

1 **Improving inference of metacommunity structure using**  
2 **multi-species occupancy models**

3 **Authors** Joseph R. Mihaljevic<sup>1\*</sup>, Maxwell B. Joseph<sup>1</sup>, Pieter T.J. Johnson<sup>1</sup>

4 <sup>1</sup> *Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO,*  
5 *80309*

6 *\* Corresponding Author*

<sup>7</sup> **Abstract**

<sup>8</sup> Blah blah blah

## 9 Introduction

Broadly speaking, the metacommunity concept seeks to understand how spatial patterns of community composition emerge as a product of both local (e.g. competition) and regional (e.g. dispersal) dynamics (Leibold et al. 2004, Holyoak et al. 2005, Chase 2005). To date, metacommunity research can be separated into two nearly distinct categories: mechanism-based approaches and pattern-based approaches. The mechanism-based approach employs mechanistic modeling or controlled experiments to generate and test hypotheses related to, for example, the coexistence of competitors across local sites (Holyoak et al. 2005, Cottenie 2005, Urban and De Meester 2009, Economo 2011, Logue et al. 2011, Pillai et al. 2011, Carrara et al. 2012). The pattern-based approach relates empirical patterns of species occurrences across a landscape to hypothesized biological processes via the exploration of covariate effects on observed structure (Leibold and Mikkelsen 2002, Presley and Willig 2010, Presley et al. 2010, López-González et al. 2012, Henriques-Silva et al. 2013, Meynard et al. 2013, Richgels et al. 2013, Erős et al. 2014). While both approaches are valuable, there is room for synthesis and improvement of these often disparate lines of research.

The pattern-based metacommunity approach, often referred to as the ‘elements of metacommunity structure’ (EMS) paradigm (Leibold and Mikkelsen 2002, Presley et al. 2010), utilizes observed occurrences of species among ‘patches’ of habitat in the metacommunity (e.g. field sampling sites). These data are typically compiled into a site-by-species incidence matrix and statistical analyses determine if the metacommunity exhibits any of twelve categorical structures (Presley et al. 2010). Based on the specific structure observed, inferences are made as to how the metacommunity assembles along dominant environmental axes. However, the efficacy of this pattern-to-process approach depends heavily upon the quality of the observed data (Gotelli and Graves 1996, ???). For instance, problems with species detection could lead to inaccurate incidence matrices and, therefore, inaccurate assessments of metacommunity structure. Fortunately, a decade of advancements in occupancy modeling have led to powerful methods to deal with problems such as species detectability (MacKenzie et al. 2002, Dorazio et al. 2010, Burton et al. 2012, ???). Integrating occupancy modeling with the pattern-based EMS paradigm should improve our ability to assess metacommunity structure and our ability to infer structuring processes.

Occupancy modeling relies on repeated sampling surveys to distinguish between the probability of a species occurring at a site and the probability of a species being detected at a site in which it occurs (MacKenzie et al. 2002, ???). This approach allows for an estimation of ‘true’ occupancy at each sampled site. Community occupancy models also facilitate estimation of environmental covariate effects on the occurrence of each species at each site (Dorazio et al. 2010, Burton et al. 2012). By allowing for better estimates of site-specific occurrence and covariate effects, using occupancy models has the potential to improve our assessment and understanding of metacommunity structure. Indeed, Dorazio and colleagues (2010) specifically identify a closer union of multi-species occupancy modeling with metacommunity theory as a priority in the field.

In this article, we first briefly review current EMS methods and introduce the advantages of community occupancy models, including overcoming detection errors and estimating species-specific covariate effects. We then present a simulated example of how occupancy

modeling could be used to better assess metacommunity structure. Next, to quantitatively support our overall verbal arguments, we use simulation to ask whether, in general, using occupancy models can indeed reduce bias in assigning metacommunity structure, and then to demonstrate how an occupancy model's integration of species-specific covariate effects can inform us of when and where different metacommunity structures might arise. We believe that using occupancy modeling will facilitate a more complete understanding of metacommunities by improving metacommunity structure inference and improving our understanding of these structures' biological underpinnings.

