Using multi-species occupancy models to improve the classification and understanding of metacommunity structure

Authors   Joseph R. Mihaljevic[1]*, Maxwell B. Joseph[1], Pieter T.J. Johnson[1]

[1] *Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 80309*

* *Corresponding Author*

8 Abstract

Introduction

Over the last few decades, metacommunity ecology has integrated ecological theory across spatial scales in an effort to better understand local and regional community dynamics. Particularly, metacommunity theory explores how regional processes, such as dispersal, feedback with local species interactions to affect both local and regional patterns of species coexistence (Leibold et al. 2004, Holyoak et al. 2005, Chase 2005). One dominant goal of metacommunity research is to empirically characterize multispecies distributional patterns across space, with the ultimate goal of linking these patterns to underlying biotic and abiotic gradients or processes. In this effort, the elements of metacommunity structure (EMS) framework (Leibold and Mikkelson 2002, Presley et al. 2010) was developed as a set of analytical tools to categorize metacommunity structure. The EMS framework has been used to better understand metacommunity dynamics across a variety of habitats and taxa (Presley and Willig 2010, López-González et al. 2012, Henriques-Silva et al. 2013, Meynard et al. 2013, Richgels et al. 2013, Erős et al. 2014). For instance, EMS were used to show that the structure of South American small-mammal communities is most consistent with trends in historical biogeogrpahy, particularly endemism, rather than current anthropogenic impacts (Sancha et al. 2014). Still, the EMS framework could be further established as a particularly valuable tool in metacommunity ecology by synthesizing its foundations of pattern-detection with recent advancements in the statistical modeling of species' occupancy.

The EMS framework utilizes observed species occurrence data aggregated across 'patches' of habitat in a metacommunity (e.g. field sampling sites), which are compiled into a site-by-species incidence matrix. Three summary statistics (coherence, turnover, and boundary clumping) are derived from this matrix to determine which of six core categorical structures the metacommunity exhibits (Presley et al. 2010). Based on the specific structure observed, inferences can be made as to how the metacommunity assembles, and further statistical analyses can be performed to associate the observed structure with a dominant environmental gradient (e.g. elevation, landcover, etc.). However, the efficacy of this pattern-to-process approach depends heavily upon the quality of the data (Gotelli and Graves 1996, Ulrich and Gotelli 2013). For instance, problems with species detection could lead to incomplete 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, Royle and Dorazio 2008, Dorazio et al. 2010, Burton et al. 2012).

Occupancy modeling was designed to reduce the bias inherent in community data sets that results from imperfect detection, and could therefore lead to more accurate assignment of metacommunity structure and subsequently improve our ability to infer structuring processes. Indeed, Dorazio and colleagues (2010) specifically identify a closer union of multi-species occupancy modeling with metacommunity theory as a priority in the field. Occupancy models rely 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, Royle and Dorazio 2008). This approach allows for an estimation of 'true' occupancy at each sampled site. Multi-species occupancy models also estimate species-specific environmental covariate effects on occurrence probability (Dorazio et al. 2010,

Burton et al. 2012) This means that community-level distribution patterns could be associated with species-specific responses to environmental gradients, an important advancement that is lacking in the EMS framework. Thus, by providing better estimates of occurrence and by allowing estimation of species-specific covariate effects, occupancy modeling should improve our characterization and understanding of metacommunity structure.

Here, we review current EMS methods and introduce a metacommunity framework that integrates multi-species occupancy models, which will allow for more accurate categorization of structure and the estimation of species-specific responses to potential environmental gradients. To help illustrate the efficacy of this approach, we use simulated data to show how occupancy modeling can be used to better assess metacommunity structure. We also quantitatively evaluate the degree to which occupancy models reduce bias in structure classification. We expect that this approach will facilitate a more complete understanding of metacommunities by improving metacommunity structure inference and by revealing how species-specific responses to environmental gradients might scale up to community-level patterns.

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 empirical metacommunities tested thus far exhibit detectable structure (e.g. Leibold and Mikkelson 2002). Here we provide an overview of these methods (see also Leibold and Mikkelson 2002, Presley et al. 2010) and discuss some problems that can arise.

Species occurrence data (or abundance data) are assembled into a site-by-species 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 theoretically represents how species assemblages are structured along a dominant environmental axis (i.e. gradient). For example, Mexican bat species form discrete assemblages that turnover along a humidity gradient (López-González et al. 2012). 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 six core categorical metacommunity structures (Figure 1, Leibold and Mikkelson 2002, Presley et al. 2010).

The first metric of metacommunity structure is coherence, which reflects 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, which 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.

