The Internal Structure of Metacommunities

Authors and Affiliations:

Mathew A. Leibold, Department of Biology, University of Florida. 32611

mleibold@ufl.edu

F. Javiera Rudolph, Department of Biology, University of Florida.

javiera.av@ufl.edu

F. Guillaume Blanchet, Département de Biologie, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, Québec, J1K 2R1, Canada.

guillaume.blanchet@usherbrooke.ca

Luc De Meester, Laboratory of Aquatic Ecology, Evolution and Conservation, University of Leuven, 3000 Leuven, Belgium; Leibniz Institut für Gewässerökologie und Binnenfischerei (IGB), 12587 Berlin, Germany; Institute of Biology, Freie Universität Berlin, Königin-Luise-Strasse 1-3, 14195 Berlin, Germany; Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Königin-Luise-Str. 2-4, 14195 Berlin, Germany

Luc.DeMeester@kuleuven.be

Dominique Gravel, Département de Biologie, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, Québec, J1K 2R1, Canada.

Dominique.Gravel@USherbrooke.ca

Florian Hartig, Theoretical Ecology, University of Regensburg, Germany, florian.hartig@ur.de

Pedro Peres-Neto, Department of Biology, Concordia University.

pedro.peres-neto@concordia.ca

Lauren Shoemaker, Botany Department, University of Wyoming.

lshoema1@uwyo.edu

Jonathan M. Chase, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena Leipzig; Department of Computer Sciences, Martin Luther University Halle-Wittenberg; jonathan.chase@idiv.de

Running title: Internal Metacommunity Structure

Keywords: Interspecific competition, site dynamic theory, hierarchical community models, metacommunity paradigms, variation partitioning.

Data accessibility statement: Should this manuscript be accepted, the data and code supporting the results will be archived in an appropriate public repository such as Github and will be mirrored at Zendo or similar locations.

Abstract

Current analyses of metacommunity data largely focus on global attributes across the entire metacommunity, such as mean alpha, beta, and gamma diversity, as well as the partitioning of compositional variation across all species and sites. This view neglects that different species and sites in the landscape can vary widely in how they contribute to these metacommunity-wide attributes as a function of traits and site attributes. We argue that the study of this internal structure can help define and describe the complex and interactive relations between process and pattern in metacommunities. To illustrate this general idea, we create synthetic data using a simple colonization-extinction metacommunity model, and quantify variation between species and sites (what we call the ‘internal structure’ of metacommunities) using Joint Species Distribution Models. We find that this perspective on internal structure of metacommuniites

provides useful information about the distinct ways that different species and different sites contribute to metacommunity attributes. We conclude with some discussion about how these realistic complexities regarding internal structure can be incorporated into a more cohesive metacommunity theory.

Introduction

Recent developments in community ecology indicate that the field is undergoing an important renaissance in both its concepts and tools. One of the more exciting and important elements of this renaissance is in the use of the metacommunity concept, which recognizes the feedback between local communities and the broader-scale regional biota (Hanski and Gilpin 1991, Leibold et al. 2004 and reviewed in Leibold and Chase 2017). Earlier work focused on specific scenarios where such feedbacks could occur (e.g., Levins and Culver 1971, Horn and MacArthur 1974, Levin 1974, Sloan-Wilson 1992, Leibold 1998, Hubble 2001, Amarasekare and Nisbet 2001), which was synthesized into a useful categorization of the different ways in which metacommunities could be structured (Leibold et al. 2004). However, it is apparent that there is a much more complex and nuanced spectrum of possibilities regarding the mechanisms and processes underlying the structure of metacommunities (Leibold and Chase 2017). Ongoing developments, including both more sophisticated theoretical (e.g., Shoemaker and Melbourne 2016, Fournier et al. 2017, Ovaskainen et al. 2019, Thompson et al. 2020) and analytical (e.g., Legendre and DeCaceres 2013, Hui et al. 2013, Ovaskainen et al. 2017, Ohlman et al. 2018, Jabot et al. 2020) approaches, promise a new level of understanding and synthesis of the regional-local community-level feedbacks. It seems increasingly apparent that understanding the feedbacks between local communities and the regional biota is a major challenge for community ecology in general (see Leibold and Chase 2017) as well as for the application of these insights to applied issues of environmental and health concerns (e.g., Bengtsson 2009, Schiesari et al. 2019, Miller et al. 2019, Brown and Barney 2020).

While important progress has been made in the study of metacommunities, most work tends to assume that processes act similarly on all species and sites in a metacommunity. However, variation among species and among sites can strongly influence metacommunity-level properties of biotas. This is particularly relevant to local-regional feedbacks because species and sites vary across metacommunities. For instance, the spatial structure of environmental features among sites almost always changes different explicitly determined landscapes due to spatial contingency (Peres-Neto et al. 2012). In contrast with many previous studies (e.g., Blanchard et al. 2020, Jabot 2020), we emphasize that different species can be heterogenous in how they contribute to metacommunity level properties and that different sites can also be heterogenous in how they contribute to these patterns (but see Pandit et al. 2009, Legendre and De Cáceres 2013). Using this concept as a starting point, we highlight the importance of understanding the ‘internal structure’ of metacommunities, which can help resolve more complex dynamics that could not be resolved without taking this different perspective. Community assembly in metacommunities is a complex process that involves mechanisms related to environmental effects, species interactions, dispersal, and stochasticity as they act on a regionally-defined pool of species within a specific landscape (e.g. Vellend 2010, 2015, Weiher et al. 2011). Current analytical tools for inferring the effects of these mechanisms from metacommunity patterns focus on global metrics that describe the entire metacommunity, such as diversity metrics and coexistence patterns and variation partitioning analysis (e.g., Borcard et al. 1992, Gotelli and McCabe 2002, Leibold and

Mikkelson 2002). Recent efforts have used several of these global metrics to separate metacommunity processes (Ovaskainen et al. 2019, Guzman et al. 2020). While these approaches provide insights into the processes that drive species distributions and determine their levels of interaction within metacommunities, they only characterize them in very broad terms at the larger metacommunity scale, which we consider to be the 'external' structure of metacommunities. Here, we focus on the 'internal structure' instead and thus on resolving how individual species and sites or patches contribute to such broad patterns by dissecting their individual contribution to the global metacommunity pattern (see also Fournier et al. 2017, Suzuki and Economo 2021).

To illustrate the advantages of studying the internal structure of metacommunities, we create synthetic data from a process-based metacommunity model of competition-extinction dynamics. We evaluate species-level and site-level variation using joint species distribution models (JSDMs; see review in Warton et al. 2016). We find that there can be tremendous heterogeneity between the contributions of individual species and sites to overall metacommunity structure; and that some of this heterogeneity can be identified by JSDMs and be related to attributes such as dispersal and species associations for species, or environmental uniqueness for sites.

Previous work has shown that particular examples of such effects can occur but here we seek a more general approach. For example, Pandit et al. (2009) showed that species can be heterogenous in their responses to environmental and stochastic factors depending on their degree of habitat specificity. Others have argued that JSDMs can provide important insights into the drivers of such variation in species distributions ((Hui et al. 2013, 2016, Pollock et al. 2014, Ovaskainen et al. 2017 see also Ovaskainen and Abrego 2020; but see Poggiato et al. 2021). Similarly, the heterogeneity among sites has long been identified as driving individual species distributions (see Guisan and Thulliers 2005, Soberon and Peterson 2005, Elith and Leathwick 2006) as well as driving overall variation among sites in global metrics of community structure (e.g. diversity patterns, etc.) as characterized by the field of landscape ecology (Turner et al. 2001). Further suggestions that individual sites might vary in how they contribute to metacommunity dynamics include the concept of 'keystone communities' (Mouquet et al. 2012, Resetarits et al. 2017, Yang et al 2020) and metacommunity approaches to spatial networks. These disparate approaches (species vs sites) to metacommunities are likely closely related to each other and can be linked by the emerging methodologies of methods such as JSDMs to a more nuanced metacommunity ecology that recognizes a plurality of mechanisms and processes underlying community assembly. Nevertheless, we discuss some remaining important challenges to resolve in making process-pattern linkages in metacommunities (see also Poggiato et al. 2021, Miele et al. 2021).

Quantifying the link between process and pattern using a simple metacommunity simulation and refined statistical approach

To test and exemplify our ability to infer individual species and site contributions, we simulate data from a process-based model, which allows us to create observations with full knowledge about the underlying mechanisms. Our process-based model is based on a spatial implementation of spatially implicit site occupancy models (e.g., Levins and Culver 1971, Horn and MacArthur 1972, Levin 1974, Hastings 1980, Hanski 1991) to describe dynamics in heterogeneous metacommunities. We focused here on a model for predicting presence-absence (and not abundance; but see Supporting Information) because it is the most widely available type of empirical data for metacommunity analysis. For each species in each patch, we model occupancy using two key equations (details in the Supporting Information). The first of these describes the colonization of patch *z* by species *i* during a discrete time interval, Δt:

$$P\left(X_{i,z,t+\Delta t} = 1 | X_{i,z,t} = 0\right) \; = \; I_{i,z,t} S_{i,z,t} C_{i,z,t} \qquad (1)$$

where $X_{i,z,t}$ is a stochastic variable representing the occurrence of species i at location z at time t, $I_{i,z,t}$ is the number of immigrants of species, $S_{i,z,t}$ is the effect of environmental filtering on the probability of establishing a viable local population, and $C_{i,z,t}$ is the effect of ecological interactions on the establishment probability. Second, we consider the alternative possibility: the extinction of species *i* in patch *z* during the time interval Δt:

$$P\left(X_{i,z,t+\Delta t} = 1 | X_{i,z,t} = 0\right) \; = \; M_{i,z,t} E_{i,z,t} \qquad (2)$$

where $M_{i,z,t}$ and $E_{i,z,t}$ are the responses of the extinction probability to the local environment and to ecological interactions, respectively.

At steady state the solution to this model is:

$$\log\frac{\hat{p}^{i,z}}{1-\hat{p}^{i,z}} = \log I^{i,z} + \log\frac{S^{i,z}}{M^{i,z}} + \log\frac{C^{i,z}}{E^{i,z}} \tag{3}$$

where $\hat{P}^{iz}$ is the expected probability that site *z* is occupied by species *i*. This formulation assumes that immigration($I^{i,z}$), 'environmental selection' ($S^{i,z}$ *and* $M^{i,z}$) and interactions ($C^{i,z}$ *and* $E^{i,z}$) can be separated into distinct effects that do not interact.

Equation 3 can be analyzed by a JSDM to separate the contributions of these effects into spatial effects (driven by immigration), environmental filtering (driven by abiotic selection) and species co-distribution unrelated to either space or environment, with an additional fraction quantifying residuals resulting from stochasticity in the case of a finite number of patches (see also Shoemaker et al. 2020). Furthermore, the likelihood of every observation can be marginalized over each species (by summing the likelihoods for a given species across all patches) to describe the variation among species. Alternatively, the likelihood can be marginalized by sites (by summing the likelihoods for a patch across all species) to describe the variation across the metacommunity landscape. Doing so, we can quantify the importance of environment, species co-distribution, and space for predicting metacommunity structure as a whole, as well as their importance for predicting the presence-absence (or, in theory, the abundance) of individual species or community composition at individual patches.

Analyzing the inferred quantities, we can study the internal structureof the metacommunity. This can clarify how these key processes work together in a diverse and complex metacommunity to explain species distributions within the metacommunity (Figure 1). We also emphasize that one main advantage of the proposed framework is to allow ecologists to identify the contributions of individual communities and species which are often units of personal interest and knowledge (e.g., natural history of species, place-based ecological knowledge).

To illustrate the utility of this approach in a more realistic framework that includes stochasticity and spatially explicit landscapes, we implemented the key processes of drift, environmental filtering, dispersal, and species interactions (Vellend 2010, 2016) in a flexible simulation version of the model (described in more details in the Supporting Information). The simulation model allows us to vary each process separately for each species in a heterogeneous spatially explicit landscape. It simulates the dynamics of a metacommunity across a set of patches and generates a spatial network that specifies the connectivity among patches. The state variables of the simulation are the occupancy of each species in every patch (i.e. presence/absence, though future implementations could also address abundance data, e.g. Rybicki et al. 2018, Ovaskainen et al. 2019, Thompson et al. 2020). Each patch can be colonized from nearby patches depending on their location in the landscape, dispersal rate of the species and proximity of extant populations in neighboring patches. Each species in each patch is subject to extinctions that reflect demographic and/or environmental stochasticity. Patches can differ in local environmental conditions that differentially influence baseline colonization and extinction probabilities. Species interactions are modeled in two ways. First, the presence of other species in a patch can modify baseline colonization probability (a reduction in the case of competition). Second, co-occurring species can modify baseline extinction probability (an increase in the case of competition).

We next apply a JSDM to the resulting distribution of species among patches. Specifically, we use the HMSC R package (Ovaskainen et al. 2017, Ovaskainen and Nerea 2020) that models species distributions as a function the environment, spatial autocorrelation and species co-distributions. After fitting the model using HMSC, we partition variation between the model processes into four components (or fractions) using an approach akin to classic variation partitioning (Borcard et al. 1992; Peres-Neto et al. 2006). Details about how this type of variation partitioning is computed through HMSC is given in the Supporting Information. Specifically, the variation was partitioned to quantify the effects of environment (labeled [E]), spatial patterning (labeled [S]), co-distribution among species (labeled [C]) and residual (unexplained) variation that cannot be attributed to any of the three previously mentioned fractions (i.e. sets of predictors). This is expressed as $1-R^2$, where R^2 is the proportion of variation explained by the model and includes fractions [E], [S], and [C].