## Elements of metacommunity structure

The EMS paradigm follows a step-wise procedure to determine which of twelve categorical metacommunity structures are exhibited by a data set of species occurrences observed across multiple sites. Although the procedure can determine that no orderly structure exists (i.e. random structure), most metacommunities tested thus far exhibit detectable structure (e.g. Leibold and Mikkelsen 2002). Here we provide an overview of these methods (see also Leibold and Mikkelsen 2002, Presley et al. 2010) and then discuss some potential problems that can arise.

First, species occurrence data are assembled into a site-by-species incidence matrix and this matrix is ordinated, typically using reciprocal averaging. Reciprocal averaging is a type of correspondence analysis that uses an algorithm to generate ordination scores based on the sites' similarities in species composition and the species' similarities in distribution among sites. The original matrix is rearranged (i.e. ordinated) based on the primary ordination axis scores to group similar sites and similarly distributed species. This ordinated matrix, thus, theoretically represents how species assemblages are structured along a dominant environmental axis (i.e. gradient). For example, species may form discrete assemblages that turnover along an elevational gradient. Then, from the ordinated matrix, statistics are calculated to summarize the three elements of metacommunity structure: coherence, turnover, and boundary clumping. These statistics are used to assign one of the twelve categorical metacommunity structures [Leibold and Mikkelsen (2002); Presley et al. (2010); Figure 1].

The first metric of metacommunity structure is coherence, which represents whether the majority of species in the metacommunity respond to the same axis of variation, often assumed to be an environmental gradient. Coherence is the foundation of structure, as without either positive or negative coherence, the community is said to be randomly structured (i.e. species do not structure along a common axis of variation; Figure 1). Coherence is estimated using the number of embedded absences in the ordinated matrix. Embedded absences occur in areas of the matrix where a species is absent at a site in which it would be expected to occur based on the ordination. The observed number of embedded absences are then compared to a null distribution of embedded absences generated from ~1000 simulated matrices.

If a metacommunity exhibits positive coherence, meaning most species respond to the dominant axis of variation, two more metrics are calculated to further describe the metacommunity structure: turnover and boundary clumping. Turnover represents how species composition changes along the theoretical environmental gradient, estimated using the number of species

replacements observed in the ordinated matrix. Negative turnover (significantly fewer replacements than the null) is indicative of nested subsets, while positive turnover represents shifting composition along the gradient. Boundary clumping, estimated with Morisita's index, helps to further distinguish structures by determining if discernible groups of species form along the gradient. For instance, with positive turnover, significant clumping would indicate Clementsian structure, where discrete species groups turnover along the gradient. Additionally, if there is no significant turnover, various quasi-structures are assigned to the metacommunity, depending on the trend observed in turnover and boundary clumping (*sensu* Presley et al. 2010).

After the metacommunity structure is discerned using the three metrics above, researchers typically seek to determine which environmental covariates might be driving the primary axis of variation in the ordinated community. This analysis takes various forms. In most cases, the ordination score of the primary axis is extracted for each sampled site. Then univariate correlation tests are run for each covariate of interest, and subsequently researchers theorize how significantly correlated covariates might be responsible for structuring the metacommunity (e.g. Henriques-Silva et al. 2013, Meynard et al. 2013). Another, related approach is to use, for example, canonical correspondence analysis to relate site compositions to multiple covariates simultaneously (e.g. López-González et al. 2012). A more recent approach is to combine one of these previously discussed analyses with a variance partitioning analysis to evaluate the relative contribution of classes of covariates, such as 'local' and 'spatial' or 'abiotic' and 'biotic' (Henriques-Silva et al. 2013, Dallas 2014).