Significantly negative coherence is indicative of a checkerboard pattern, demonstrating negative pair-wise species associations. However, if a metacommunity exhibits positive coherence, 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 distinct 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 characterized using the three metrics above, researchers typically seek to determine which environmental covariate explains 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 against the ordination scores, 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 (CCA) to relate site compositions to multiple covariates simultaneously (e.g. López-González et al. 2012). Some recent studies 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 and Presley 2014). Finally, emerging research demonstrates the utility of combining heirarchical cluster analysis with CCA to determine which combination of sites represent distinct metacommunity compartments, and how environmental covariates might influence the formation of these compartments along a gradient (Sancha et al. 2014).

Difficulties with the EMS approach

A problem with detection error

The EMS approach relies on occurrence data, which often suffer from imperfect detection owing to issues of sampling design and effort, low species' abundances and idiosyncracies in species' ecologies (e.g. cryptic or crepuscular organisms) (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 misclassification errors (analagous 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, Ulrich and Gotelli 2013).

Inferring structuring mechanisms from covariates

Leibold and Mikkelson (2002) emphasized 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 hypothesized 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 speculative in the EMS paradigm given that these methods do not estimate species-specific covariate effects. Rather, EMS methods tend to rely on correlating the ordination scores from a single axis of variation – which could be leaving out valuable information about community structure – to environmental covariates.

Linking EMS and multi-species occupancy models

The occupancy modeling framework can help ameliorate both issues addressed above by: (1) disentangling occupancy and detection probabilities of each species, and (2) estimating species-specific covariate effects, which can complement the use of ordination scores and will allow for empirical tests of 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, Royle and Dorazio 2008). 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 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.

6

Formulating the multi-species occupancy model

We use a multi-species occupancy model with mulitple surveys at each site over a single time period (e.g. three surveys over a month). More complex models can be created, but we wish to keep our analyses tractable to demonstrate the utility of occupancy models in the EMS paradigm with a simple example.

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, $\mathbf{Z}$. True occurrence states arise as Bernoulli random variables with probability, $\psi_{i,k}$, the probability of occurrence:

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

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 practice:

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

, where β_{i0} is the species-specific intercept, β_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, $\mathbf{Y}$. 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 facilitates estimation of species-specific detection probabilities. 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:

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

Here we adopt a Bayesian approach for inference and parameter estimation, utilizing Markov chain Monte Carlo (MCMC) sampling in which we iteratively sample from the posterior distribution of each true occupancy state, $z_{i,k}$. For each draw from the posterior, we then obtain a site-by-species incidence matrix $\mathbf{Z}_{post}$, which consists of the elements $z_{i,k}$.

For all of our simulations below, we used the open-source statistical software, R (Team 2014). For Bayesian analyses we used the open-source software JAGS (Just Another Gibbs Sampler; http://mcmc-jags.sourceforge.net/). For metacommunity analyses, we used the R package *metacom* (Dallas 2014), which relies heavily on the R package *vegan* (Oksanen et al. 2013). All of our code and simulation data is available on GitHub *link*.

Integrating EMS and occupancy models: A simulated example

To illustrate how occupancy modeling can be effectively integrated into EMS methods, we simulated a metacommunity of 12 species and 75 sampled sites, assuming each site was sampled 3 times.

In order to create a coherent metacommunity, we imposed a dominant environmental covariate to which species had variable responses. In this way, the community composition shifts along a gradient of the covariate, analogous to say, an elevational gradient. We set $\beta_{i0} = logit(0.60)$ for all species, and we assumed that species-specific covariate effects, β_i, follwed a normal distribution with mean=0 and standard deviation=1. Site-specific covariate values followed a normal distribution with mean=0 and standard deviation=2. To emphasize how detection errors can obscure true metacommunity structure, we assumed the species-specific probabilities of detection, $logit(p_i)$, followed a normal distribution with mean=$logit(0.5)$ 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.

For this simulated metacommunity, we thus had a known occurrence matrix, $\mathbf{Z}$, and an observed occurrence matrix, $\mathbf{Y}$, and we used EMS methods to characterize metacommunity structure of $\mathbf{Z}$ compared to $\mathbf{Y}$. In this example, the known metacommunity $\mathbf{Z}$ displayed Clementsian structure, while the observed metacommunity, $\mathbf{Y}$, displayed random (i.e. no discernible) structure, based on standard EMS methods (Figure 2). Here, detection error resulted in more embedded absences and a concomitant rearrangement of species and sites in the ordinated matrix. It is worth noting that following standard EMS methods, in this example, the only result is that the metacommunity displays random structure, which we know is incorrect.