Compared to the classical method of variation partitioning (Borcard et al. 1992, Peres-Neto et al. 2006), this approach provides two major advances. The first is that we can infer a third fraction that quantifies co-variation (or co-distribution) among species using latent variables (fraction [C]), which is distinct from either the environment effect (fraction [E]) or the spatial effects (fraction [S]). In our model this fraction reflects the consequences of species associations rather than biotic interactions, and we emphasize that this interpretation must be done with caution (see Dormann et al. 2018 and Blanchet et al. 2020 for critical reviews of this issue). In nature, this latent fraction may also reflect other sources of variation such as those from unmeasured environmental variables. Similar points have been raised in metacommunity analysis regarding the interpretation of model variation due to spatial autocorrelation as these could be related to spatialized environmental effects due to unmeasured environmental factors (Peres-Neto et al. 2012). The second advance is that all fractions are calculated at the species level, where species-specific variation detail the internal structure of the metacommunity from the perspective of each individual species. Fractions of the variation at the metacommunity level are then calculated by averaging across all species to obtain metacommunity (global) level descriptors as those obtained from variation partitioning. This allows us to subsequently examine potential drivers (environment, space and joint co-distribution) underlying the variation in these different effects among species. In addition, we implemented a partitioning scheme to quantify how each site separately contributes to different drivers as described in the Supplement Information. This way to partition variation among sites informs us about the individual contribution each site has to variation in the data for each set of drivers. The ability to resolve species-specific and site-specific components means that we can study the 'internal structure' of the metacommunity in ways that were absent in classic variation partitioning.

Although the analytical model described in equations 1-3 suggests that making links between processes and patterns using JSDMs are possible, we wished to evaluate if this was also likely in less idealized situations such as those used in our metacommunity model. To do this, we simulated a number of scenarios that vary the strength of environmental selection, dispersal and competition. Comparing scenarios with varying niche breadth and competition (scenarios A-D, Fig. 2 and Fig 3), and a more complex case where species compete and vary in both dispersal and in their responses to the environment (scenario G, Figure 4, scenarios E and F in the SI) highlight how HMSC provides an avenue for distinguishing between underlying processes based on abundances of species across metacommunity patches. Using our framework, our goal here is to illustrate how links between pattern and process might be made in metacommunities. In doing so, we leave a more extensive and systematic evaluation of the model’s components (e.g., performance of JSDMs under multiple complex scenarios) for future work (but see Ovaskainen and Nerea 2020).

Simulation experiments

In a first set of simulations, we considered a situation where species had distinct environmental optima along an environmental gradient and had limited dispersal (Figure 2, Table 1). We contrasted the case where the environmental niches were narrow (steep changes in baseline colonization success and extinction rates with small deviations in environment) with the case with identical optima, but with wide environmental niches (much weaker changes in colonization and extinction with environmental value). As expected, we find that these differences in environmental niche breadth had strong effects on the relative importance of environmental filtering (fraction [E]) versus spatial patterning (fraction [S]). Specifically, we find stronger spatial patterning when niche breadths were broad and stronger environmental filtering when niches were narrower (Figure 2, Table 1). We also found that the R^2 values were higher for the case with narrow niches than with wide niches. Finally, we found non-zero (though relatively weak) variance components for co-distributions (fraction [C]) in both cases, especially when niches were broad even though our analytical model would predict the absence of such variation components since colonization and extinctions were not affected by species interactions in these simulations.

We next simulated metacommunities with identical parameters as above, except with added interspecific competition effects (Figure 3, Table 1). As in the case without species interactions (compare with Figure 2), narrow niches enhanced the relative strength of environmental filtering (fraction [E]) and reduced spatial patterning (fraction [S]) when compared to wide niches. In these simulations, however, the co-distribution components (fraction [C]) were much more substantial than without species interactions. We also found that adding interspecific competition substantially increased the total amount of variation explained (R^2) by the model (i.e. due to the joint component of co-distribution).

We conducted a number of other simulations to explore if that describe interspecific variation on environment (fraction [E]), space (fraction [S]), and co-distributions (fraction [C]) depend on the dispersal, niches breadth, and interactions. Illustrative examples are shown in the supplemental information and summary statistics are shown in Table 1. In Figure 4, we present the results from one of these examples that includes heterogenous dispersal to show how the internal structure can reveal how dispersal variation affects species distributions. We found that one could distinguish species by the degree to which their distributions are related to environment (fraction [E]), space (fraction [S]) and co-distributions (fraction [C]) (Figure 5a) and we found that this could be related to their traits (i.e., species optima in our simulation framework). Species with higher dispersal ability and more specialized environmental niche positions had distributions better predicted by the environment than those that were dispersal limited and had distributions that presented a higher level of spatial autocorrelation (fraction [S]). Species with optima closer to the middle of the environmental gradient also had a larger fraction [C] than those with more extreme optima.

Sites also differed in how their species composition was related to environmental (fraction [E]) and spatial effects (fraction [S]) as well as co-distributions (fraction [C] - Figure 4b). Some sites tended to be occupied by locally dominant species (in the lower left of the ternary plot, nearer to fraction [E]), while others were occupied by species found in nearby sites (lower right of the ternary plot, nearer to fraction [S]). Some sites were also occupied by combinations of species that were differentially associated with each other regardless of environment or dispersal (upper apex of the ternary plot, nearer to fraction [C]). As can be seen in Figure 4b, there were also a wide range of intermediate conditions. A major driver of this variation is the local environmental condition, especially in relation to how distinct the local environment is from the overall mean environment of the metacommunity (Figure 5b).

We further investigated the structure of the species co-distribution (fraction [C]). This covariation can be directly attributable to species interactions because we explicitly model the processes underlying metacommunity dynamics. However, even in our model, species co-distribution may not directly link to pairwise interaction coefficients, but rather may emerge as a complex relationship between species interactions and environmental conditions (Cazelles et al. 2015, Blanchet et al. 2020). To illustrate this, we show the co-distribution among species as a heat map separately for each of the five individual simulations presented in Figure 4 and compared them to the actual interaction matrix that describes interspecific competition in our model (Figure 6; similar heat maps obtained with the other scenarios are shown in the Supplement Information). Despite the fact that the same interaction matrix was used for all five of these simulations, the resulting co-distribution patterns are inconsistent in their details. However, these matrices show that there is consistency in several features of the co-distribution pattern. For example, they all share the predominance of strong negative correlations along the main diagonal that match the interaction matrix we used. They also share a strong 'checkerboard' pattern with alternating negative and positive co-distributions between species when these are ranked against their environmental optima. Given the simple scheme of species interactions we used (Fig. 5 and SI), these results are consistent with the predictions that direct interactions are stronger than indirect ones and tend to weaken with the number of links in indirect chains even if the details of these effects are less predictable.

Discussion

While metacommunity ecology has made great progress in the past decades, the assumptions that species and sites were relatively homogeneous in their underlying processes (reviewed in Leibold and Chase 2017) have limited the applicability of metacommunity theory to more realistic species pools within complex landscapes. Here, by combining a tractable process-based model with emerging analytical methods, we provide a general quantitative approach that accounts for multiple interacting processes that may operate differently among species or in different parts of landscapes.

Although there are some important challenges to consider, our study illustrates important insights about the internal structure of metacommunities, including:

1) Variation partitioning using JSDMs (here implemented using HMSC) can be linked to basic processes of community assembly at the species and sites levels (e.g. environmental selection, dispersal, biotic selection, and drift).

2) Quantifying co-distributions of species in metacommunities can improve predictive ability even when the processes that generate these distributions are complex (stochasticity, complex spatial landscapes, and species interactions).

3) Species can have distributions that vary in the degree to which they are determined by combinations of the basic community assembly processes depending on features of their ecology (e.g. dispersal and environmental preferences); and

4) The predominant assembly processes that determine local communities can differ among adjacent sites in a metacommunity (e.g. sites that are occupied by species most fit for environmental conditions vs sites occupied by species in nearby sites that may differ in environmental conditions).

It is important to emphasize that there remain some substantial challenges in moving forward with the overall approach we advocate in this paper. This includes some technical issues, such as the estimation of parameters and interpretation of results in more complex metacommunity models in JSDMs, as well as conceptual ones, such as accounting for other processes such as speciation, local adaptation, historical biogeography. Nevertheless, we see that such endeavor will be fruitful, allowing a deeper understanding of ecological dynamics in more realistic, but necessarily complex, spatial landscapes.

Our analytical framework simplifies several potentially complex processes (e.g. non-linearities and interactive effects of mechanisms) into an approximation involving colonization-extinction dynamics. It is possible that more realistic and complex mechanisms driving these processes will weaken associations between pattern and process or create biases in the partitioning of the variation revealed by JSDMs. However, the developments of JSDMs are still progressing, and we anticipate that future developments will solve some of these problems (see Wilkinson et al. 2020).

The co-distribution component of the JSDMs (fraction [C]) is particularly concerning. We find that this component can be biased, especially when species have broad environmental niches (Figure 2). It is widely recognized that there are particular challenges in quantifying and interpreting species co-distributions (reviewed by Blanchet et al. 2020). Perhaps the most obvious challenge is that this component can reflect any process that is inadequately quantified by the environmental (fraction [E]) and spatial components (fraction [S]). In addition to species interactions, this would include, for example, unmeasured environmental factors (see Blanchet et al. 2020) or inadequately quantified landscape attributes. Teasing apart the effects of species interactions from these confounding factors should thus be a major focus for future work.

Nevertheless, it is important to understand that including the co-distribution component in our analyses allows us to account for them, rather than lumping them with residual variation where they have likely given a greatly exaggerated impression of stochasticity.

The basic framework we used to address the process to pattern links is most transparent for mechanisms that focus on interspecific competition because of the analogies with evolutionary genetics (Vellend 2010, 2016). However, species interactions in metacommunities are much more variable and include consumer-resource interactions, mutualisms, and facilitative interactions. Although such interactions can easily be incorporated in simulations, the interpretation that might link process to pattern in such cases are likely to become more complex. Likewise, future work could include local (co-)evolutionary dynamics (see Urban et al. 2020) and historical effects of biogeography and speciation (e.g. Leibold and Chase 2018, Overcast et al. 2020). Here, we have also retained a simple two-level perspective on spatial scale (local discrete sites in a broader regional landscape). It is increasingly apparent that metacommunity dynamics occur over multiple nested scales and that habitats can be continuous and/or nested, rather than discretely patchy (e.g. Munkenmuller et al. 2012, Rybicki et al 2018, Ovaskainen et al. 2019, Viana and Chase 2019, König et al. 2021)Refining our approach to address multiple spatial scales is a logical next step.

Finally, it is increasingly clear that temporal dynamics of community change in metacommunities can provide critical insights about the mechanisms that drive metacommunity patterns (e.g. Jabot et al. 2020, Blanchard et al. 2020, Guzman et al. 2021). We imagine future work on the internal structure of metacommunities as being very amenable to incorporating temporal changes (see for example, Ovaiskainen et al. 2017 for an initial step in this direction). Here we have focused on purely spatial approaches because there are still remarkably few studies that might permit sufficiently structured data to permit temporal analyses and because the limitations and challenges of such analyses are not yet clear.

We have argued that a focus on the internal structure of metacommunities, by examining site-specific and species-specific variation components, can enhance our study of the links between process and pattern in the distributions of species and the occupancies of sites. We also argue that continued work on this focus is essential if metacommunity ecology is to address the dynamics and structure of realistic metacommunities that have typically high biodiversity and occur in complex landscapes. This focus on internal structure represents a shift from traditional approaches that used descriptors of overall variation components at the metacommunity scale that generalizes previous ad hoc approaches to similar internal variation in metacommunity patterns (e.g. Pandit et al. 2009, Legendre and De Cáceres 2013). The dynamics and structure of distributions of realistically diverse species in a realistically structured landscape of sites likely involves the interaction of community assembly processes including environmental filtering, dispersal, and drift, and these are unlikely to be adequately described by simple metacommunity level metrics (see Ovaskainen et al. 2019). Consequently, dissecting the internal structure of metacommunities on the basis of species and site contributions could provide key insights into the processes underlying metacommunity assembly. Such insights might be particularly useful in the management of landscapes and metacommunities for conservation purposes since they focus on particular units (species or sties) that are often the focus of concern in such cases.

Acknowledgments: This paper emerged from workshops funded with the support (to JMC) from the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the German Research Foundation (FZT 118). MAL would like to acknowledge the

Alexander von Humboldt Foundation for a Research Award that helped fund this work. LDM acknowledges KU Leuven Research Fund project C/2017/002 and FWO projects G0B9818 and G0C3818. We also thanks F. Engel, C. Rakowski, K. Taylor, R. Pelinson, X. Zhao, L. Juen, T. Michelan and A. Rudolf for comments on earlier versions of the manuscript.

References

Amarasekare, P. , & Nisbet R. M. (2001). Spatial heterogeneity, source-sink dynamics, and the local coexistence of competing species. *The American Naturalist,* 158, 572-584.

Bengtsson J. (2009). Applied (meta)community ecology: diversity and ecosystem services at the intersection of local and regional processes. In: Verhoef HA and Morin PJ (Eds), Community Ecology: Processes, Models, and Applications. Oxford, UK: Oxford University Press.

Blanchard, G., Birnbaum, P., & Munoz, F. (2020). Extinction–immigration dynamics lag behind environmental filtering in shaping the composition of tropical dry forests within a changing landscape. *Ecography,* 43, 869-881.

Blanchet, F.G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–1055.

Boulangeat, I., Gravel, D., & Thuiller, W. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters,* 15, 584-593.

Brown, B.L., & Barney, J.N. (2021). Rethinking Biological Invasions as a Metacommunity Problem. *Frontiers in Ecology and Evolution,* 8, 584701.

Brown, B.L., Sokol, E.R., Skelton, J., & Tornwall, B. (2017). Making sense of metacommunities: dispelling the mythology of a metacommunity typology. *Oecologia*, 183, 643–652.

Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics: Meta-analysis of metacommunities. *Ecology Letters*, 8, 1175–1182.

