CONCEPTS & SYNTHESIS

EMPHASIZING NEW IDEAS TO STIMULATE RESEARCH IN ECOLOGY

Using multispecies occupancy models to improve the characterization and understanding of metacommunity structure

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

Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309 USA

Abstract. Two of the most prominent frameworks to develop in ecology over the past decade are metacommunity ecology, which seeks to characterize multispecies distributions across space, and occupancy modeling, which corrects for imperfect detection in an effort to better understand species occurrence patterns Although their goals are complementary, metacommunity theory and statistical occupancy modeling methods have developed independently. For instance, the elements of metacommunity structure (EMS) framework uses species occurrence data to classify metacommunity structure and link it to underlying environmental gradients. While the efficacy of this approach relies on the quality of the data, few studies have considered how imperfect detection, which is widespread in ecological surveys and the major focus of occupancy modeling, affects the outcome. We introduce a framework that integrates multispecies occupancy models with the current EMS framework, detection error-corrected EMS (DECEMS). This method offers two distinct advantages. First, DECEMS reduces bias in characterizing metacommunity structure by using repeated surveys and occupancy models to disentangle species-specific occupancy and detection probabilities, ultimately bringing metacommunity structure classification into a more probabilistic framework. Second, occupancy modeling allows estimation of species-specific responses to environmental covariates, which will increase our ability to link species-level effects to metacommunity-wide patterns. After reviewing the EMS framework, we introduce a simple multispecies occupancy model and show how DECEMS can work in practice, highlighting that detection error often causes EMS to assign incorrect structures. To emphasize the broader applicability of this approach, we further illustrate that DECEMS can reduce the rate of structure misclassification by more than 20% in some cases, even proving useful when detection error rates are quite low (~10%). Integrating occupancy models and the EMS framework will lead to more accurate descriptions of metacommunity structure and to a greater understanding of the mechanisms by which different structures arise.

Key words: Bayesian inference; Bayesian models; community structure; detection error; metacommunity ecology; occupancy models; simulation.

Introduction

Over the last decade, metacommunity ecology has integrated ecological theory across spatial scales in an effort to better understand local and regional community dynamics. In particular, metacommunity theory explores how regional processes, such as dispersal, combine with local dynamics, such as species' interac-

tions and niche partitioning, to affect species' coexistence across scales (Leibold et al. 2004, Chase 2005, Holyoak et al. 2005). Metacommunity research often seeks to empirically characterize multispecies spatial distribution patterns 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 identify and classify structural patterns in community data sets. 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-

[1] Present address: Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637 USA.
E-mail: joseph.mihaljevic@colorado.edu

Silva et al. 2013, Meynard et al. 2013, Richgels et al. 2013, Erős et al. 2014). For instance, by associating unique community structures with areas of endemism and historical refugia, de la Sancha et al. (2014) showed that the structures of South American Atlantic Forest small-mammal communities were most consistent with trends in historical biogeography rather than current anthropogenic impacts. However, the value of the EMS framework could be substantially enhanced by synthesizing its foundations of pattern detection with recent advancements in the statistical modeling of species occupancy.

The EMS framework uses 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 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 help 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); problems with species detection could lead to incomplete incidence matrices and inaccurate assessments of metacommunity structure (e.g., Olito and Fox 2014, who also corrected for such issues). Fortunately, a decade of advancements in occupancy modeling has led to powerful methods to overcome problems such as imperfect detection (MacKenzie et al. 2002, Royle and Dorazio 2008, Dorazio et al. 2010, Burton et al. 2012).

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. Multispecies 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.

We review current EMS methods and introduce a metacommunity framework that integrates multispecies occupancy models, which we term detection error-corrected EMS (DECEMS). This method aims to allow for more accurate categorization of structure and the estimation of species-specific responses to environmental

gradients. To help illustrate the efficacy of this approach, we introduce a simulated case study, which shows how detection error can bias standard EMS approaches and how DECEMS can effectively overcome this problem. We also quantitatively demonstrate that DECEMS can significantly reduce the effect of imperfect detection on structure misclassification across a wide range of simulated detection error rates. By estimating species-specific covariate effects, DECEMS could enhance mechanistic understanding of metacommunity structure. 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 stepwise procedure to determine which of six core 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). 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 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 order 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 (Fig. 1; Leibold and Mikkelson 2002, Presley et al. 2010). It should be noted that although abundance data can be used to ordinate the matrix, only occurrence is used to calculate the three elements of structure.