Next, we applied the occupancy model to estimate $\mathbf{Z}$, based on the observed data, $\mathbf{Y}$. We used uninformative priors, ran the model with a 1000 iteration adaptive 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 $\mathbf{Z}_{post}$ across 3 MCMC chains. We assessed chain convergence using the potential scale reduction factor, $\hat{R}$ (Gelman 1996).

For 1000 $\mathbf{Z}_{post}$ from the Bayesian occupancy model, we calculated metacommunity metrics, assembling psuedo-posterior distributions for each metric (Figure 3b-d). These are 'psuedo'-posteriors because we used frequentist statistics, based on standard null (i.e. randomized) matrix simulations, to calculate the significance of the metacommunity metrics for each $\mathbf{Z}_{post}$; however, these $\mathbf{Z}_{post}$ were estimated in the Bayesian framework. We thus determined the metacommunity structure of each $\mathbf{Z}_{post}$ ($n = 1000$).

In this example, we found that most (65.3%) $\mathbf{Z}_{post}$ matched the true metacommunity structure of $\mathbf{Z}$ (Clementsian), rather than the incorrect random metacommunity structure (23.7%) associated with observed $\mathbf{Y}$ (Figure 3e). An additional 6.7% of $\mathbf{Z}_{post}$ showed Quasi-Clementsian structure. We also created a new ordinated matrix based on the species- and site-specific probabilities of occurence, estimated as the proportion of occurrences observed across the 1000 $\mathbf{Z}_{post}$ (Figure 3a). Importantly, although the occupancy model does not find every

$\mathbf{Z}_{post}$ to match $\mathbf{Z}$, this integration of methods now puts metacommunity structure into a probabilistic framework.

The model was also able to accurately estimate the species-specific probabilities of detection, p_i, and species-specific covariate effects, β_i (Figure 4). Having these species-specific effect estimates allows us to see how species-level responses can scale up to metacommunity-wide patterns. For instance, the Clementsian structure observed in $\mathbf{Z}$ is likely driven by a few species that have strong responses (either positive or negative) to the dominant covariate. This would preclude them from habitats with more extreme covariate values, leading to a clumped distribution. The occupancy model shows that, in our example, Species B, E, and K have relatively strong responses compared to the rest of the metacommunity members. This matches the pattern observed in the metacommunity ordination, showing that indeed these three species' responses to the underlying gradient tend to drive the Clementsian pattern.

This example shows that detection error can bias the assignment of metacommunity structure and that, at least in this case, the occupancy model can provide a more accurate picture of metacommunity structure. Furthermore, it demonstrates the utility of estimating species-specific covariate effects, which helps us elucidate how species-level responses can influence overall metacommunity structure.

Does an occupancy model generally reduce bias?

Although an occupancy model was useful in the simulated example above, we wanted to quantitatively explore whether occupancy modeling reduces the misclassification rate in assigning metacommunity structure across different p_i distributions. In other words, we wanted to answer the question of whether occupancy models are useful generally or only when detection errors are more extreme.

To address this question, we simulated 1000 unique metacommunities 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 acheive coherence in most cases. In order to simulate different metacommunity structures, we randomly varied the distribution of species-specific covariate effects and covariate values (e.g. by varying the distribution type – normal or uniform – and variablility – standard deviation or range).

For each simulated metacommunity ($n = 1000$) we thus had a known occurrence matrix, $\mathbf{Z}$, and an observed occurrence matrix, $\mathbf{Y}$. We used EMS methods to categorize the metacommunity structure for each $\mathbf{Z}$ and $\mathbf{Y}$. For each metacommunity, we then used an occupancy model to estimate the posterior distribution of $\mathbf{Z}$ by drawing 500 iterations of the posterior estimated occurrences, $\mathbf{Z}_{post}$, and determining the metacommunity structure for each $\mathbf{Z}_{post}$. We conducted this full simulation three times, fixing the mean probability of detection at three values: 0.9, 0.7, or 0.5, with a standard deviation of 0.75 (e.g. $logit(p_i) \sim N(logit(0.9), 0.75)$. 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.

To determine whether the occupancy model reduced bias in estimating metacommunity structure, we compared the percentage of cases in which the assigned metacommunity structure of observed occurrence matrix, $\mathbf{Y}$, deviated from known occurrence matrix, $\mathbf{Z}$ – a point estimate – to the median (and its bootstrapped 95% confidence interval) of the distribution of percentages of times that the structure of $\mathbf{Z}_{post}$ matched that of $\mathbf{Z}$. Effectively, this tested whether the occupancy model method, which corrects for imperfect detection, signficantly reduced the misclassification rate of the 1000 simulated metacommunities compared to the standard EMS approach.