## Difficulties with the EMS approach

### *A problem with detection error*

Although EMS methods have yielded various insights into the structuring of metacommunities, the EMS approach relies on occurrence data. Empirical occurrence data suffer from imperfect detection, and the EMS paradigm so far has not attempted to account for this problem. Sampling design and effort, low species' abundances and peculiarities in species' ecologies (e.g. cryptic or crepuscular organisms) can all lead to less than perfect detection of species in a given community (MacKenzie et al. 2002). Detection error can in turn influence the ordination of the community incidence matrix, the calculated EMS metrics, and the accuracy of structural inference based on null matrices. For example, if species detection is imperfect, the calculated number of embedded absences from the ordinated matrix may be overestimated, which could lead to type II errors where metacommunities are assigned incorrect structures or are incorrectly assigned random structure. Additionally, imperfect detection influences the form of null matrices, as most methods of null matrix generation utilize the raw data on row and/or column sums (Gotelli 2000, ???).

### *Inferring structuring mechanisms from covariates*

While introducing the EMS methods in their seminal paper, Leibold and Mikkelsen (2002) emphasize that the EMS methods can identify patterns, but cannot necessarily elucidate the processes that lead to pattern. For example, metacommunities that exhibit Gleasonian

and Clementsian structure are assumed to be structured differently based on species-specific responses to a dominant environmental gradient (Clements 1916, Gleason 1926, Gilpin and Diamond 1982). It is often stated that Gleasonian structure arises from idiosyncratic species responses, whereas Clementsian structure arises from groups of species that respond similarly to each other, but differently from other groups of species in the metacommunity. Alternatively, Clementsian structure could arise from negative associations between species pairs or groups that arise along the gradient (Gilpin and Diamond 1982). However, these mechanistic interpretations of structure remain more or less speculative in the EMS paradigm, as EMS methods do not estimate species-specific covariate effects. Rather, EMS methods rely on correlating the ordination scores from a single axis of variation to environmental covariates, which could be leaving out valuable information about community structure.

## A solution: Community occupancy models

The occupancy modeling framework can help ameliorate both issues addressed above by: (1) circumventing problems with detection by disentangling occupancy and detection probabilities of each species, and (2) incorporating species-specific covariate effects, which avoids the use of ordination scores and allows for empirical tests of verbal hypotheses related to structuring mechanisms.

Occupancy models were developed to estimate a species' probability of occurring at a site while correcting for the fact that species may go undetected in a survey (MacKenzie et al. 2002, ???). These models use data from repeat surveys, conducted in a time period during which the true occupancy status of a site is assumed to be constant (i.e. the occurring species are not transient). This allows one to disentangle detection and occurrence probabilities in order to estimate true occupancy at each site and to obtain unbiased estimates of a species' response to environmental covariates. More recently, these models have been extended to multi-species and multi-timepoint (i.e. longitudinal) surveys (Dorazio et al. 2010, Burton et al. 2012). These models incorporate species-, site- and time-specific estimates of detection probability, occurrence probability, and covariate effects. Additionally, with longitudinal surveys, one can estimate probabilities of persistence at a site and colonization of previously unoccupied sites. In occupancy models, parameter estimates are derived by Bayesian statistical techniques, which offers specific advantages to the EMS approach, described below.

Here, we introduce a simple community occupancy model that we will use to: (1) show how occupancy modeling can be used in practice with the EMS approach, (2) explore quantitatively whether an occupancy model can generally reduce the bias in assigning metacommunity structure that is the result of imperfect detection, and (3) explore how the distributions of species-specific covariate effects and site-specific covariate values influence metacommunity structure.

## The occupancy model

We use a community occupancy model with multiple surveys at each site over a single time period (e.g. three surveys over a month). More complex models can be created, as described

above, but we wish to keep our analyses tractable to demonstrate the utility of occupancy models in the EMS paradigm in a general sense.