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 significant positive or negative coherence,
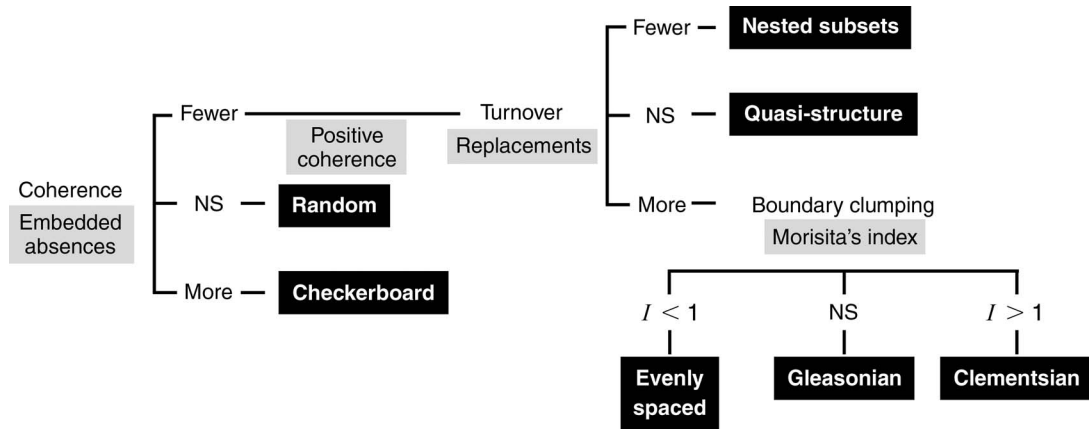
FIG. 1. Flow chart used to determine metacommunity structure. Text in gray boxes represents the metric used to estimate the corresponding element of structure (e.g., embedded absences are used to determine coherence). NS means nonsignificant based on simulated null matrices.

the community is said to be randomly structured (i.e., species do not structure along a common axis of variation; Fig. 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 is then compared to a null distribution of embedded absences generated from ~1000 simulated matrices.

Significantly negative coherence is indicative of a checkerboard pattern, suggesting negative pairwise 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 hypothetical 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 (i.e., a core local assemblage with species subtractions along the gradient), while positive turnover represents more substantial shifting composition (i.e., species additions and subtractions to local communities along the gradient). Boundary clumping, estimated with Morisita's index, helps to further distinguish structures by determining whether distinct clusters of species aggregate along the gradient or whether there is a more gradual, random shift in structure. For instance, with positive turnover, significant clumping would indicate Clementsian structure, where discrete species groups turnover along the gradient, whereas a lack of clumping would be consistent with Gleasonian structure, where species respond idiosyncratically to the hypothetical gradient. A quasi-structure is assigned in recognition that the community exhibits nonrandomness (i.e., significant coherence) but has an indistinguishable pattern of turnover. Classification of these structures relies on the trend towards positive or negative turnover

and the subsequent analysis of boundary clumping (Fig. 1). While plausible mechanisms exist to explain quasi-structures (Presley et al. 2010), it is also possible that these quasi-structures are more often misclassified due to errors in accurately assessing patterns of turnover.

After the metacommunity structure is characterized, 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 correlations are run for each covariate of interest against the ordination scores to explore how covariates might be responsible for structuring the metacommunity (e.g., Henriques-Silva et al. 2013, Meynard et al. 2013), although multivariate models could (and perhaps should) be used here as well. A complementary approach involves using canonical correspondence analysis (CCA) to relate the incidence matrix 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 of the incidence matrix 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 hierarchical 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 (de la Sancha et al. 2014).

Challenges inherent to 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 idiosyncrasies in species' ecologies (e.g., cryptic or crepuscular

organisms; MacKenzie et al. 2002). Detection error results in false absences, which can in turn influence the ordination of the community incidence matrix, the calculated EMS metrics, and the form of null matrices generated from observed occurrences. False absences can alter the ordination score assigned to each species and each site, altering the final form of the ordinated matrix, especially if rare species are often missed in surveys. Additionally, false absences inflate the number of embedded absences, increasing the chance of finding noncoherent or negatively coherent structures. Furthermore, when detection errors occur at the range boundaries of species, which is likely if abundances are lower at the range extremes, this can bias turnover and boundary clumping metrics. Additionally, imperfect detection will incorrectly restrict 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). In turn, these biases could lead to misclassification errors (analogous to type II errors), where metacommunity structures are assigned incorrectly.