Dormann, C.F., Bobrowski, M., Dehling, M. et al. (2018). Biotic interactions in species distribution modelling: ten questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, 27,1004-1016.

Elith, J., & Leathwick, J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.

Fournier, B., Mouquet, N., Leibold, M.A., & Gravel, D. (2017). An integrative framework of coexistence mechanisms in competitive metacommunities. *Ecography*, 40, 630–641.

Gotelli, N.J., & McCabe, D.J. (2002). A meta-analysis of J. M. Diamond's assembly rules model. *Ecology*, 83, 2091-2096.

Gravel, D., Canham, C.D., Beaudet, M., & Messier, C. (2006). Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, 9, 399–409.

Gravel, D., & Massol, F. (2020). Toward a general theory of metacommunity ecology. *In*: McCann, K.S. & Gellner, G. (eds.). Theoretical Ecology. Oxford University Press. *In press.*

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.

Guzman, L.M., Germain, R.M., Forbes, C., Straus, S., O'Connor, M. I., Gravel, D. Srivastava, D.S., & Thompson, P.L. (2019). Towards a multi-trophic extension of metacommunity ecology. *Ecology Letters,* 22, 19-33.

Guzman, L.M., Thompson, P.L., Viana, D.S., Vanshoenwinkel, B., Horvath, Z., Ptacnik, R., Jeliazkov, A., Gascon, S., Lemmens, P., Anton-Pardo, M., Langenheder, S., De Meester, L., & Chase, J.M. (2021). Disentangling metacommunity processes using multiple metrics in space and time. https://www.biorxiv.org/content/10.1101/2020.10.29.361303v1

Hanski, I. (1991). Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society of London. 42, 17-38.*

Hanski, I., & M. Gilpin. (1991). Metapopulation dynamics: brief history and conceptual domain. *Biological Journal of the Linnean Society of London. 42, 3-16.*

Hastings, A. (1980). Disturbance, coexistence, history, and competition for space. *Theoretical Population Biology*, 18, 363–373.

Horn, H.S., & MacArthur, R.H. (1972). Competition among fugitive species in a harlequin environment. *Ecology*, 53, 749–752.

Hubbel, S. P. (2001). The Unified Neutral theory of Biodiversity and Biogeography. Princeton University Press. Princeton, N.J.

Hui, F.K.C., Warton, D.I., Forster, S.D., & Dunston, P.K. (2013). To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. *Ecology,* 94, 1913-1919.

Hui, F.K.C. (2016). boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, 7, 744–750.

Jabot, F., Laroche, F., Massol, F., Arthaud, F., Crabot, J., Dubart, M., Blanchet, S., Munoz, F., David, P., & Datry, T. (2020). Assessing metacommunity processes through signatures in spatiotemporal turnover of community composition. *Ecology Letters*, 23, 1330-1339.

König, C., Wuest, R.O., Graham, C.H., Karger, D.N., Sattler, T., Zimmerman, N.E. & Zurell, D. (2021). Scale dependence of joint species distribution models challenges interpretation of biotic interactions. *Journal of Biogeography, in press.*

Lawton, J.H. (1999). Are there general laws in ecology? *Oikos*, 84, 177.

Legendre, P., & De Cáceres, M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters*, 16, 951-963.

Leibold, M.A. (1998). Similarity and coexistence in regional biotas. *Evolutionary Ecology,* 12, 95-110.

Leibold, M.A., & Chase J.M. (2018). *Metacommunity ecology*. Princeton University Press, Princeton.

Leibold, M.A., Economo, E.P., & Peres-Neto, P. (2010). Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecology Letters*, 13, 1290–1299.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601–613.

Leibold, M.A., & McPeek, M.A., (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology*, 87, 1399-1410.

Leibold, M.A., & Mikkelson, G.M. (2002). Coherence, species turnover, and boundary clumping: elements of meta-community structure. *Oikos*, 97, 237–250.

Levin, S.A. (1974). Dispersion and population interactions. *American Naturalist*, 108, 207-228.

Levins, R., & Culver, D. (1971). Regional coexistence of species and competition between rare species. *Proceedings of the National Academy of Sciences, U.S.A.*, 68, 1246-1248.

Lindstrom, E.S., & Langenheder, S., (2012). Local and regional factors influencing bacterial community assembly. *Environmental Microbiology Reports*, 4, 1-9.

Logue, J.B., Mouquet, N., Peter, H., & Hillebrand, H. (2011). Empirical approaches to metacommunities: a review and comparison with theory. *Trends in Ecology & Evolution*, 26, 482–491.

Miller, E.T., & Bohannan, B. (2019). Life between patches: Incorporating microbiome biology alters the predictions of metacommunity models. *Frontiers in Ecology and Evolution,* 7, 276.

Mouquet, N., Gravel, D., Massol, F., & Calcagno V. (2012). Extending the conept of keystone species to communities and ecosystems. *Ecology Letters*, 16, 1-8.

Münkemüller, T., de Bello, F., Meynard, C. N., Gravel, D., Lavergne, S., Mouillot, D., Mouquet, N., & Thuiller, W. (2012). From diversity indices to community assembly processes: a test with simulated data. *Ecography*, 35, 468–480.

Ohlman, M., Mazel, F., Chalmandrier L., Bec, S., Coissac, E., Gielly, L., Pansu, J., Schilling, V., Taberlet, P., Zinger, L, Chave, J., & Thuiller, W. (2018). Mapping the imprint of biotic interactions on b-diversity. *Ecology Letters,* 21, 1660-1660.

Ovaskainen, O., & Abrego, N., (2020). Joint Species Distribution Modelling With Applications in R. Cambridge University Press.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., *et al.* (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576.

Ovaskainen, O., Rybicki, J., & Abrego, N. (2019). What can observational data reveal about metacommunity processes? *Ecography*, 42, 177-1886.

Overcast, I., Ruffley, M., Rosindell, J., Harmon, L., Borges, P.A.V., Emerson, B.C., Etienne, R.S., Gillespie, R., Krehenwinke, H., Mahler, D.L., Massol, F., Parent, C.E., Patino, J., Peter, B., Week, B., Wagner, C., Hickerson, M.J., & Rominger A. (2020). A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. bioRxiv 10.1101/2020.01.30.927236.

Pandit, S.N., Kolasa, J., & Cottenie, K. (2009). Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology,* 90, 2253–2262.

Poggiato, G., Munkenuller, T., Bruystova, D., Arbel, J., Clark, J.S., & Thuiller, W. (2021). On the interpretation of Joint Modelling in community ecology. *Trends in Ecology and Evolution.* (in press).

Peres-Neto, P.R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, 87, 2614–2625.

Peres-Neto, P.R., Leibold, M.A., & Dray, S. (2012). Assessing the effects of spatial contingency and environmental filtering on metacommunity phylogenetics. *Ecology*, 93, S14–S30.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O’Hara, R.B., Parris, K.M., et al. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5, 397–406.

Resetarits, E.J., Cathey, S.E., & Leibold M.A. (2017). Testing the keystone community concept: effects of landscape, patch removal, and environment on metacommunity structure. *Ecology,* 99, 57-67.

Ricklefs, R.E. (1987). Community diversity: Relative roles of local and regional processes. *Science*, 235, 167–171.

Rybicki, J., Abrego, N., & Ovaskainen, O., (2019). Habitat fragmentation and species diversity in competitive communities. *Ecology Letters*, 23, 506-517.

Schiesari, L., Matias, M.G., Prado, P.I., Leibold M.A., Albert, C.H., Howeth, J.G., Leroux S.H., Pardini, R., Siqueira, T., Brancalion, P.H.S., Cabeza, M., Coutinho, R.M. Diniz-Filho, J.A.D.F., Fournier, B. Lahr, D.J.G., Lewinsohn, T.M., Martins, A., Morsello, C., Peres-Neto, P.R., Pillar, V.D., & Vazques, D.P. (2019). *Perspectives in Ecology and Conservation*, 17, 172-181.

Shoemaker, L.G., & Melbourne B.A. (2016). Linking metacommunity paradigms to spatial coexistence mechanisms. *Ecology,* 97, 2436–2446.

Shoemaker, L.G., Sullivan, L.L., Donohue, I., Cabral, J.S., Williams, R.J. Mayfield M.M., Chase, J.M., Hcu, C. Harpole, W.S., Huth, A., HilleRisLambers, J., James, A.R.M., Kraft, N.J.B., May, F., Muthukrishnan, R., Satterlee, S., Taubert, F., Wang, X., Wiegand, T., Yang,Q., & Abbott, K.S. (2020). Integrating the underlying structure of stochasticity into community ecology. *Ecology*, 101, e02922.

Sloan-Wilson, D. (1992). Complex Interactions in Metacommunities, with Implications for Biodiversity and Higher Levels of Selection. *Ecology,* 73. 1984-2000.

Soberon, J., & Peterson, A.T. (2005). Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics*, 2, 1-10.

Suzuki, Y., & Economo, E.P. (2021). From species sorting to mass effects: spatial network structure mediates the shift between metacommunity archetypes. *Ecography,* 44, 1-12.

Thompson, P.L., Guzman, L.M., De Meester, L., Horvath, Ptacnik, R., Vanschoenwinkle, B., Viana, D.S., & Chase, J.M. (2019). A process-based framework for metacommunity ecology. *Ecology Letters*, 9, 1314-1329.

Turner MG. (2005). Landscape ecology: what is the state of the science? *Annual Reviews of Ecology, Evolution and Systematics,* 36*,* 319–44.

Urban, M.C., Strauss, S.Y., Pelletier, F., Palkovacs, E.P., Leibold M.A., Hendry A.P., De Meester, L., Carlson, S.M., Angert, A.L., & Giery, S.T. (2020). Evolutionary origins for ecological patterns. *Proceedings of the National Academy of Sciences, U.S.A.*, in press.

Vellend, M. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology*, 85, 183–206.

Vellend, M. (2016). The theory of ecological communities. Princeton University Press.

Viana, D.S., & Chase, J.M. (2019). Spatial scale modulates the inference of metacommunity assembly processes. *Ecology,* 100, e02576.

Warton, D. I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. and Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution* 30, 766–779.

Weiher, E., Freund, D, Bunton, T.,Stefanski, A., Lee, T., & Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society: B Biological Sciences*, 366, 2403-2413.

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R., & McCarthy, M.A. (2020). Defining and evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution*, 12, 494-404.

Wiegand, T., Uriarte, M., Kraft, N.J.B., Shen, G., Wang, X., & He, F. (2017). Spatially explicit metrics of species diversity, functional diversity, and phylogenetic diversity: Insights into plant community assembly processes. *Annual Review of Ecology, Evolution, and Systematics*, 48, 329–351.

Yang, X, Tan, J., Sun, K.H., & Jiang, L. Experimental demonstration of the importance of keystone communities for maintaining metacommunity biodiversity and ecosystem functioning. *Oecologia*, 193, 437-447.

Table 1: Summary of metacommunity level variation components for the seven different scenarios modeled in this study.

		Fractions			
Scenario	Corresponding Figure	E (SD)	S (SD)	C (SD)	Residuals $1\text{-}R^2$ (SD)
A	2, upper panels	0.75 (0.11)	0.026 (0.019)	0.044 (0.062)	0.18 (0.11)
B	2, lower panels	0.019 (0.015)	0.15 (0.037)	0.019 (0.036)	0.81 (0.04)
C	3, upper panels	0.42 (0.31)	0.21 (0.2)	0.14 (0.16)	0.23 (0.12)
D	3, lower panels	0.018 (0.026)	0.35 (0.3)	0.28 (0.33)	0.35 (0.33)
E	Supplement	0.63 (0.21)	0.067 (0.069)	0.13 (0.15)	0.18 (0.089)
F	Supplement	0.74 (0.097)	0.035 (0.025)	0.047 (0.057)	0.18 (0.097)
G	4	0.5 (0.17)	0.08 (0.072)	0.2 (0.16)	0.22 (0.12)

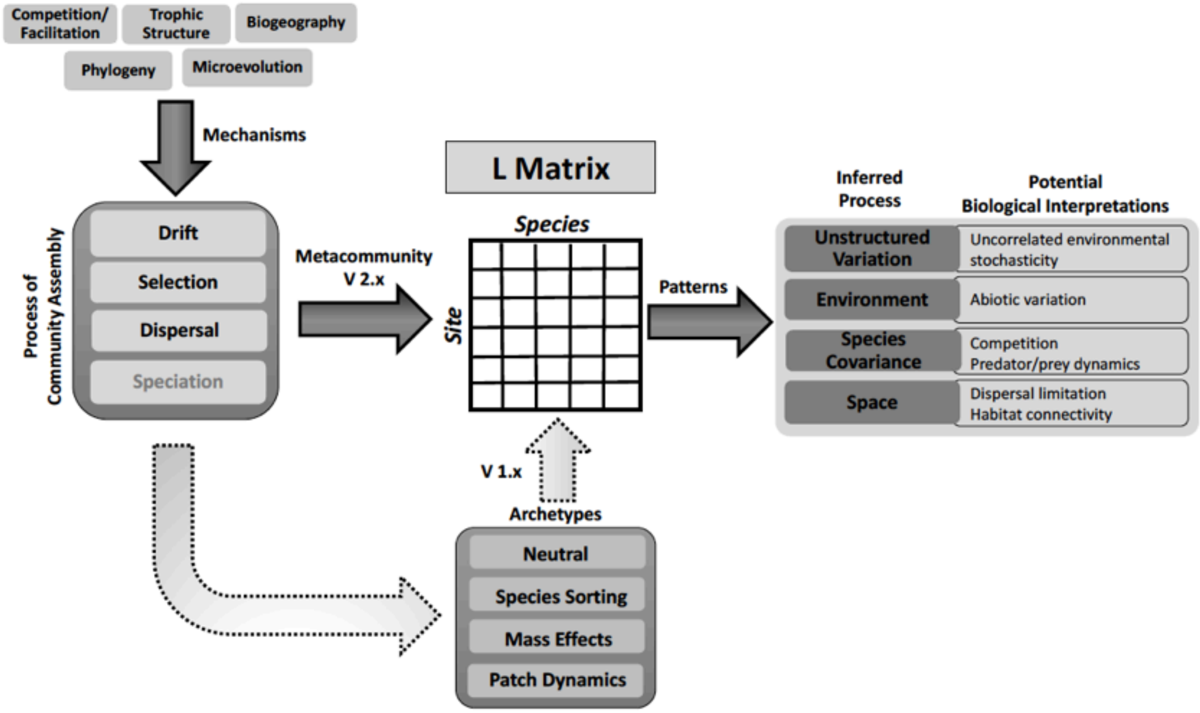


Figure 1: A summary of the metacommunity problem. Species distributions, denoted by the species-by-sites **L** matrix, are the outcome of drift, selection, dispersal, and speciation. These basic processes can be influenced by species interactions, food web structure, biogeography, phylogeny and micro-evolution. Metacommunity theory mainly focuses on drift, selection and dispersal. We view previous approaches based on the four archetypes of Leibold et al. (2004) as being much more indirect and idealized. Instead, we call for a more direct evaluation of how the basic processes affect the **L** matrix, and how to dissect the consequences to the distributions of different species and the occupancy of different sites, for example by using a JSDM to identify main effects and interspecific variability in the importance of unstructured, biotic, environmental, and spatial effects on **L**. This approach allows us to recognize and address the effects of heterogeneities among species and among patches on the overall structure of the metacommunity.