Let  $z_{i,k}$  represent the true occurrence of species  $i$  at site  $k$ , where  $z_{1,2} = 1$  means that Species 1 is present at Site 2. These  $z_{i,k}$  values can be compiled into a ‘true’ metacommunity occurrence matrix,  $Z$ . True occurrence is estimated as a Bernouli trial with probability,  $\psi_{i,k}$ , the probability of occurrence:

$$z_{i,k} \sim \text{Bern}(\psi_{i,k})$$

We assume that the probability of occurrence,  $\psi_{i,k}$ , is related to a single continuous environmental covariate, though any number of covariates could be used in normal practice:

$$\text{logit}(\psi_{i,k}) = \beta_{i0} + \beta_i x_k$$

, where  $\beta_{i0}$  is the species-specific intercept,  $\beta_i$  is the effect of covariate  $x$  on species  $i$ , and  $x_k$  is the value of covariate  $x$  at site  $k$ .

We assume that multiple surveys are conducted at each site and observations are compiled into a species-by-site occurrence matrix,  $Y$ . For example, if Species 1 is observed in two out of three surveys at Site 2,  $y_{1,2} = 2$ . The number of observed occurrences of each species out of the total number of re-surveys at each site allows the model to estimate species-specific probabilities of detection. So, let  $p_i$  represent the probability of detection of species  $i$ . Although detection probability can be related to covariates similarly to occurrence probability, for simplicity we did not impose any such covariate effects.

The observed occurrences are thus binomially distributed, influenced by both the detection and occurrence probabilities and the number of surveys conducted at each site,  $J_k$ :

$$Y_{i,k} \sim \text{Binom}(J_k, z_{i,k} p_i)$$

Estimating all of these species- and site-specific parameters can be computationally expensive. In order to optimize parameter estimation, we use a heirarchical Bayesian model, which relies on Markov chain Monte Carlo (MCMC) sampling. In the Bayesian framework, a posterior distribution of each true occupancy state,  $z_{i,k}$ , is generated. For each draw from the posterior, the estimated  $z_{i,k}$  values can thus be assembled into a posterior  $Z$  matrix,  $Z_{\text{post}}$ . Each  $Z_{\text{post}}$  thus represents an estimate of the true occupancy of the entire metacommunity,  $Z$ , using the observed occurrences,  $Y$ .

For all of our simulations below, we used the open-source statistical software, *R* [Citation]. For Bayesian analyses we used the program JAGS interfaced with the *R* package, *rjags* [Citations]. For metacommunity analyses, we used the *R* package *metacom* (???), which relies heavily on the *R* package *vegan* (???). All of our code and simulation data is available on GitHub [link](#).

## A simulated example

To show how this occupancy model could be used with EMS methods, we first simulated a metacommunity of 12 species and 75 sampled sites, assuming each site was sampled 3

times during a relevant time period. We believe this represents a well-sampled, speciose metacommunity. We assumed a dominant environmental covariate to which most species responded, in order to achieve coherence. We set  $\beta_{i0} = \text{logit}(0.60)$  for all species, and we assumed that species-specific covariate effects,  $\beta_i$ , followed a normal distribution with mean=0 and standard deviation=1. We assumed site-specific covariate values followed a normal distribution with mean=0 and standard deviation=2. In order to emphasize that detection errors can obscure true metacommunity structure, we assumed the probability of detection,  $\text{logit}(p_i)$ , followed a normal distribution with mean= $\text{logit}(0.50)$  and a standard deviation=0.75. This represents a community in which many species are difficult to detect, a scenario under which an occupancy model might be most useful. Later, we explore how influential  $p_i$  values are on the occupancy model's ability to generally reduce bias in determining metacommunity structure.

