niche and neutral processes. – Ecography 33: 648–655.

Soininen, J. et al. 2007. The distance decay of similarity in ecological communities. – Ecography 30: 3–12.

Tucker, C. M. et al. 2016. Differentiating between niche and neutral assembly in metacommunities using null models of β-diversity. – Oikos 125: 778–789.

Tuomisto, H. et al. 2003. Dispersal, environment and floristic variation of Western Amazonian forests. – Science 299: 241–244.

Vanschoenwinkel, B. et al. 2007. The role of metacommunity processes in shaping invertebrate rock pool communities along a dispersal gradient. – Oikos 116: 1255–1266.

Vellend, M. 2010. Conceptual synthesis in community ecology. – Q. Rev. Biol. 85: 183–206.

Warton, D. I. et al. 2015. So many variables: joint modeling in community ecology. – Trends Ecol. Evol. 30: 766–779.

Weinstein, B. G. et al. 2014. Taxonomic, phylogenetic and trait beta diversity in South American hummingbirds. – Am. Nat. 184: 211–224.

Whittaker, R. H. 1975. Communities and ecosystems. – Macmillan.

Zurell, D. et al. 2010. The virtual ecologist approach: simulating data and observers. – Oikos 119: 622–635.

Zurell, D. et al. 2018. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? – Ecography 41: 1812–1819.