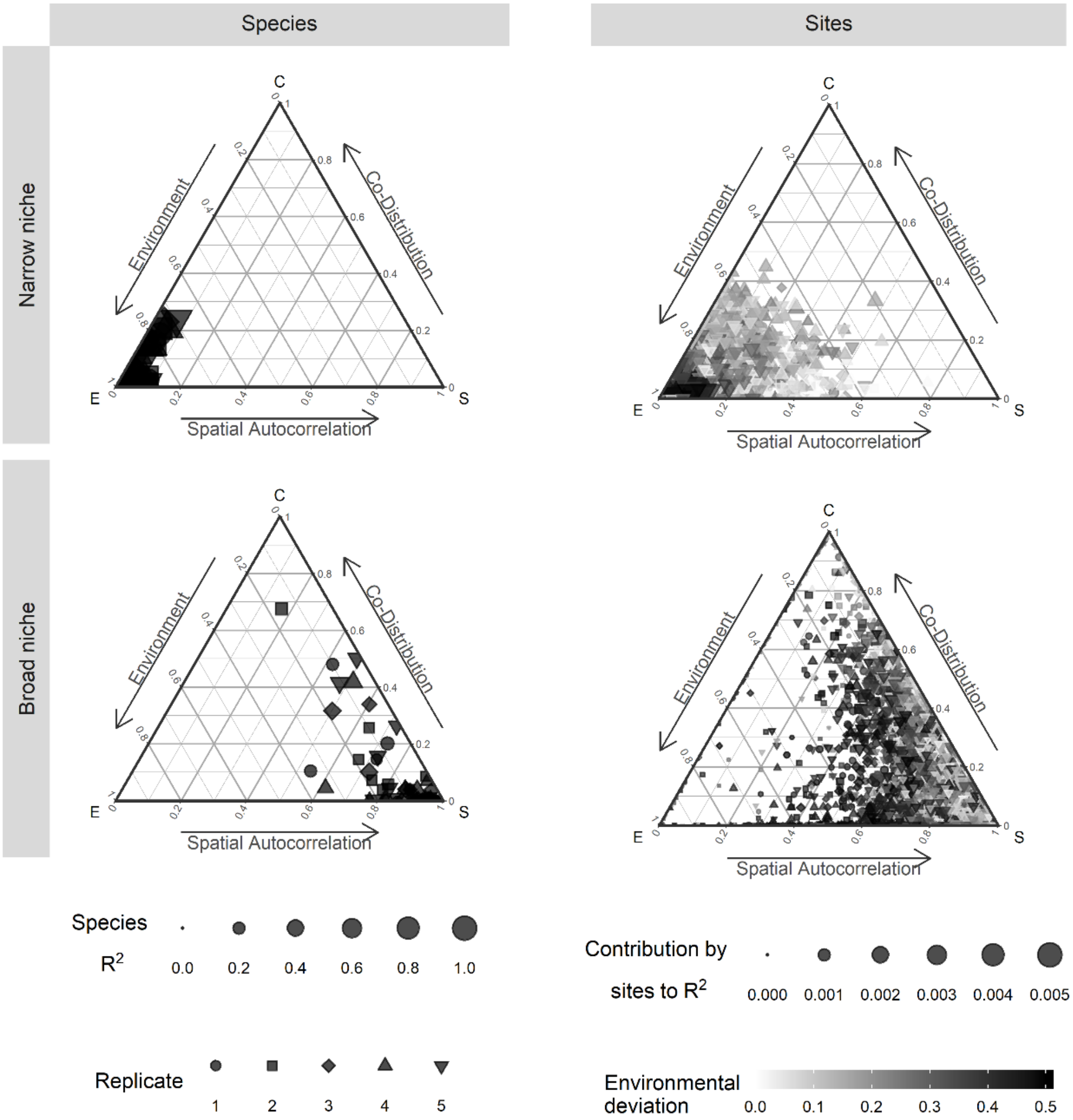


Figure 2: Ternary plots describing the three components of metacommunity internal structure for two different simulation scenarios with no species interactions (independent metapopulations): The upper panels correspond to narrow environmental niches whereas lower panels correspond to wide environmental niches. Each dot represents a species (left panels) or a site (right panels). The size of the symbol is proportional to the R^2 of the model (note the different scales used for species and sites) and the location indicates the proportion of explained variation attributed to environmental factors (*E* - lower left), spatial effects (*S* - lower right) and remaining co-distributions (*C* - upper apex) (see SI for details). In the species panels (left side) different symbols indicate different replicate simulations; generally, these indicate that the distribution of species responses are variable within replicates but that the overall variation among replicates are

repeatable. In the site panels (right side), the shading indicates how central (lighter) or extreme (darker) the local environmental conditions are on the gradient; these also show substantial variation but indicate that more extreme environmental conditions increase the effects of local environment on occupancy patterns than more central conditions.

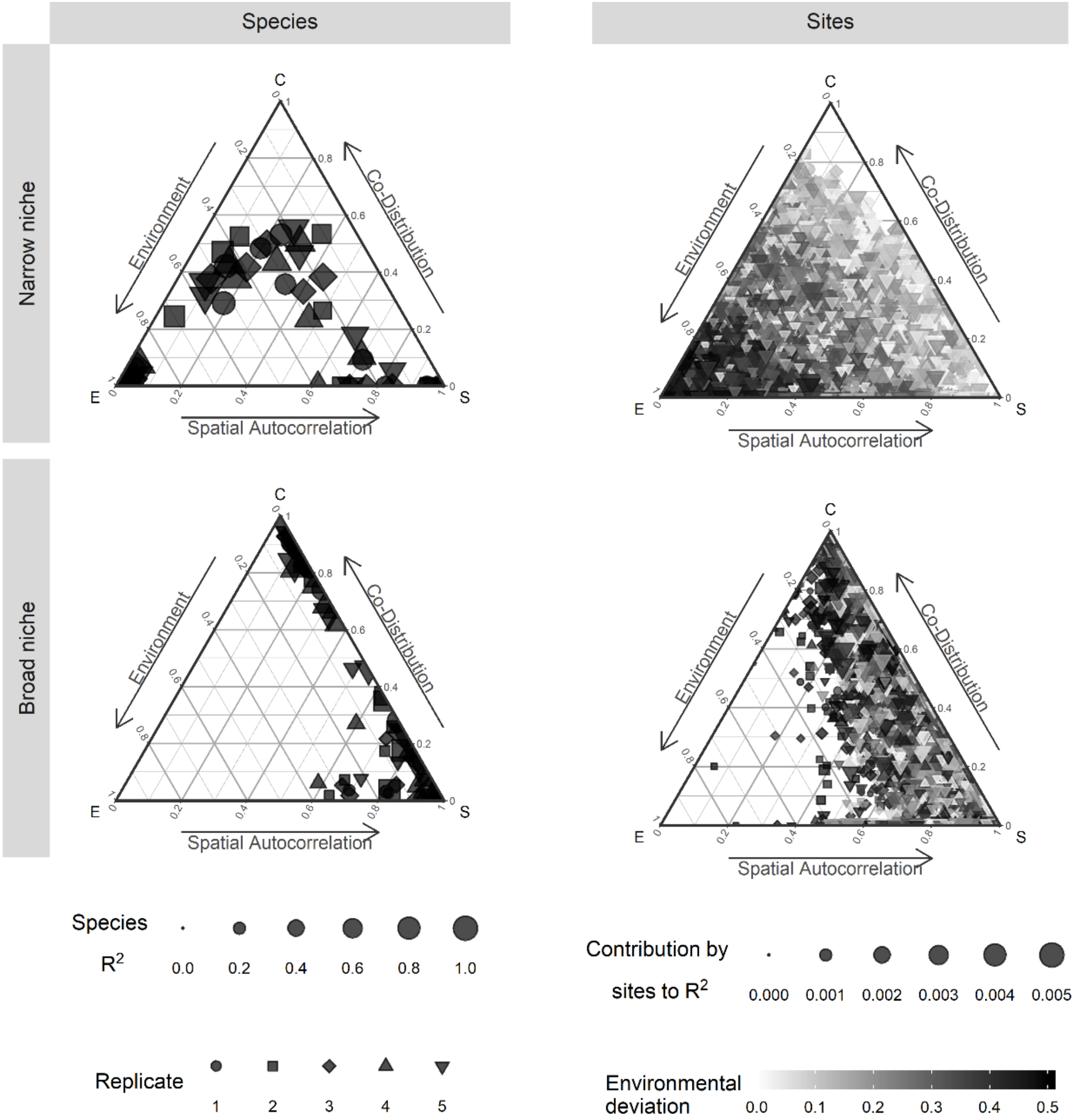


Figure 3: Ternary plots describing the three components of metacommunity internal structure for different simulation scenarios with competition among the species. Notation is the same as in

Figure 2. The upper panels correspond to narrow environmental niches whereas the lower panels correspond to wide environmental niches. Left-hand panels show variation components for different species whereas panels on the right-hand side of the figure correspond to variation components for different sites.

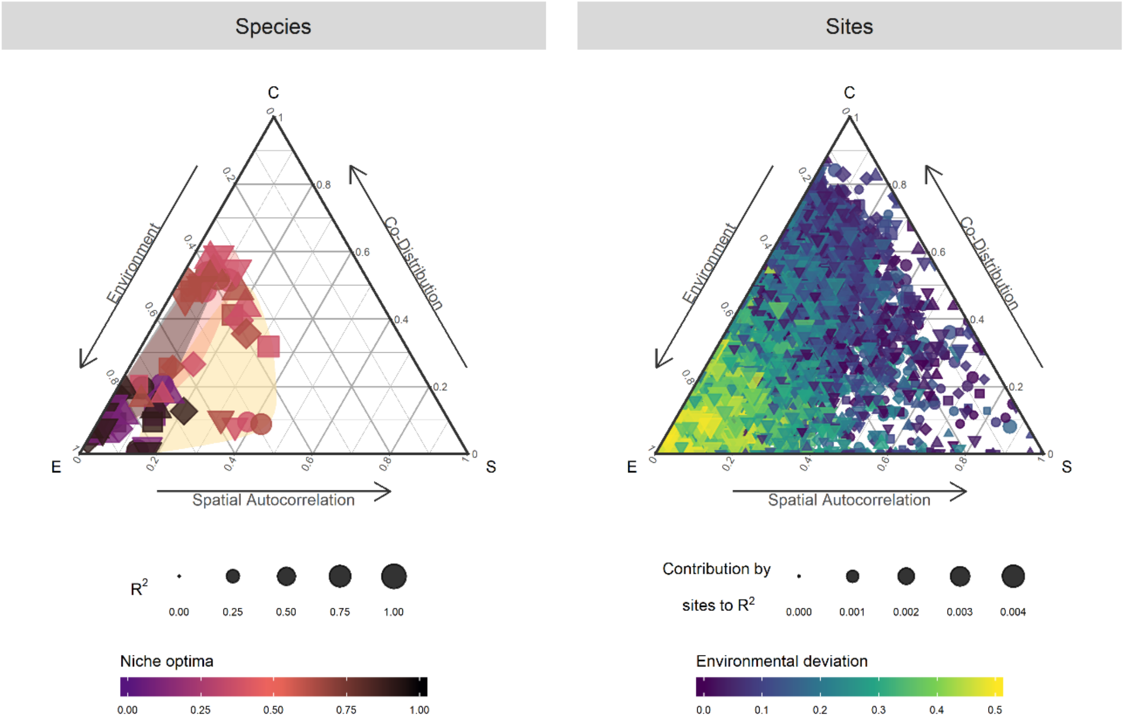


Figure 4: Ternary plots for species (left panel) and sites (right panel) for simulations with species that differ in environmental position along the gradient and dispersal ability. The size of the symbol indicates the R^2 of the model for each species or site). In the left panel (species) the color indicates the preferred local environmental conditions for species (yellow for species that prefer centrally located environmental conditions, purple or magenta for species with more extreme environmental optima). The symbol indicates the dispersal rate of the species (circles are more dispersal limited, squares are least dispersal limited and triangles are intermediate). In the right-hand panel the color indicates the degree of deviation from centrality along the environmental gradient (as in Figure 2).