For this simulated metacommunity, we thus had a known occurrence matrix,  $Z$ , and an observed occurrence matrix,  $Y$ , which was generated by imposing imperfect detection. We then used EMS methods to determine the categorical metacommunity structure of  $Z$  compared to  $Y$ . Then, we used the Bayesian occupancy model to estimate the known metacommunity occupancy,  $Z$ . We used uninformative priors, ran the model with a 1000 iteration adaptation phase, followed by a 5000 iteration burn-in period. After the burn-in period, we ran the model for 10,000 iterations, thinning the MCMC chains by 10 iterations, for a final sample of 1000  $Z_{\text{post}}$  across 3 MCMC chains. We assessed chain convergence using the potential scale reduction factor,  $\hat{R}$  (??).

In this example, the known metacommunity  $Z$  displayed Clementsian structure, while the observed metacommunity,  $Y$ , displayed random (i.e. no discernible) structure, based on standard EMS methods (Figure 2a-b). This shows that detection error can bias the assignment of metacommunity structure.

Next, for each  $Z_{\text{post}}$  from the Bayesian occupancy model, we calculated metacommunity metrics, assembling psuedo-posterior distributions for each metric (Figure 2c-e). These are 'psuedo'-posteriors because we used frequentist statistics, based on standard null matrix simulations, to calculate the significance of the metacommunity metrics for each  $Z_{\text{post}}$ ; however, these  $Z_{\text{post}}$  were estimated in the Bayesian framework. We thus determined the metacommunity structure of each  $Z_{\text{post}}$  ( $n = 1000$ ).

We found that most (65.3%)  $Z_{\text{post}}$  matched the true metacommunity structure of  $Z$  (Clementsian), rather than the incorrect random metacommunity structure (23.7%) assigned to observed  $Y$  (Figure 2f). An additional 6.7% of  $Z_{\text{post}}$  showed Quasi-Clementsian structure. The model was also able to estimate the species-specific probabilities of detection,  $p_i$ , and species-specific covariate effects,  $\beta_i$  (**Supplement XXX**). Although the occupancy model does not find every  $Z_{\text{post}}$  to match  $Z$ , the model is able to put metacommunity structure into a probabilistic framework (Figure 2f). It is worth noting that following standard EMS methods, in this example the only possible outcome would be that the metacommunity displays random structure, which is incorrect. This example demonstrates that in at least some cases, the occupancy model can provide a more accurate picture of metacommunity structure.



## Does an occupancy model generally reduce bias?

Although an occupancy model is designed to account for detection errors, we wanted to quantitatively explore whether using an occupancy model can indeed reduce the type II error rate in assigning metacommunity structure in cases of imperfect detection, across different  $p_i$  distributions.

To address this question, we simulated 1000 metacommunities, similar to above, again using a global pool of 12 species and 75 sites surveyed 3 times each. We assumed a dominant environmental covariate to which species responded, in order to achieve coherence in most cases. We randomly varied the distribution of covariate effects and covariate values in order to simulate many different metacommunity structures (see *How do different metacommunity structures arise?* for more information).

For each simulated metacommunity ( $n = 1000$ ) we thus had a known occurrence matrix,  $Z$ , and an observed occurrence matrix,  $Y$ . We used EMS methods to determine the categorical metacommunity structure for each  $Z$  and  $Y$ . We used an occupancy model to estimate the posterior distribution of each  $Z$  by drawing 1000 iterations of the posterior estimated occurrences,  $Z_{post}$ , and determining the metacommunity structure for each  $Z_{post}$ . Then, for each metacommunity, we calculated the percent of cases in which the metacommunity structure of estimated  $Z_{post}$  matched the metacommunity structure of known  $Z$ . Thus for each  $Z$  and  $Y$  combination ( $n = 1000$ ), we also determined the metacommunity structure of 1000  $Z_{post}$  (i.e.  $N = 1,001,000$  metacommunity calculations). We conducted the full simulation three times, fixing the mean probability of detection at three values: 90%, 75%, and 50% (e.g.  $\text{logit}(p_i)$  followed a normal distribution with mean= $\text{logit}(0.90)$ , standard deviation=0.75). Due to the computational intensity of these simulations, this work utilized the Janus supercomputer, which is supported by the National Science Foundation (award number CNS-0821794) and the University of Colorado Boulder. The Janus supercomputer is a joint effort of the University of Colorado Boulder, the University of Colorado Denver and the National Center for Atmospheric Research.