Importantly, misclassifying metacommunity structures has implications for biological interpretation. For instance, if one found a truly coherent metacommunity to be noncoherent due to detection errors, one would incorrectly infer that some species in this metacommunity tend to respond to different environmental gradients, instead of one common gradient. Additionally, for example, Gleason's and Clements's views of community assembly are quite different. Thus, incorrectly assigning Gleasonian structure to a Clementsian metacommunity (or vice versa) due to false absences at species' range boundaries could lead to incorrect inferences about species interactions and niche partitioning. However, not all misclassifications are equally egregious. For instance, incorrectly assigning a nested metacommunity as quasi-nested does not drastically affect our biological interpretation of the system. Thus, it is important to understand how detection error affects the likelihood of different types of misclassification errors (Appendix).

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). Gleasonian structure is believed to arise 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, however, Clementsian structure could arise from negative associations between species pairs or groups that arise along the gradient (Gilpin and Diamond 1982). These mechanistic inter-

pretations 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 multispecies occupancy models

The occupancy modeling framework can help ameliorate the issues addressed above by disentangling occupancy and detection probabilities of each species, and 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 repeated 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 multispecies and multi-time-point (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 occupied sites and colonization of previously unoccupied sites.

Formulating the multispecies occupancy model

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

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

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 practice

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

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 resurveys 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

$$y_{i,k} \sim \mathrm{Binom}(J_k, z_{i,k}p_i).$$

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 $\mathbf{Z}_{\mathrm{post}}$, which consists of the elements $z_{i,k}$.

For all of our following simulations, we used the open-source statistical software R (R Development Core Team 2014). For Bayesian analyses, we used the open-source software JAGS (*available online*).[2] For metacommunity analyses, we used the R package metacom (Dallas 2014), which relies heavily on the R package vegan (Oksanen et al. 2013). Sample code and simulation data are provided as a Supplement and are available from the corresponding author upon request.

Integrating EMS and occupancy models: a simulated example of DECEMS

We use a simulated case study to introduce DE-CEMS, which integrates occupancy modeling and standard EMS. Further analyses of DECEMS are provided in the Appendix. First, we simulated a metacommunity of 12 species and 75 sampled sites, assuming each site was sampled three 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 $\beta_{i0} = \mathrm{logit}(0.60)$ for all species, and we assumed that species-specific covariate effects, β_i, followed a normal distribution with mean = 0 and standard deviation = 1. Site-specific covariate values followed a normal distribution with mean = 0 and SD = 2. To emphasize how detection errors can obscure true metacommunity structure, we
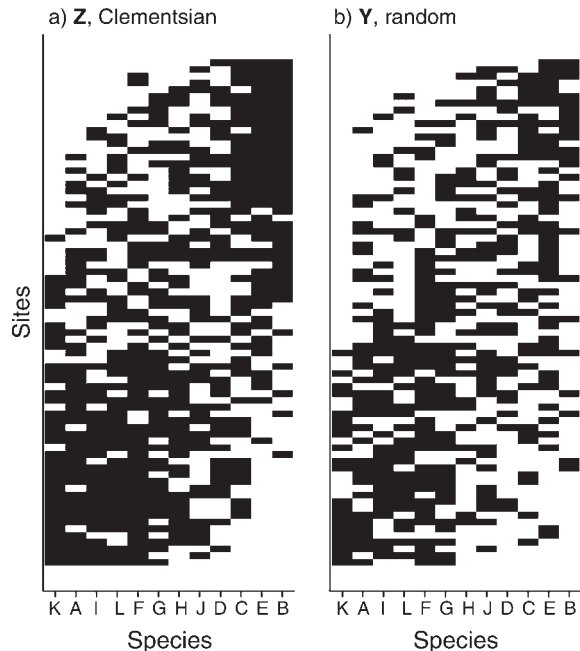
FIG. 2. The ordinated form of the (a) known occurrence matrix $\mathbf{Z}$ and the (b) corresponding observed occurrence matrix $\mathbf{Y}$ with imposed detection error from our simulated example.