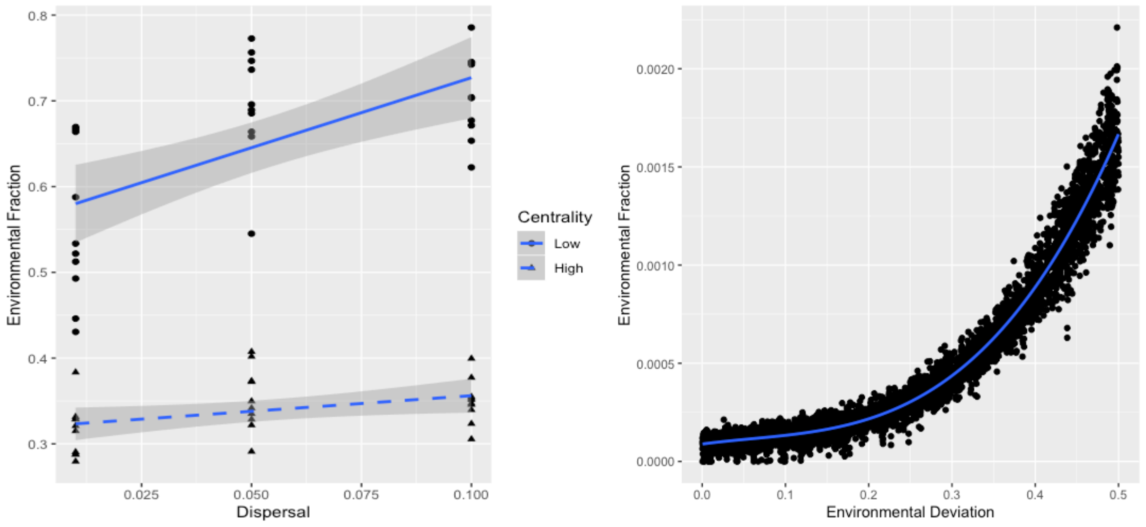


Figure 5: Effects of species traits (i.e. species optima; left panel) and site attributes (right panel) on the environmental fraction of variation in species distributions and site occupancy. A) Higher dispersal ability and lower niche centrality (i.e. greater deviation from mean niche value) enhance the degree to which different species (individual symbols) have distributions that correlate with environmental variation. B) Sites that differ more from the mean environmental value (environmental deviation) are more likely to be occupied by species with niche traits that are locally favored.

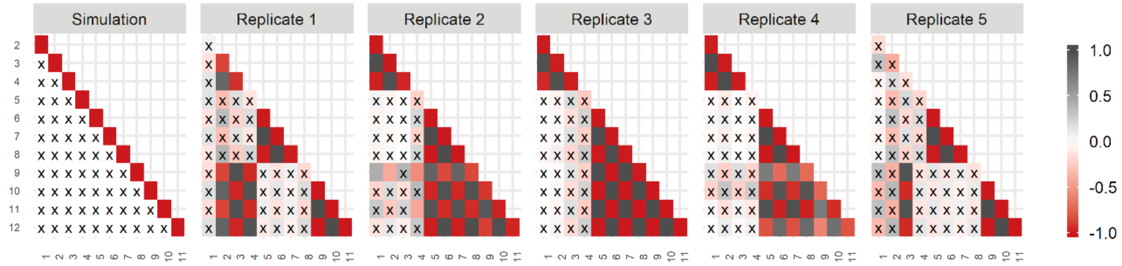


Figure 6: Comparisons of the interaction matrix (Simulation) with the co-distribution of species in five replicate runs (Iteration 1-5) of the scenario with interspecific variation in dispersal and competition among species. In each panel, species are ranked by the position of their environmental optima along the environmental gradient. The co-distributions are shown as heat maps with the strength of the covariation proportional to the intensity of color and the color indicating negative (green) or positive (gray) covariation among pairs of species. These can be compared to the pattern of direct species interactions (left panel called Simulations). The Xs denote no significant association although the color indicates the trend.

The Internal Structure of Metacommunities: Supplementary Information

Mathew A. Leibold, Javiera Rudolph, Pedro Peres-Neto, Dominique Gravel, Luc De Meester, Lauren Showmaker, Florian Hartig, F. Guillaume Blanchet, Jonathan M. Chase

Table of Contents

Description of Model

Patch locations and environmental variables

In our model, the metacommunity consists of N patches distributed over a spatially heterogeneous landscape, with multiple environmental variables (although, the current simulations only have one environmental variable D across 1000 patches) that could either be randomly distributed or spatially autocorrelated. Each patch has a set of coordinates in a two-dimensional space, and all possible coordinates are feasible such that this is a continuous space model that is not restricted to a lattice or some other kind of regular spatial arrangement of spatial units. A patch may be empty or be occupied by a single or by several species. We define $X_{i,z,t}$ as a stochastic variable representing the occurrence of species i at location z and time t. Occurrence, $X_{i,z,t}$, takes a value of 1 when species i is present and a value of 0 when it is absent. Similarly, we define $\mathbf{Y}_{z,t} = X_{1,z,t}, X_{2,z,t}, \dots, X_{R,z,t}$ as a vector containing the presence-absence of each species from the regional pool R.

The model only tracks patch occupancy (not population densities). Spatial dynamics occurs because of colonization events, in both empty patches and patches that are occupied by other species, and because of extinction events. The emerging species co-distributions are a result of a dynamic balance between these events. Ecological interactions can impact either or both the colonization and the extinction probabilities. For instance, the presence of a competitor pre-empting a patch can reduce the colonization probability by another competitor. Alternatively, the presence of a competitor in a patch could increase the extinction probability of another species. Similarly, the environment could influence both the colonization and the extinction probabilities.

Patch Colonization

We consider a discrete-time Markovian process to represent the dynamics of presence-absence of all species and we incorporate the effect of dispersal, environmental filtering and ecological interactions in such a way that we could cover all possible scenarios wherein species differ in any combination of these mechanisms and processes. We can include interspecific competition along with other types of spatial dynamics such as predator-prey interactions (Gravel et al. 2011), priority effects (Shurin et al. 2004), or mutualistic interactions (e.g. Gilarranz et al. 2015). In this paper, we focused on competition only though. Following a colonization event from time t to $t + \Delta$ corresponds to:

$$P(X_{i,z,t+\Delta t} = 1 \,\middle|\, X_{i,z,t} = 0) = I_{i,z,t} S_{i,z,t} C_{i,z,t}$$

where $I_{i,z,t}$ is the number of immigrants of species i reaching patch z at time t, $S_{i,z,t}$ is the effect of environmental filtering on the probability of establishing a viable local population and $C_{i,z,t}$ is the effect of ecological interactions on the establishment probability. We note that because we represent a stochastic process, the product of these three functions has to be bounded between 0 and 1. We consequently define these quantities:

The effect of immigration is given by:

$$I_{i,z,t} = \frac{\sum k(z,\omega) X_{i,\omega,t}}{\sum k(z,\omega)}$$

which is a weighted average of the occurrence probability of species i in the neighborhood of z. The function $k(z,\omega)$ is a dispersal kernel that depends on the location of patch z and the neighborhood ω. For convenience, we considered an exponential function of the Euclidean distance between localities. We added to the kernel a low distance and neighborhood-independent constant m to account from immigration from outside the simulated metacommunity. This assumption is required to prevent total extinction by drift under pure neutral dynamics.

The effect of the environment is given by a product of the establishment performance over all environmental variables E_n:

$$S_{i,z,t} = \prod f(E_{n,z}\, \mu_{i,n}\, \sigma_{i,n})$$

In our simulations, for convenience, we considered that the function f has a quadratic form for all species and all environmental variables, though the model is flexible and general enough to consider other (non-linear) responses that could also differ among species.

Ecological interactions on establishment probability

To incorporate all possible ecological interactions, we started by representing the interaction network by a community matrix $\mathbf{A}$ of R species. The elements α_{ij} of $\mathbf{A}$ quantify the effect of species j on the dynamics of species i. When α_{ij} is negative, the colonization probability of species i decreases and/or its extinction probability increases when j is found locally. Inversely, when α_{ij} is positive, the colonization probability increases and/or the extinction probability decreases. To account for the cumulative effects of local interactions on transition probabilities, we made colonization and extinction probabilities community dependent. As explained above, at a time t, the $\mathbf{Y}_{z,t}$ vector gives the local assemblages. We calculated the sum of interactions at any time and for each species as $v = \mathbf{A}_{z,t}\mathbf{Y}_{z,t}$. Our approach can be interpreted as a spatial analogue to the generalized Lotka–Volterra model because it takes into account the impact of the whole network of interactions on each species dynamics and can deal with any type of interaction. We now define the function:

$$C_{i,z,t} = g(v_i, z, t)$$

representing the total effect of ecological interactions on the colonization probability. For convenience, we will use a sigmoid function, with g ranging between c_{min} at high negative interactions and c_{max} at high positive interactions, where c_{max} should be interpreted as the maximal colonization probability when the environmental conditions are optimal and there are no dispersal limitations.

Patch Extinction

The definition of the extinction probability follows exactly the same rules as for colonization, except that extinction is independent of the neighborhood composition. We follow the same logic to define the effect of ecological interactions and of variation in the environment. Consequently, we get the Markovian process:

$$P\left(X_{i,z,t+\Delta t} = 1 \middle| X_{i,z,t} = 0\right) = M_{i,z,t}E_{i,z,t}$$

where $M_{i,z,t}$ and $E_{i,z,t}$ are the responses of the extinction probability to the local environment and to ecological interactions, respectively. The difference with the colonization functions defined in the previous section is that the extinction probability must be larger when interactions are negative and smaller when they are positive. In addition, the extinction rate should be minimal (instead of maximal) at environmental optimum.

Interpretation

To interpret the model, note that, at steady state, for each species, we obtain the expected occurrence probability ($\hat{P}$) at each site as:

$$\frac{\hat{P}^{iz}}{1-\hat{P}^{iz}} = \frac{I^{iz} \cdot S^{iz} \cdot C^{iz}}{M^{iz} \cdot E^{iz}}$$

After a log transformation, this yields:

$$\log\left(\frac{\hat{P}^{iz}}{1-\hat{P}^{iz}}\right) = \log\left(I^{iz}\right) + \log\left(\frac{S^{iz}}{M^{iz}}\right) + \log\left(\frac{C^{iz}}{E^{iz}}\right)$$

This last equation can be interpreted as a macroscopic description of the expected species distributions pattern (Thuiller et al. 2013). In this formulation, $\log(I)$ describes the tendency of a patch to resemble other nearby patches due to the spatial contagion by dispersal, $\log\left(\frac{S}{M}\right)$ describes the tendency of sites to be occupied by species with similar fitness responses to environmental gradients, and $\log\left(\frac{C}{E}\right)$ describes the remaining influence of other species on co-occurrence due to interactions among species. The values for these indices will depend on what choices are made for the components of eq. 1 (see Supporting Information for details on how we implemented this simulations model).

This modeling framework can represent the classical archetypes but also permits more intricate (and likely far more realistic) metacommunity scenarios and predictions. For example, we could use the model to examine how species traits (and environmental context) link to metacommunity dynamics. Moreover, continuous mixtures of different metacommunity extremes (archetypes) can be represented by appropriate parameter choices for dispersal, competitive abilities, and environmental preferences. For instance, species sorting would require a relatively large colonization to extinction ratio along with species-specific environmental requirements and regional similarity (sensu Mouquet and Loreau, 2002). Alternatively, coexistence within competition-colonization trade-offs requires species to have similar responses to the environment and appropriate heterogeneities in the I, C and E functions, but no environmental preferences.

The implemented mechanisms in the simulation model can be partially mapped onto variation partitioning components. For instance, at equilibrium, we could expect dispersal limitation (the $\log(I)$ term in equation 3) to create positive spatial autocorrelation at the dispersal scale (the [S/E] fraction in variation partitioning, i.e., spatial variation independent of environmental selection). Environmental selection (the $\log\left(\frac{S}{M}\right)$ term in the last equation) should lead to a correlation between composition and environment (the [E/S] fraction in variation partitioning). The last term in equation 3, however, describing the effect of interactions on distribution (the $\log\left(\frac{C}{E}\right)$), is novel and has no equivalent in the context of classical variation partitioning.

There are some interesting properties to point out regarding our proposed variation partitioning scheme. First, by considering the combined effects of environmental selection, dispersal and interactions, the final residuals (unexplained sources of variation) in the model leading to this new partition variation scheme is (in principle) solely related to non-spatialized independent species variation. Second, in our variation partitioning, the interaction component is due to species co-variation (i.e., a joint component among species distributions). In empirical community data, however, this interpretation can only be made if all the environmental variation (predictors) underlying environmental selection in empirical community data has been incorporated (as pointed out in the main manuscript). If not, then the spatial and species interaction components could be measuring variation related to unmeasured environmental variables that are either spatialized (i.e. characterized by the spatial component in variation partitioning) or shared among species (i.e. joint component).