In order to determine if the occupancy model reduced bias in estimating metacommunity structure, we compared the percent of cases in which the assigned metacommunity structure of observed occurrence matrix,  $Y$ , deviated from known occurrence matrix,  $Z$  (i.e. a point estimate) to the median of the distribution of percentages of  $Z_{post}$  that matched  $Z$  (e.g. Figure 3).

We found that...

## How do different metacommunity structures arise?

Given that an occupancy model is able to estimate species-specific covariate effects, we wanted to explore verbal hypotheses about how metacommunity structures can arise, in a quantitative framework. We used simulation to ask how the distribution of species-specific covariate effects and the distribution of covariate values observed among sampled sites affect resulting metacommunity structure. To do this we simulated known metacommunities,  $Z$ , of

12 species distributed across 75 sites. We were interested in true metacommunity structures, and therefore did not impose detection errors (i.e. we were not concerned about observed occurrences,  $Y$ ).

Importantly, we found that the range and distribution of site-specific covariate values sampled across the landscape can alter the observed metacommunity structure, independent of the species-specific covariate effects. In other words, the selection of sampling sites in a study could alter the observed metacommunity structure in a way that may reflect statistical phenomena rather than biological ones, complicating inferences about structuring mechanisms. Furthermore, with our chosen set of parameter values, we were overall more likely to observe Clementsian structure than Gleasonian structure, even when species-specific covariate effects were drawn from a broader distribution and were essentially idiosyncratic (Table 1). Additionally, we were able to produce nested structures when we forced covariate values to be only positive (or negative), even though covariate effects did not change from the Gleasonian/Clementsian examples (Table 2).

Our findings above suggest that under some circumstances metacommunity structure might arise from statistical phenomena in addition to or in place of biological processes. Similar to our findings, Locey and White (2013) explored the ‘feasible set’ of species area distributions (SAD) as constrained by total species richness and total abundance of species, showing that the ubiquitous hollow-curve SAD may be the result of a statistical phenomenon. Admittedly, our simulations did not attempt to encompass the entire feasibility space of the occupancy model, as this would require altering many parameters. However, this occupancy modeling simulation framework could be used as a type of sensitivity test to help us understand when structures might be generated by sampling issues versus biological processes.

---

## Additional benefits of a merger (To include or not?)

### *Covariate effects*

Another issue with the EMS strategy that can be ameliorated by a merger with occupancy modeling is the way in which covariate effects are explored. As discussed above, many researchers extract the first ordination axis score for each sampled site and associate these scores to any number of site-specific covariates using multiple univariate correlation tests. This approach is flawed due to inflation of the overall type I error rate caused by conducting multiple tests. Additionally, regardless of the type of statistical test, using the ordination scores from a single axis could be leaving out valuable information, and these ordination scores are extracted from a matrix that is most likely flawed by detection error in the first place. Occupancy modeling helps to remedy these issues by estimating species-specific covariate effects on occurrence probability based on the raw data, not ordination scores. Then, using Bayesian model selection and model averaging, dominant covariates and their effects can be estimated without the need for multiple tests.

The influence of confounding covariate effects on detection and occurrence probabilities can also be explored within the occupancy modeling framework. For example, if a single covariate,

such as vegetation density, has positive effects on species' occupancy but negative effects on species' detectability, how might this influence our ability to detect or accurately assign metacommunity structure? The simulation method described above could be used to explore this question in a future study by altering the correlation between the covariate's effect on each species' detection and occurrence probabilities and observing how metacommunity structure changes.