We found that for all three values of mean p_i, the occupancy model significantly reduced bias in characterizing metacommunity structure (Figure 5). Even when the average species-specific detection probability was very high (i.e. mean $p_i = 0.9$), detection error resulted in 94.7% of observed occurrence matrices' ($\mathbf{Y}$) structures matching the true occurrence matrices' ($\mathbf{Z}$) structures, meaning a 5.3% misclassification rate. However, using the occupancy model, the median (and 95% CI) of the proportion of posterior occurrence matrix estimates $\mathbf{Z}_{post}$ whose structure matched the true $\mathbf{Z}$ structure was 99.8% (99.5% - 99.9%), showing a significant improvement and reducing the misclassification rate to $< 1\%$ (Figure 5a). Furthermore, the benefit of using the occupancy model increased as the mean probability of detection in the community decreased (Figure 5b-c). Specifically, for mean $p_i = 0.7$, the structure of $\mathbf{Y}$ matched that of $\mathbf{Z}$ 77.5% of the time, while the median for $\mathbf{Z}_{post}$ matching $\mathbf{Z}$ was 93.4% (91.3% - 95.0%); for mean $p_i = 0.5$, the structure of $\mathbf{Y}$ matched that of $\mathbf{Z}$ 59.3% of the time, while the median for $\mathbf{Z}_{post}$ matching $\mathbf{Z}$ was 81.0% (78.4% - 83.4%).

These results demonstrate that surprisingly, even when there are relatively high detection probabilities among species, incorporating an occupancy model should increase the accuracy in assigning metacommunity structure. Using this approach could be especially important in cases of low average detection, for instance in microbial or symbiont communities (e.g. Mihaljevic 2012), in which species might be particularly cryptic due to small size or be rare due to aggregation patterns among hosts (e.g. negative binomial distribution).

Moving Forward

Here we showed that integrating multi-species occupancy models into the EMS framework can reduce bias in assigning metacommunity structure when there is error in species detection, which is a ubiquitous problem in occurrence data (MacKenzie et al. 2002). In fact, we found that even when detection error is quite low, occupancy modeling still significantly reduces metacommunity structure misclassification. Furthermore, given that an occupancy model estimates species-specific covariate effects, this method can be used to better inform how species-level responses can scale up to affect metacommunity-wide patterns of occurrence. Given these benefits, we recommend that metacommunity ecologists transition to using occupancy models in their assessment of metacommunity structure and in determining the environmental covariates that might lead to structure.

Utilizing an occupancy model will require associated changes to data collection protocols, although these changes should not necessarily be associated with an increase in effort. The model requires repeat surveys over a time period in which it can be safely assumed that

10

community member composition is not changing. This design allows for the estimation of species-specific detection probabilities. The design of repeat surveys will depend on the biology of the system and the available resources. For instance, a researcher could conduct two surveys per week (if appropriate), or multiple observers could conduct indpendent surveys on the same day. Thus collecting data to accomodate an occupancy model does not necessarily have to increase effort and the end result is improved accuracy.

Future extensions to the occupancy modeling framework we present could further improve how we assess the influence of environmental covariates on metacommunity structure. For instance, Jackson et al. (2012) present a maximum likelihood method using multi-level models - for which occupancy models are a specific example - to determine how environmental variation leads to changes in community composition. These authors show that estimating species-specific covariate effects can outperform common methods used to assess the influence of environmental covariates, such as canonical correspondence analysis (CCA) and nonmetric multidimensional scaling (NMDS). The methods presented by Jackson et al. (2012) could easily be integrated into the occupancy modeling framework with the additional benefit of simultaneously estimating (and therefore correcting for) species-specific detection probability. In this way, the same model could simultaneously estimate metacommunity structure and determine the relative influences of environmental drivers.

Occupancy modeling could also be used in theorectical metacommunity studies to help us understand when and where metacommunity structures might arise. Given that an occupancy model is able to estimate species-specific covariate effects, one could explore hypotheses about how metacommunity structures arise in a quantitative framework. For instance simulation could be used to ask how the distribution of species-specific covariate effects and the distribution of covariate values observed among sampled sites affect resulting metacommunity structures. Such studies would stregthen the linkages between pattern-based metacommunity studies and mechanism-based theory.

The benefits of correcting for imperfect detection are increasingly appreciated in the fields of ecology and biogeography (Royle et al. 2012, Fitzpatrick et al. 2013, Iknayan et al. 2014, Lahoz-Monfort et al. 2014). Although integrating occupancy modeling with current methods to assess metacommunity structure perhaps comes at the cost of learning new computational tools, the benefits include enhanced accuracy and the emergence of opportunities to method- ologically improve our study of metacommunities. We believe that such a merger is worth the costs and will greatly benefit the field of metacommunity ecology.