Description of the Statistical Framework

Hierarchical Community Models

In their simplest form, Hierarchical Community Models (HCMs) resemble standard species distribution models that regress species presences/absences against environmental predictors (i.e., logit link). However, to reduce model complexity, HCMs assume that all species in a metacommunity will react to environmental heterogeneity following a similar response function (e.g., linear vs quadratic or Gaussian). The same assumption is made in common variation partitioning (see Peres-Neto et al. 2006). To model the spatial component (i.e., due to spatialized dispersal), either spatial variables such as Moran's eigenvectors maps (MEM, Dray et al. 2006) or spatially auto-correlated latent variables (Ovaskainen et al. 2016b) can be incorporated to the model. To account for biotic interactions, non-spatially auto-correlated latent variables are used. If we use a linear specification approach (here, this can also include quadratic terms that capture Gaussian responses to environment as imposed in our model), we can write:

$$\mathbf{L}^{zi} = \mathbf{X}^{zk}\mathbf{B}^{ki} + \boldsymbol{\epsilon}^{zi}$$

with

$$\mathbf{B}^{ki} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

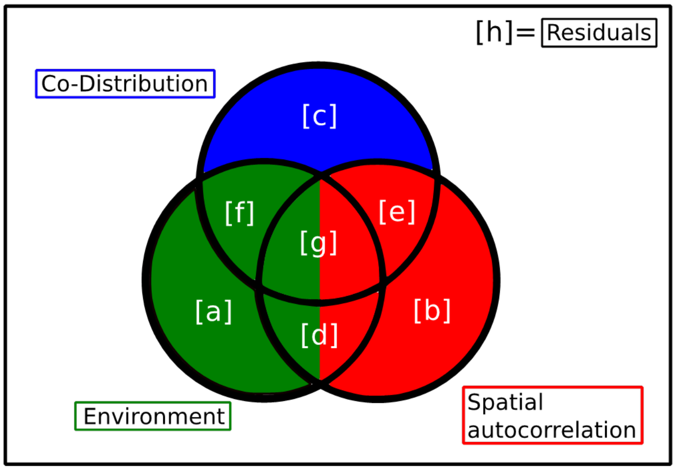
where $\mathbf{L}^{zi}$ is the presence (or absence) of species i (out of m species) at patch z (out of n patches), $\mathbf{X}^{zk}$ is the value of the environmental variable k (out of p variables) at site z, $\mathbf{B}$ is a matrix of regression parameters, $\boldsymbol{\mu}$ is a vector of length p that describes the mean environmental response of all species, Σ is a $p \times p$ covariance matrix that describes how species vary (diagonal) and co-vary (off-diagonal) around the mean environmental response (Ovaskainen and Soininen 2011), and $\boldsymbol{\epsilon}^{zi}$ is a residual value. Estimating species parameters hierarchically around a community mean reduces the degrees of freedom and makes the model easier to fit with limited data. Note that both $\boldsymbol{\mu}$ and Σ can be further informed or constrained by species traits or phylogeny if desired (Ovaskainen et al. 2017). To account for biotic interactions, we consider latent variables $\mathbf{H}^{zl}$ (where l refers to a latent variable measured at site z) and their associated parameters $\boldsymbol{\Lambda}^{li}$ (Ovaskainen et al. 2016a). This yields:

$$\mathbf{L}^{zi} = \mathbf{X}^{zk}\mathbf{B}^{ki} + \mathbf{H}^{zl}\boldsymbol{\Lambda}^{li} + \boldsymbol{\epsilon}^{zi}$$

Note that it is not necessary to always include all of these components in one model; they can be considered in any combination deemed relevant for a particular question. In this paper, we used Moran's Eigenvector Maps (MEMs; Dray et al. 2006), a powerful and commonly used method to model spatial autocorrelation in statistical models involving species distributions.

Calculating Variation Partitioning for the HCM

As in any generalized linear mixed effect model, we can now partition the explained variation into different components, notably environmental heterogeneity, space, co-distribution (biotic interactions), and unexplained variation (Figure S1). To estimate the contributions of each of these four fractions for each species, we calculated semi-partial coefficients of determination (i.e., based on Type III sum-of-squares as specified in Peres-Neto et al. 2006) using the implementation suggested by Tjur (2009) as being more appropriate for presence-absence data (i.e., logit link) than the traditional variation partitioning based on an identity link. To adjust for the number of variables used to quantify each fraction of the variation partitioning analysis, we applied the adjustment to the coefficient of determination proposed by Gelman and Pardoe (2006) in the variation partitioning analysis, which is designed for hierarchical models. As shown in Figure S1, the different fractions were combined so that a unique value was associated to environment (fractions [a], [d]/2, [f] and [g]/2), co-distribution (fraction [c]), space (fractions [b], [d]/2, [e] and [g]/2) and the unexplained portion of the variation (fraction [h]). Latent variables are quite powerful to isolate structure in the data. As such, in the calculation of the variation partitioning, latent variables will capture almost all (if not all) variation associated to the environment and space, giving an artificial inflation of the overlapping partitions between co-distribution and environment and co-distribution and space. For this reason, all partitions overlapping with co-distribution (fractions [e], [f] and [g]) were assigned to either environment (fractions [f] and [g]) or space (factions [e] and [g]). In this calculation, a unique measure of explained variation (akin to adjusted R^2) is associated to co-distribution (fraction [c]) but this is not the case for environment and space. To associate a unique value to environment and space, and represent the results as we did in Figure 2 and 3 (main manuscript), we divided the fractions overlapping environment and space between these two components. As such, the sum of fractions [a], [f], half of fraction [d] and half of fraction [g] were used to measure the effect of the environment while fractions was considered [b], [e], half of fraction [d] and half of fraction [g] were used to measure the effect of space. This scheme in which half of common variation is assigned to two or more common components is commonly used in hierarchical partitioning (Chevan & Sutherland 1991).



.

Supplementary Figure 1 – variation partitioning scheme used to estimate the importance of each matrix of predictors.

Calculation of the coefficient of determination

1) Classic coefficient of determination

The coefficient of determination, R^2, that was partitioned in the variation partitioning analysis (Appendix XX) is calculated for any given species j as:

$$R_j^2 = 1 - \frac{\sum_{i=1}^{n}(y_{ij} - \hat{y}_{ij})^2}{\sum_{i=1}^{n}(y_{ij} - \overline{y}_{ij})^2}$$

where y_{ij} is the data (presence-absence) associated with species j (out of p species) at site i (out of n sites), $\hat{y}_{ij}$ is the model (predicted value) associated to species j at site i and $\overline{y}_j$ is the average of the data (i.e., sum of presences divided by *n*) for species j across all sites.

2) Community-level coefficient of determination

Although having an R_j^2 for each species j can be highly informative and is part of our framework on the internal structure of metacommunities, it can be also useful to estimate the contribution of single communities R^2 to the entire metacommunity. This is obtained by averaging all R_j^2:

$${}^{c}R^2 = \frac{\sum_{j=1}^{p} R_j^2}{p}$$

where the ${}^{C}R^2$ is the community-level R^2.

3) Site contribution to the coefficient of determination

In the paper, we use the contribution of each site to R_j^2 to present how each site contributes differently to the environment, space and co-distribution for the community. The calculation of the site i contribution to the R_j^2, is calculated as:

$$R_{ij}^2 = \frac{1}{n} - \frac{(y_{ij} - \hat{y}_{ij})^2}{\sum_{j=1}^{p}(y_{ij} - \overline{y}_{ij})^2}$$

The first part of the equation where the 1 of the classic R_j^2 is divided by n is included to make sure that if we sum all R_{ij}^2 across all sites for species j, the resulting value equals to R_j^2.

More importantly, what can be noticed is that by calculating R_{ij}^2, the contribution of sites to each species R_j^2, we obtain a matrix that has the same dimension as the site by species matrix (an $n \times p$ matrix). Using this matrix, if we sum across all sites, we obtain the R_j^2.

However, if we average across the species for site i we obtain the site's contribution to the community-level ${}^{C}R^2$ or ${}^{C}R_i^2$.

The amount of variation expressed by the R_j^2, the ${}^{C}R^2$, the R_{ij}^2 or the ${}^{C}R_i^2$ can all be partitioned in its environmental, spatial and co-distribution component following the procedure presented in the section "Calculating Variation Partitioning for the HCM" above.

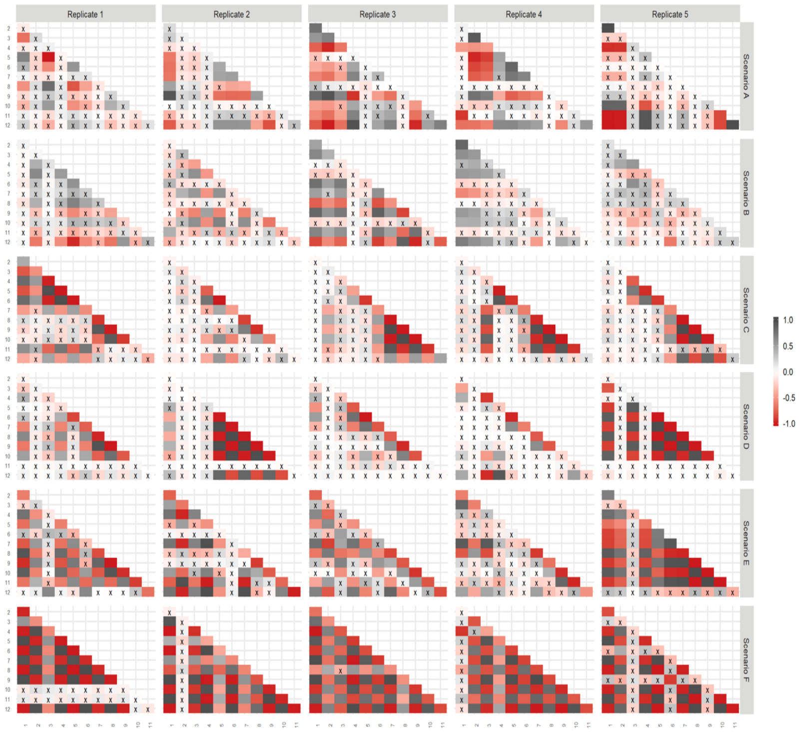
Parameterization and Simulation Scenarios

We simulated metacommunity dynamics with a landscape of 1000 patches over 200 time steps and an initial occupancy of 0.8. Patches were placed randomly in a two-dimensional plane with coordinates drawn from a uniform distribution with a minimum of 0 and a maximum of 1. The environment varied spatially, with values drawn from a random distribution between 0 and 1. In the specific simulations we studied in the paper, colonization was the only component of the species that were affected by the environment (i.e. $E_{i,z,t} = 1$). Specifically, colonization reacted to the environment following a quadratic curve.

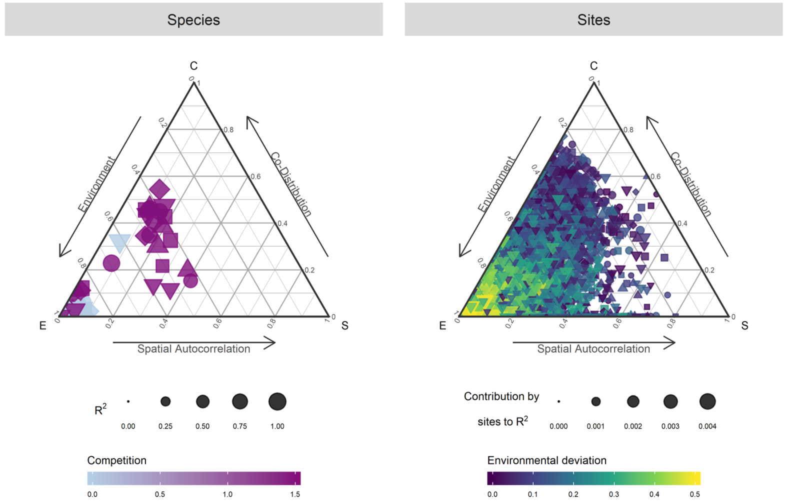
For all scenarios considered, we simulated 12 species. Niche optimums for the species were evenly distributed between 0.1 and 0.9 while niche breadth was set to 0.8 for simulations with narrow niches (scenarios A, B, E, F, G), and to 2 for simulations where niche was assumed to be broad (scenarios C and D). For dispersal, we considered an exponential dispersal kernel, with a distance-independent immigration probability of 0.001 and an α parameter of 0.05. For scenario G, where we have variable dispersal kernels (Figure 3), α was 0.01 for 1/3 of species, 0.05 for 1/3 of species, and 0.1 for the other 1/3 of species. We used a sigmoid function to relate the total number of interactions with colonization and extinction coefficients following the implementation by Cazelles et al. (2016). Colonization probability in the absence of interactions was set at 0.4, which tends to zero as negative interactions tend to infinity, while it asymptotes at a 1 with infinite positive interactions. All other aspects of the colonization-interaction curve were the same for all scenarios. Similarly, extinction in the absence of interactions was set at 0.025, and tended to 1 with infinite negative interactions, while its asymptote tended to 0 with infinite positive interactions. In both cases, the parameter setting the shape of the sigmoid function was set to 0 for the scenarios without competition (scenarios A, B, and F), and 1.5 in the presence of competition (scenarios C, D, E, G) . If there were interactions, then a focal species only interacted with the two species that had the closest niches. For all scenarios, five sets of metacommunities were simulated and analyzed to obtain the results found in Figures 2, 3, and 4 in the main text, and supplementary figures.

All scenarios have been implemented in R and the project's repository can be found here: github.com/javirudolph/testingHMSC

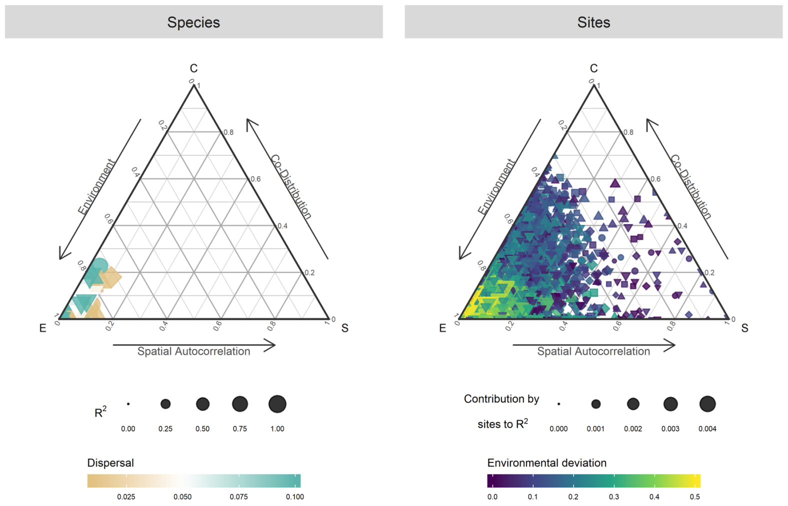
Supplementary Figure 2. Species interactions for all scenarios.



Supplementary Figure 3. In which we have half of the species with interactions and the other without.