### *Temporal dynamics*

With longitudinal metacommunity data becoming more and more available, an interesting question that arises is how does metacommunity structure vary across time points and what are the driving mechanisms for these changes. For example, Erős et al. (2014) explored how metacommunity structure of stream fishes changed over time and how this structure was influenced by environmental covariates. Dynamic occupancy modeling has a few complementary features that could aid in a more cohesive approach to studying such questions.

Occupancy modeling could be used to distinguish if changes in metacommunity structure are more related to changing covariate values at sites over time, whether the dominant structuring covariate changes identity over time (e.g. from altitude to pH), or if metacommunity structure changes due to idiosyncratic changes in how species respond to the same covariates over time. Additionally, Dorazio et al. (2010) showed how multi-species occupancy modeling can be used to estimate species- and time-specific covariates on detection, occurrence, colonization and persistence probabilities. Understanding colonization and persistence (i.e. extinction) dynamics could aid in inferring more specific mechanisms for metacommunity structural changes.

## **Discussion**

... Shortcomings... - Still relies on single axis ordination for metacommunity analyses - Sensitivity to priors for metacommunity parameter distributions? (e.g. I assigned some parameters from uniform distributions, but their priors were assumed to be normally distributed)

Our simulations are simple in many regards and future studies could more fully explore the sensitivity of metacommunity structure to various model assumptions. One important extension would be to determine the effects of discrete versus continuous covariates and the resulting patterns when both types of covariates influence occurrence probabilities. Additionally, we held our metacommunity size constant, in terms of the number of species and the number of sites sampled. It is well established that matrix size influences the power to detect patterns using null models (Gotelli and Graves 1996, Gotelli 2000), so understanding the influence of these parameters in occupancy models will be important. For tractability, we unrealistically assumed that base-line occurrence and detection probabilities were fixed among species. Understanding how the distributions of these probabilities (and their species-specific covariance) influences overall metacommunity structure should be a leading question.

## Table and Figure Legends

Figure 1. Flow chart that is followed to determine metacommunity structure. Text in gray boxes represents the metric used to estimate the corresponding element of structure (e.g. embedded absences are used to determine coherence). NS = non-significant based on simulated null matrices.

Figure 2. (a) The ordinated matrix from simulated occurrences and (b) the ordinated heat map generated from 200 ordinated  $Z_{post}$  estimated with a heirarchical Bayesian occupancy model. Species-specific base-line occurrence probabilities and detection probabilities were both drawn from a normal distribution with mean=0.5 and SD=0.75. Species-specific covariate effects were drawn from a uniform distribution with high=1.5 and low=-1.5 (on the logit scale). Site-specific covariate values were randomly assigned from a uniform distribution with high=4 and low=-4.

Figure 3. Posterior distributions of the elements of metacommunity structure for our simulated metacommunity, based on the 200 ordinated  $Z_{post}$  shown in Figure 2. (a) Coherence normalized z-scores comparing the number of embedded absences to 1000 null matrices created in the R package ‘metacom’, using the default null generator method, ‘r1’. (b) Turnover z-scores comparing the number of replacements as in (a). (c) Estimated boundary clumping using Morisita’s index. Based on an  $\alpha = 0.05$ , all coherence measures were significantly positive, all turnover was significantly positive, but only 23/200 Morisita’s Index scores were significantly different than 1.