assumed the species-specific probabilities of detection, logit(p_i), followed a normal distribution with mean = logit(0.5) and SD = 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 (Fig. 2), which is an expected outcome of detection error. Particularly, false negatives 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 the true occupancy states $\mathbf{Z}$, 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 $\mathbf{Z}_{\mathrm{post}}$ across three MCMC chains. We assessed convergence using the potential scale reduction factor, $\hat{R}$ (Gelman 1996).
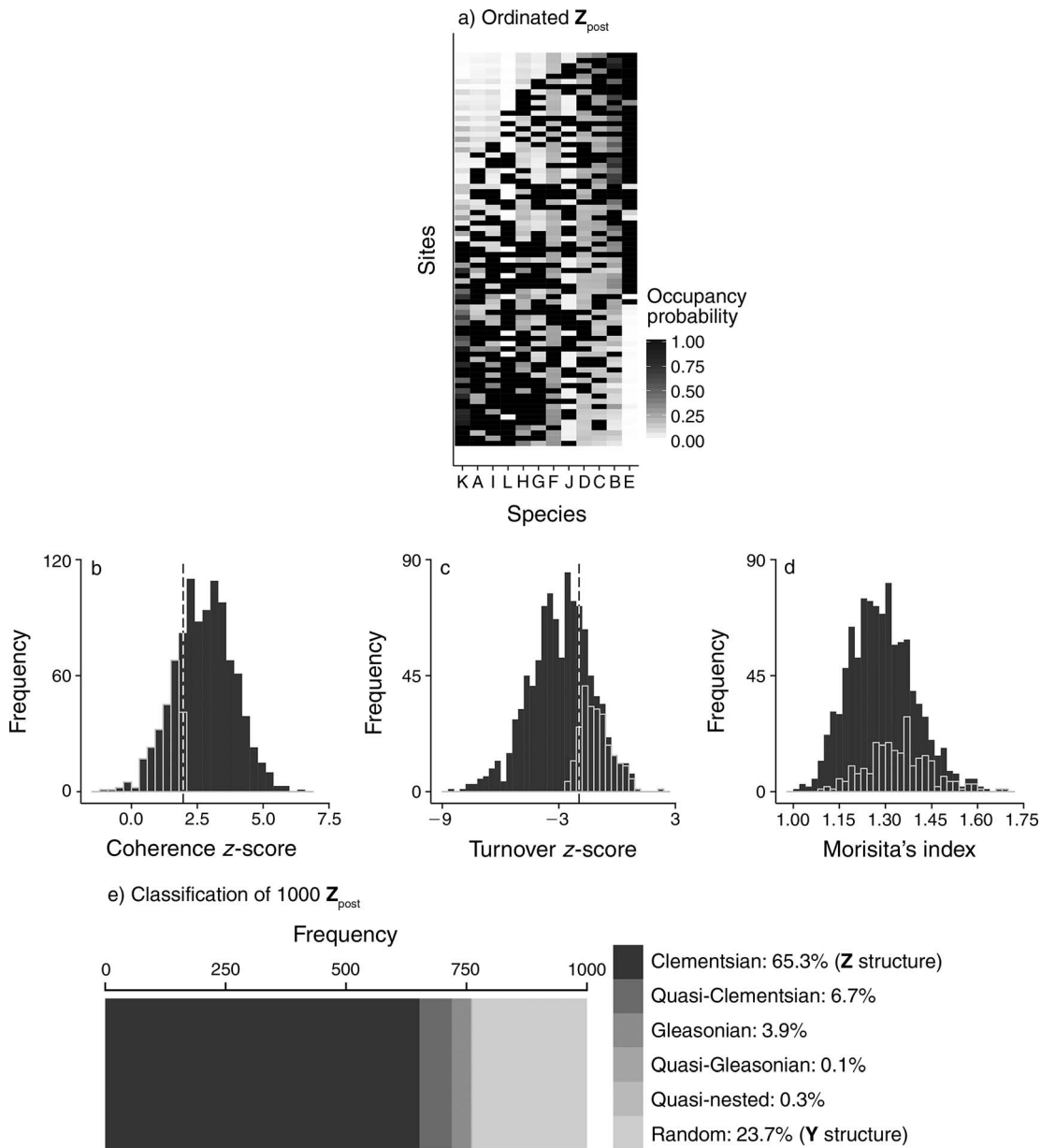
FIG. 3. Data on metacommunity structure derived from occupancy model. (a) A heat map ordination of the $\mathbf{Z}_{post}$ matrices (drawn from the posterior distribution), based on the average site- and species-specific probability of occurrence across the 1000 iterations. These values were then used 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 nonsignficant coherence (i.e., random structure). Vertical dashed lines in (b and c) delineate 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.