Supplementary Figure 4. In which we change dispersal only, α was 0.01 for 1/3 of species, 0.05 for 1/3 of species, and 0.1 for the other 1/3 of species.



Example of metacommunity simulation functions and analyses

The following example of code shows the overall processes involved in the metacommunity simulation for our model. This example shows the scenario for 20 patches and one environmental variable. The model gives the option for a random or spatially aggregated structure for the patches. In the aggregated case, we determined four clusters, denoted by `Nclusters` in the code below. The value of the environmental variable for each patch is shown with the color hue. In this case, the environmental variable is randomly distributed.

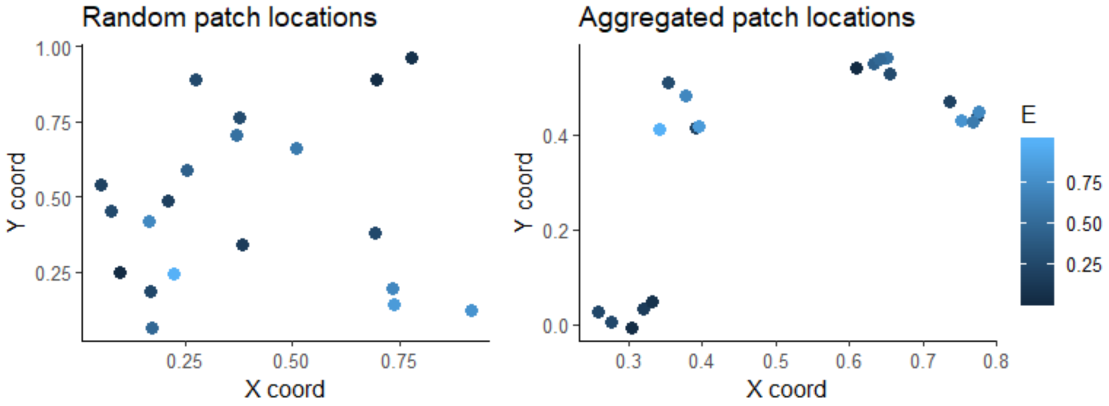
```r
set.seed(227)
# Random XY coordinates
# Each coordinate is drawn from a random uniform distribution
get_XY = function(N) cbind(runif(N),runif(N))

# Aggregation of XY coordinates
get_XY_agg = function(N, Nclusters, sd_xy) {

  Xclust = runif(Nclusters)
  Yclust = runif(Nclusters)

  X = rnorm(N, rep(Xclust,N/Nclusters), sd_xy)
  Y = rnorm(N, rep(Yclust,N/Nclusters), sd_xy)

  cbind(X,Y)
}

# Random uniform environmental values
get_E = function(D, N) matrix(runif(D*N), nr = N, nc = D)
```

Here, we set N, the number of patches to 20, and D, the number of environmental variables to one.

```r
N <- 20
D <- 1
```

```r

rXY <- get_XY(N)
agXY <- get_XY_agg(N, 4, 0.02)
E <- get_E(D = D, N = N)
```

The following figure shows a side by side comparisson between random and aggregated patches obtained from our functions.

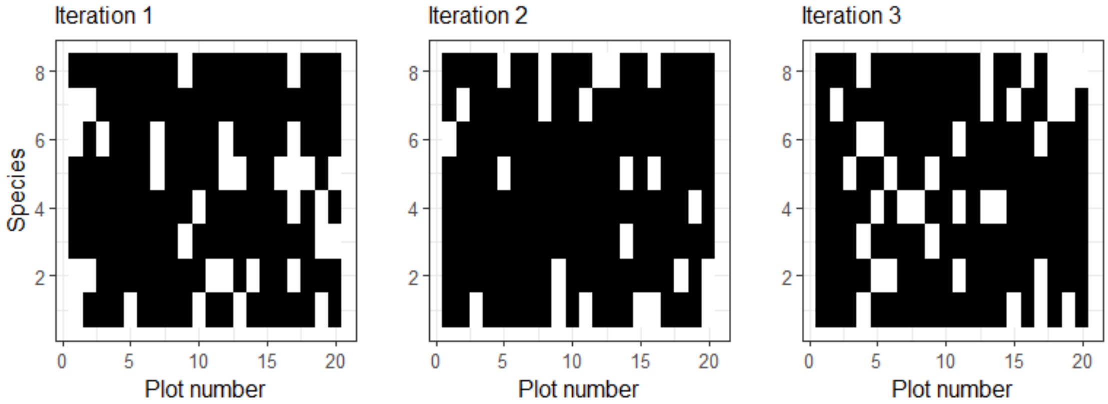


Initial Occupancy

The toy model allows for setitng the initial occupancy in the metacommunity. For example, to create the initial conditions, $t = 0$, of presence absence, species occupancy is drawn from a random uniform distribution, and values smaller than 0.8 are considered as species presence. Patches or locations z are represented by rows in the matrix, whereas each species is a column. Each cell in the matrix is $X_{i,z,t}$ and each row is $\mathbf{Y}_{z,t}$ for $t = 0$. The following figure shows three different iterations of this process, where areas in black represent occupancy = 1, and white denotes an absence.

```r
#Get your initial conditions:
R <- 8
Y0 = matrix(0, nrow = N, ncol = R)
rand = matrix(runif(N*R), nr = N, nc = R)
Y0[rand < 0.8] = 1

ggplot(melt(Y0), aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "black")+
  labs(x = "Plot number", y = "Species", subtitle = "Iteration 1") +
  theme_bw() -> A

# Second iteration for the same initial conditions
Y0 = matrix(0, nrow = N, ncol = R)
rand = matrix(runif(N*R), nr = N, nc = R)
Y0[rand < 0.8] = 1

ggplot(melt(Y0), aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "black")+
  labs(x = "Plot number", y = "" , subtitle = "Iteration 2") +
  theme_bw() -> B
```

```r

# Third iteration
Y0 = matrix(0, nrow = N, ncol = R)
rand = matrix(runif(N*R), nr = N, nc = R)
Y0[rand < 0.8] = 1

ggplot(melt(Y0), aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient(low = "white", high = "black")+
  labs(x = "Plot number", y = "", subtitle = "Iteration 3") +
  theme_bw() -> C

# Set the three iterations to be displayed side by side
ggarrange(A, B, C, ncol = 3, common.legend = TRUE, legend = "none")
```



When we consider the immigration component, we need to also consider the connectivity matrix. In the code below, `I_f` calculates the probabilty of immigration for each species, based on the occupancy matrix and the dispersal kernel, K. The argument `K` is the connectivity matrix. The argument `XY` corresponds to the patch coordinates, whereas `alpha` is the dispersal parameter associated to the exponential distribution used for dispersal. It can be computed with the following function:

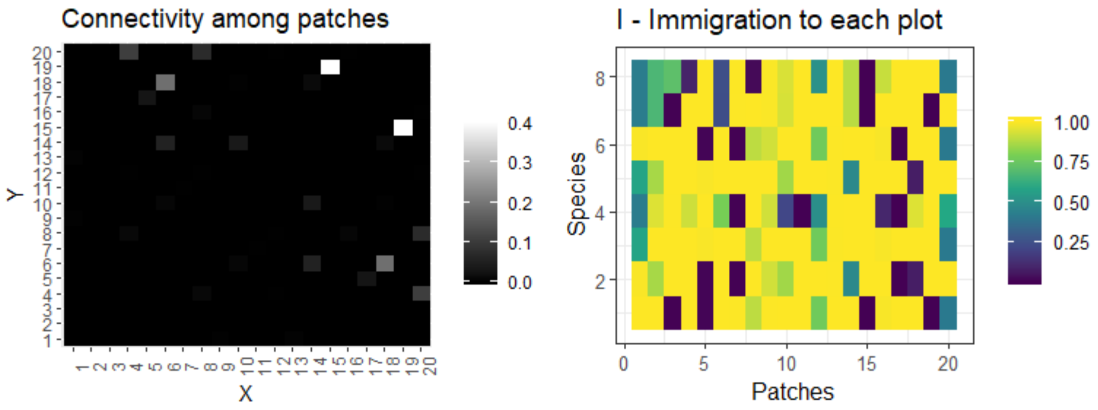
```r
# Compute the propagule pressure
I_f = function(Y, K, m) I = (1-m)*(K%*%Y)/(K%*%matrix(1,nr=N,nc=R)) + m
```

The arguments for this function are: `Y, K, m`. We calculated $\mathbb{Y}$, species presence or absence, in the previous section with the case for initial conditions. Argument `m` is set in the parameters as a value `m = 0.001` and the connectivity matrix `K` is calculated below.

```r
# Compute the connectivity matrix
get_K = function(XY, alpha) {
    N = nrow(XY)
    distMat = as.matrix(dist(XY, method = "euclidean", upper = T, diag = T))
    ConMat = exp(-1/alpha*distMat)
    diag(ConMat) = 0
    return(ConMat)
}
```

As an example, using the aggregated XY coordinates for 20 patches and our initial occupancy matrix with 8 species ,we can see the connectivity between patches, and can calculate the contribution of immigration from each species to each patch.

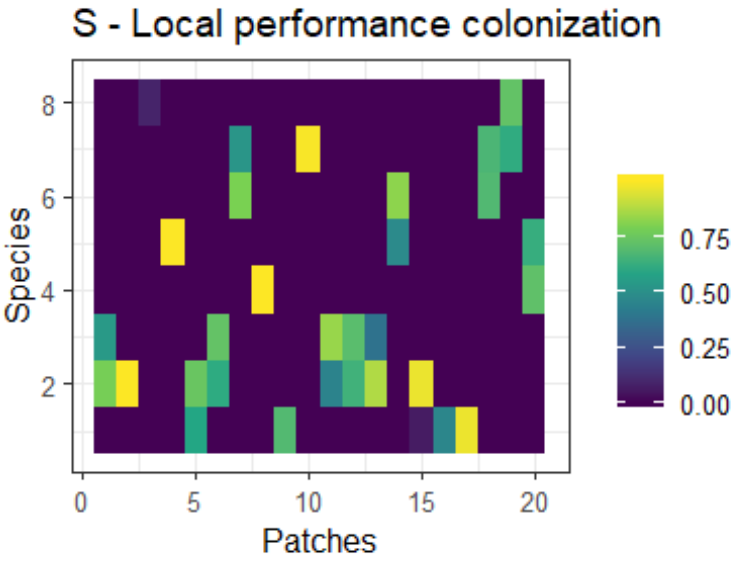
```r
# We can use the aggregated XY coordinates for this example:
XY <- agXY

# Connectivity matrix
alpha <- 0.005
K <- get_K(XY, alpha)
plot_K <- cbind(rownames(K), stack(as.data.frame(K)))
names(plot_K) <- c("X", "Fill", "Y")
plot_K %>%
  as_data_frame() %>%
  mutate(X = factor(X, levels = c(1:20)),
         Y = factor(Y, levels = c(1:20))) -> plot_K
ggplot(plot_K, aes(x = X, y = Y, fill = Fill)) +
  geom_raster() +
  scale_fill_gradient(low = "black", high = "white") +
  #scale_fill_viridis_c() +
  labs(title = "Connectivity among patches") +
  theme(legend.title = element_blank(),
        axis.text.x = element_text(angle = 90)) -> A
# Immigration
m <- 0.001
Y <- Y0
I <- I_f(Y, K, m)
rastPlot(I, title = "I - Immigration to each plot", x = "Patches", y = "Species") +
  theme(legend.title = element_blank()) -> plot_I
ggarrange(A, plot_I, ncol = 2)
```



The effect of the environment on each species, depending on each species niche optima, is computed in the following code section. The argument `E` corresponds to the vector of values for the environmental variable in each patch. The other two arguments in this function are the niche optima (`u_s`) for each species and niche breadth(`s_c`).

```r
# Compute the local performance of propagules
S_f_quadratic <- function(E, u_c, s_c) {
  R <- ncol(u_c)
  N <- nrow(E)
  D <- ncol(E)
  S <- matrix(1, nr = N, nc = R)
```

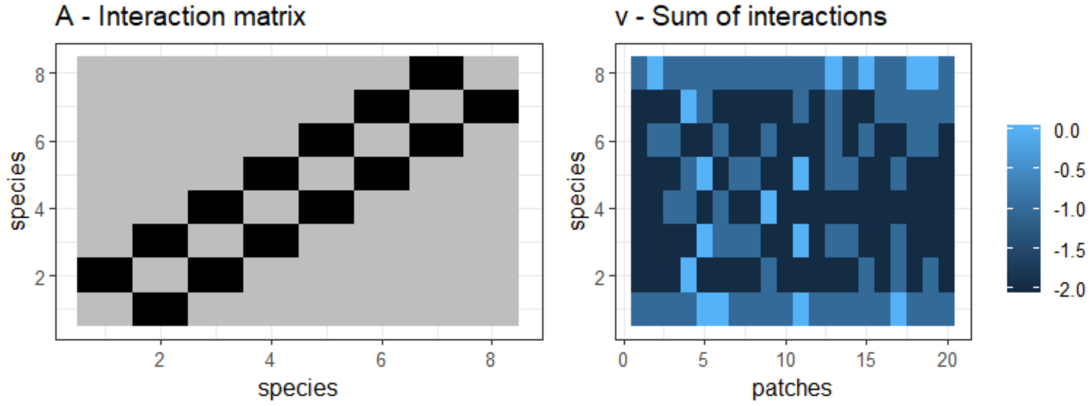
```
  for(i in 1:D){
    optima <- matrix(u_c[i,],nrow = N,ncol = R,byrow = TRUE)
    breadth <- matrix(s_c[i,],nrow = N,ncol = R,byrow = TRUE)
    S <- S * ((-1 / (breadth/2)^2) * (E[,i] - optima)^2 + 1)
    S <- ifelse(S < 0, 0 , S)

  }
  return(S)
}

# Understood as niche optima for each species, for each environmental variable
  u_c = matrix(nr = D, nc = R)
  u_c[1,] = seq(0.1,0.9, length=R)
# Understood as niche breadth
  s_c = matrix(0.2, nr = D, nc = R)

# Local performance, colonization
S <- S_f_quadratic(E, u_c, s_c)
plot_S <- rastPlot(S, title = "S - Local performance colonization", x = "Patches", y = "Specie
s") +
  guides(fill = guide_colorbar(title = ""))
plot_S
```