## References

- Burton, A., M. Sam, C. Balangtaa, and J. Brashares. 2012. Hierarchical Multi-Species Modeling of Carnivore Responses to Hunting, Habitat and Prey in a West African Protected Area. *PloS one* 7.
- Carrara, F., F. Altermatt, I. Rodriguez-Iturbe, and A. Rinaldo. 2012. Dendritic connectivity controls biodiversity patterns in experimental metacommunities. *Proceedings of the National Academy of Sciences of the United States of America* 109:5761–5766.
- Chase, J. M. 2005. Towards a really unified theory for metacommunities. *Functional Ecology* 19:182–186.
- Clements, F. E. 1916. *Plant Succession: An Analysis of the Development of Vegetation*. Page 512. Carnegie Institution of Washington.
- Cottenie, K. 2005. Integrating environmental and spatial processes in ecological community dynamics. *Ecology letters* 8:1175–1182.
- Dallas, T. 2014. *metacom: an R package for the analysis of metacommunity structure*. Ecography.
- Dorazio, R. M., M. Kéry, J. A. Royle, and M. Plattner. 2010. Models for inference in dynamic metacommunity systems. *Ecology* 91:2466–2475.
- Economo, E. P. 2011. Biodiversity conservation in metacommunity networks: linking pattern and persistence. *The American naturalist* 177:E167–80.
- Erős, T., P. Sály, P. Takács, C. L. Higgins, P. Bíró, and D. Schmera. 2014. Quantifying temporal variability in the metacommunity structure of stream fishes: the influence of non-native species and environmental drivers. *Hydrobiologia* 722:31–43.
- Gilpin, M., and J. Diamond. 1982. Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia*:75–84.
- Gleason, H. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* 53:7–26.
- Gotelli, N. J. 2000. Null Model Analysis of Species Co-Occurrence Patterns. *Ecology* 81:2606.
- Gotelli, N. J., and G. Graves. 1996. *Null Models in Ecology*. Smithsonian Institution Press Washington, D.C.
- Henriques-Silva, R., Z. Lindo, and P. R. Peres-Neto. 2013. A community of metacommunities: exploring patterns in species distributions across large geographical areas. *Ecology* 94:627–639.
- Holyoak, M., M. A. Leibold, and R. D. Holt. 2005. *Metacommunities: spatial dynamics and ecological communities*. Page 513. University of Chicago Press.
- Leibold, M. a., M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and a. Gonzalez. 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters* 7:601–613.

422 Leibold, M., and G. Mikkelsen. 2002. Coherence, species turnover, and boundary clumping:  
423 elements of meta-community structure. *Oikos* 97:237–250.

424 Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the  
425 distribution of abundance. *Ecology letters* 16:1177–85.

426 Logue, J. B., N. Mouquet, H. Peter, and H. Hillebrand. 2011. Empirical approaches to  
427 metacommunities: a review and comparison with theory. *Trends in ecology & evolution* 26.

428 López-González, C., S. J. Presley, A. Lozano, R. D. Stevens, and C. L. Higgins. 2012.  
429 Metacommunity analysis of Mexican bats: environmentally mediated structure in an area of  
430 high geographic and environmental complexity. *Journal of Biogeography* 39:177–192.

431 MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm.  
432 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*  
433 83:2248–2255.

434 Meynard, C. N., I. Boulangeat, L. Garraud, N. Mouquet, and W. Thuiller. 2013. Disentangling  
435 the drivers of metacommunity structure across spatial scales:1–12.

436 Pillai, P., A. Gonzalez, and M. Loreau. 2011. Metacommunity theory explains the emergence  
437 of food web complexity. *Proceedings of the National Academy of Sciences of the United*  
438 *States of America* 108:19293–19298.

439 Presley, S. J., and M. R. Willig. 2010. Bat metacommunity structure on Caribbean islands  
440 and the role of endemics. *Global Ecology and Biogeography* 19:185–199.

441 Presley, S. J., C. L. Higgins, and M. R. Willig. 2010. A comprehensive framework for the  
442 evaluation of metacommunity structure. *Oikos* 119:908–917.

443 Richgels, K. L. D., J. T. Hoverman, and P. T. J. Johnson. 2013. Evaluating the role of  
444 regional and local processes in structuring a larval trematode metacommunity of *Helisoma*  
445 *trivolvis*. *Ecography* 36:854–863.

446 Urban, M. C., and L. De Meester. 2009. Community monopolization: local adaptation  
447 enhances priority effects in an evolving metacommunity. *Proceedings. Biological sciences /*  
448 *The Royal Society* 276:4129–4138.