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. The ordinated form of the (a) known occurrence matrix $\mathbf{Z}$, and the (b) corresponding observed occurrence matrix $\mathbf{Y}$ with imposed detection error from our simulated example.

Figure 3. Data on metacommunity structure derived from occupancy model. (a) A heat map ordination of the $\mathbf{Z}_{post}$ matrices, based on the average site- and species-specific probability of occurrence across the 1000 iterations. These values were then used to ordinate the species and sites with detrended correspondence analysis. (b-d) Histograms of the metacommunity metrics for each $\mathbf{Z}_{post}$. Bars highlighted in light gray represent $\mathbf{Z}_{post}$ that show non-signficant coherence (i.e. random structure). Vertical dashed lines in (b-c) show the significance cut-off for the metrics' normalized z-scores at $\alpha = 0.05$. (e) Distribution of 1000 $\mathbf{Z}_{post}$ metacommunity structures estimated by the occupancy model.

Figure 4. Values of species-specific probabilities of detection, p_i, and species-specific covariate effects, β_i, as estimated by the occupancy model. Filled circles mark the medians of the posterior distributions of each parameter, while open circles mark the true (simulated) values. Thicker and thinner lines represent the 68% and 95% credible intervals of the estimates, respectively.

Figure 5. Determining if an occupancy model reduces bias in assigning metacommunity structure. Black vertical lines represent the point estimate of the proportion of 1000 simulations for which the observed metacommunity, $\mathbf{Y}$, structure matched the known metacommunity, $\mathbf{Z}$, structure. Solid and dashed gray vertical lines represent the median and bootstrapped 95% confidence intervals for the proportion of 500 estimated $\mathbf{Z}_{post}$ structures that match the known metacommunity, $\mathbf{Z}$, structure. Parameters: (a) mean $p_i = 0.90$, (b) mean $p_i = 0.70$, (c) mean $p_i = 0.50$.

Figures

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.

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.

Dallas, T. 2014. metacom: an R package for the analysis of metacommunity structure. Ecography.

Dallas, T., and S. J. Presley. 2014. Relative importance of host environment, transmission potential and host phylogeny to the structure of parasite metacommunities. Oikos.

Dorazio, R. M., M. Kéry, J. A. Royle, and M. Plattner. 2010. Models for inference in dynamic metacommunity systems. Ecology 91:2466–2475.

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.

Fitzpatrick, M., N. Gotelli, and A. Ellison. 2013. MaxEnt vs. MaxLike: Empirical comparisons with ant species distributions. Ecosphere 4.

Gelman, A. 1996. Inference and monitoring convergence. Pages 131–143 *in* Markov chain monte carlo in practice.

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 PressWashington, 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.

Iknayan, K. J., M. W. Tingley, B. J. Furnas, and S. R. Beissinger. 2014. Detecting diversity: emerging methods to estimate species diversity. Trends in ecology & evolution 29:97–106.

Jackson, M., M. Turner, S. Pearson, and A. Ives. 2012. Seeing the forest and the trees: multilevel models reveal both species and community patterns. Ecosphere 3:1–16.

Lahoz-Monfort, J. J., G. Guillera-Arroita, and B. a. Wintle. 2014. Imperfect detection impacts the performance of species distribution models. Global Ecology and Biogeography 23:504–515.

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.

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

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

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

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

Mihaljevic, J. R. 2012. Linking metacommunity theory and symbiont evolutionary ecology. Trends in ecology & evolution 27:323–329.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, O'HaraR. B., G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013. vegan: Community Ecology Package.

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

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

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

Royle, J., and R. Dorazio. 2008. Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations, and Communities. Page 464.

Royle, J., R. Chandler, C. Yackulic, and J. D. Nichols. 2012. Likelihood analysis of species occurrence probability from presence only data for modelling species distributions. Methods in Ecology and Evolution 3:545–554.

Sancha, N. U. de la, C. L. Higgins, S. J. Presley, and R. E. Strauss. 2014. Metacommunity structure in a highly fragmented forest: has deforestation in the Atlantic Forest altered historic biogeographic patterns? Diversity and Distributions.

Team, R. C. 2014. R: A language and environment for statistical computing. R Foundation for Statistical ComputingVienna, Austria.

462   Ulrich, W., and N. J. Gotelli. 2013. Pattern detection in null model analysis. Oikos 122:2–18.