When incorporating species interactions into the toy model, we use the following interaction matrix **A**, where the colored black sections show species with potential of interacting:

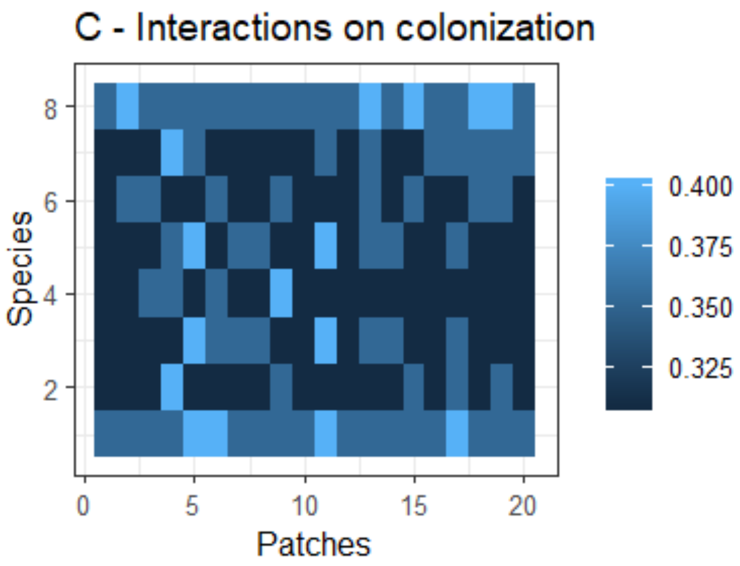
```
  # # Interaction matrix
  A = matrix(0,nr=R,nc=R)
  d = as.matrix(dist(c(1:R),upper=TRUE,diag=T))
  A[d<=1] = -1
  diag(A) = 0

  plot_A <- rastPlot(A, title = "A - Interaction matrix", x = "species", y = "species") +
    scale_fill_gradient(low = "black", high = "grey") +
  theme(legend.position = "none")
```

```r
# Compute the sum of ecological interactions for every location and every species
sum_interactions = function (A, Y) t(A%*%t(Y))
# this is considered to be "v"
v <- sum_interactions(A, Y)

plot_v <- rastPlot(v, title = "v - Sum of interactions", x = "patches", y = "species") +
  guides(fill = guide_colorbar(title = "")) +
  scale_fill_gradient()

ggarrange(plot_A, plot_v, ncol = 2)
```

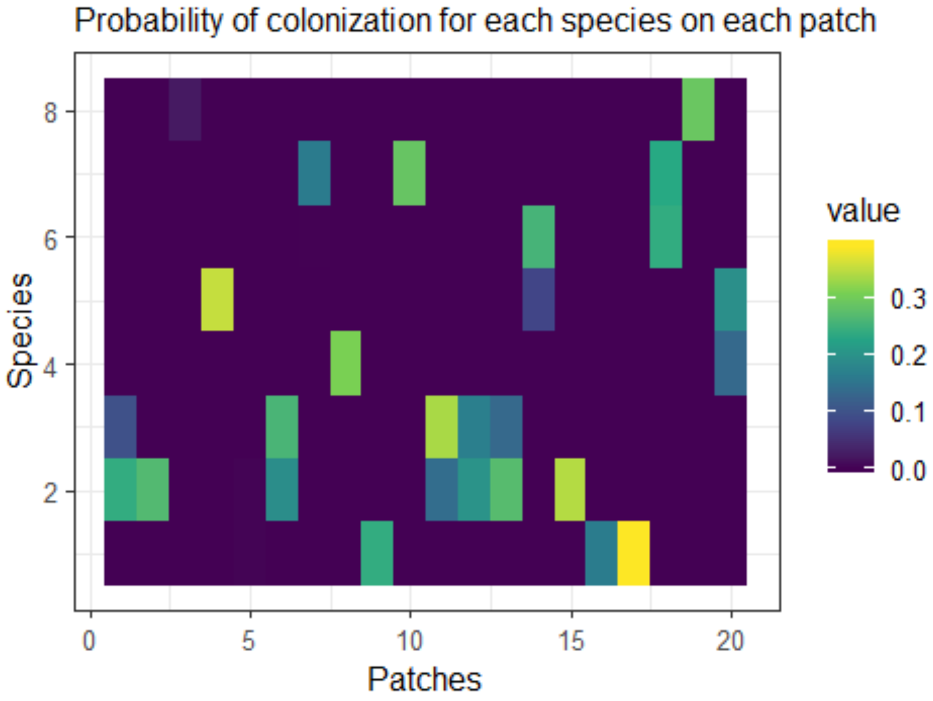


We compute the effect of ecological interactions on colonization probability with the function below.The arguments for this function are `v` as the resulting matrix from the sum of interactions, `d_c` as the sensitivity to interactions, `c_0` and `c_max` as the colonization parameters:

```r
C_f = function(v, d_c, c_0, c_max) c_max*(1 +(1/c_0 - 1)*exp(-v*d_c))^-1

  # # Colonization function
  c_0 = rep(0.4, R) # Colonization at 0 interactions
  c_max = rep(1, R) # Colonization at max interactions

  # # Sensitivity to interactions
  d_c = 0.2

C <- C_f(v, d_c, c_0, c_max)
plot_C <- rastPlot(C, title = "C - Interactions on colonization", x = "Patches", y = "Species
") +
  guides(fill = guide_colorbar(title = "")) +
  scale_fill_gradient()
plot_C
```



With all the components calculated, we can now compute the colonization probability $P\left(X_{i,z,t+\Delta t}=1 \,\middle|\, X_{i,z,t}=0\right)=I_{i,z,t}S_{i,z,t}C_{i,z,t}$



The following function calculates the effect of the environment on the extinction, with `E` being the environmental variable, and `u_e` and `u_s` being species level effect and the assymptote.

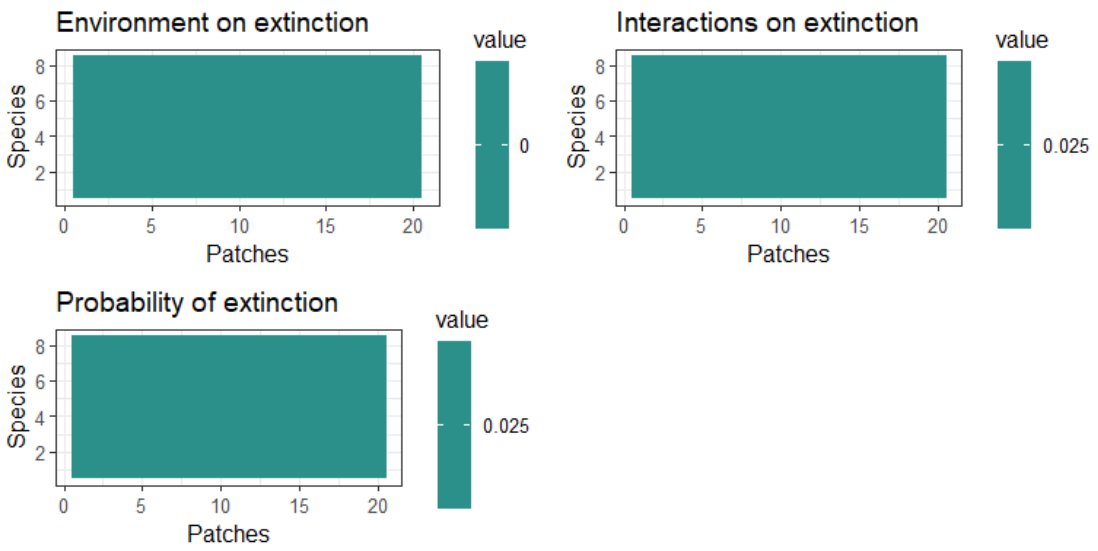
```
M_f = function(E, u_e, s_e) {
    R = ncol(u_e)
    N = nrow(E)
    D = ncol(E)
    M = matrix(1, nr = N, nc = R)
    for(i in 1:D){
      M = M*(1-exp(-(E[,i]-matrix(u_e[i,],nr=N,nc=R,byrow=TRUE))^2 / matrix(s_e[i,],nr=N,nc=R,
byrow=TRUE)^2))
      }
    return(M)
}

# Set the function arguments
  # # Effect of the environment on extinction
  u_e = matrix(nr = D, nc = R)
  u_e[1,] = c(rep(0.5, R-1), 0.05) # One species having a lower level of extinction from envir
onmental effect
  s_e = matrix(Inf, nr = D, nc = R)

  #u_e
  #head(s_e)
```

The following shows the effect of ecological interactions on extinction, using the same `v` matrix calculated above.

```
E_f = function(v, d_e, e_0, e_min) {

    e_min_mat = matrix(e_min, nr = N, nc = R, byrow=TRUE)

    e_min_mat+(1/(1-e_min_mat)+(1/(e_0-e_min_mat)-1/(1-e_min_mat))*exp(d_e*v))^-1

}


#With the arguments computed as:


# # Extinction function
  #e_0 = c(rep(0.025, R-1), 0.5) # Extinction at 0 interactions, with one species having a hig
her value.
e_0 = rep(0.025, R)
e_min = rep(0, R) # Exinction at max interactions

  # # Sensitivity to interactions

  d_e = 0
```

We can now compute the probability of extinction $P\left(X_{i,z,t+\Delta t} = 1 \,\middle|\, X_{i,z,t} = 0\right) = M_{i,z,t}E_{i,z,t}$

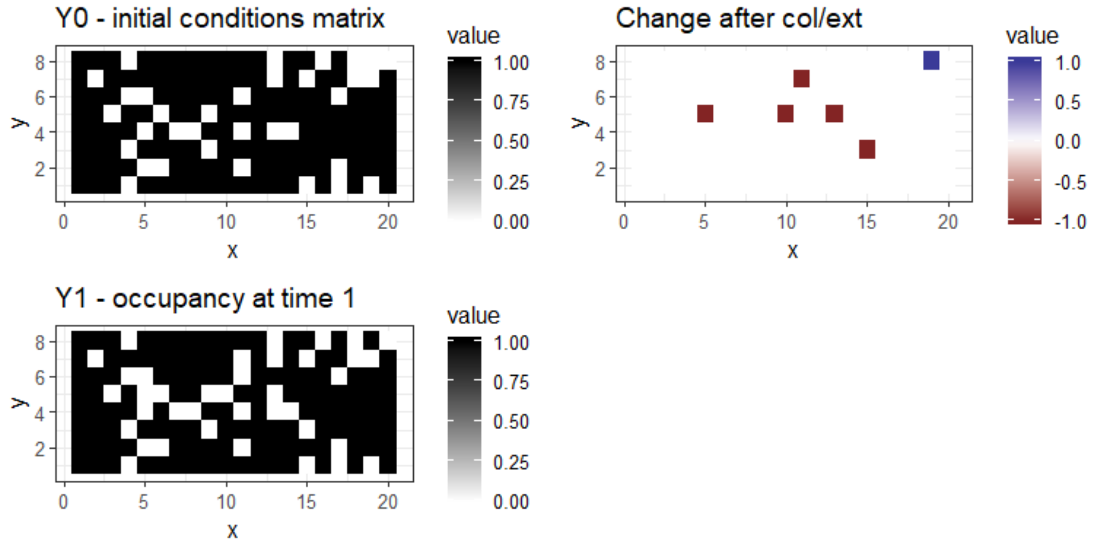
The figure shows these as identical graphs, since we have made all species have the same probability of extinction and the environment not having an effect on extinction.



The way parameters are set, the extinction component is the same for all species in every patch.

Testing and changes

```r
# Perform the test
delta <- matrix(0, nr = N, nc = R)
rand <- matrix(runif(N*R), nr = N, nc = R)
delta[Y == 0 & rand < P_col] <- 1

# Perform the test
rand = matrix(runif(N*R), nr = N, nc = R)
delta[Y == 1 & rand < P_ext] = - 1
```



References:

Chevan, A., & Sutherland, M. (1991). Hierarchical partitioning. *American Statistician*, 45, 90-96.

Dray, S., P. Legendre, & Peres-Neto, P.R. (2006). Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). *Ecological Modelling*, 196, 483-493.

Efron, B. (1978). Regression and ANOVA with zero-one data: Measures of residual Variation. J*ournal of the American Statistical Association*, 73, 113–121.

Gelman, A., & Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. *Technometrics*, 48, 241–251.

Gilarranz, L. J., Sabatino, M., Aizen, M. A., & Bascompte, J. (2015). Hot spots of mutualistic networks. *Journal of Animal Ecology*, 84, 407-413.

Gravel, D., Massol, F., Canard, E., Mouillot, D. & Mouquet, N. (2011). Trophic theory of island biogeography. *Ecology Letters*, 14, 1010–1016.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, 7, 549–555.

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91, 2514–2521.

Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, 7, 428–436.

Ovaskainen, O., & Soininen, J. (2011). Making more out of sparse data: hierarchical modeling of species communities. *Ecology*, 92, 289–295.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., et al. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576.

Peres-Neto, P.R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, 87, 2614–2625.

Shurin, J.B., Amarasekare, P., Chase, J.M., Holt, R.D., Hoopes, M.F. & Leibold, M.A. (2004). Alternative stable states and regional community structure. *Journal of Theoretical Biology*, 227, 359–368.

Tjur, T. (2009). Coefficients of determination in logistic regression models-A new proposal: the coefficient of discrimination. The *American Statistician*, 63, 366–372.