For 1000 $\mathbf{Z}_{post}$ from the Bayesian occupancy model, we calculated metacommunity metrics, assembling pseudo-posterior distributions for each metric (Fig. 3b–d). These are pseudo-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 $\mathbf{Z}_{post}$.

In this example, using DECEMS 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 $\mathbf{Y}$ (Fig. 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 occurrence, estimated as the proportion of occurrences observed across the 1000
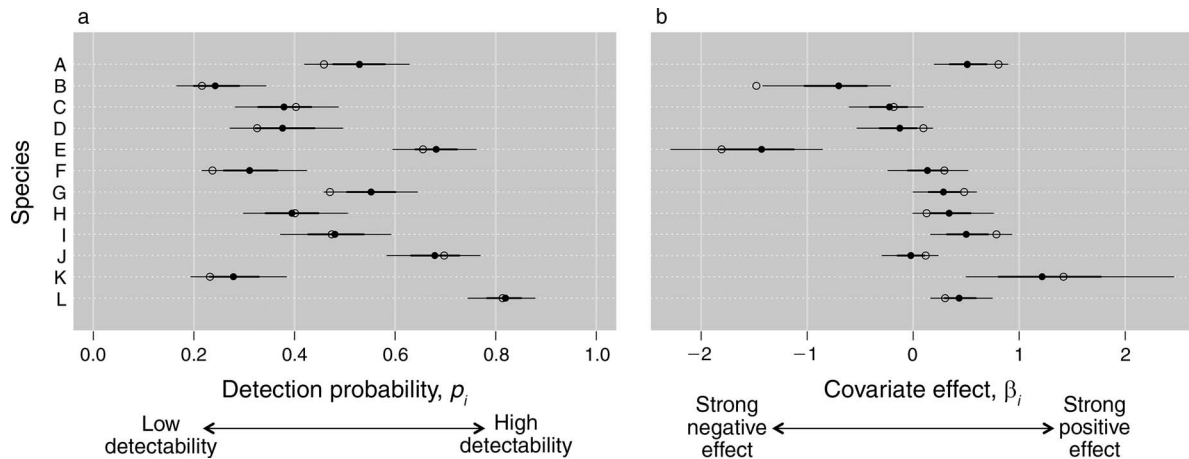
FIG. 4. Values of (a) species-specific probabilities of detection, p_i, and (b) species-specific covariate effects, β_i, as estimated by the occupancy model. Solid circles mark the medians of the posterior distributions of each parameter, while open circles mark the true (simulated) values. Thicker and thinner lines represent the 68% and 95% credible intervals of the estimates, respectively.

$\mathbf{Z}_{post}$ (Fig. 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. This is important, because a metacommunity may fit into a certain structure categorization loosely (e.g., displaying Clementsian structure, but with weak boundary clumping, bordering on Gleasonian classification). DECEMS can make this ambiguity explicit.

The model was also able to accurately estimate the species-specific probabilities of detection, p_i, and species-specific covariate effects, β_i (Fig. 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 significant positive turnover 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 contribute significantly to turnover. In metacommunities with more overall species, boundary clumping could also be predicted from the model. For instance, if species had similar predicted covariate effects, as well as similar baseline occupancy probabilities (β_{i0}), then they would likely have similar range boundaries.

This example shows that detection error can bias the assignment of metacommunity structure and, at least in this case, that 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.

Effect of detection error correction on EMS misclassification rate

Although an occupancy model was useful in the simulated example, 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 three times each. We assumed a dominant environmental covariate to which species responded, in order to achieve 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 variability, 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 SD of 0.75 (e.g., logit(p_i) ~ N(logit(0.9),0.75).

Next, we determined whether the occupancy model reduced bias in estimating metacommunity structure. First we calculated the percentage of cases in which the
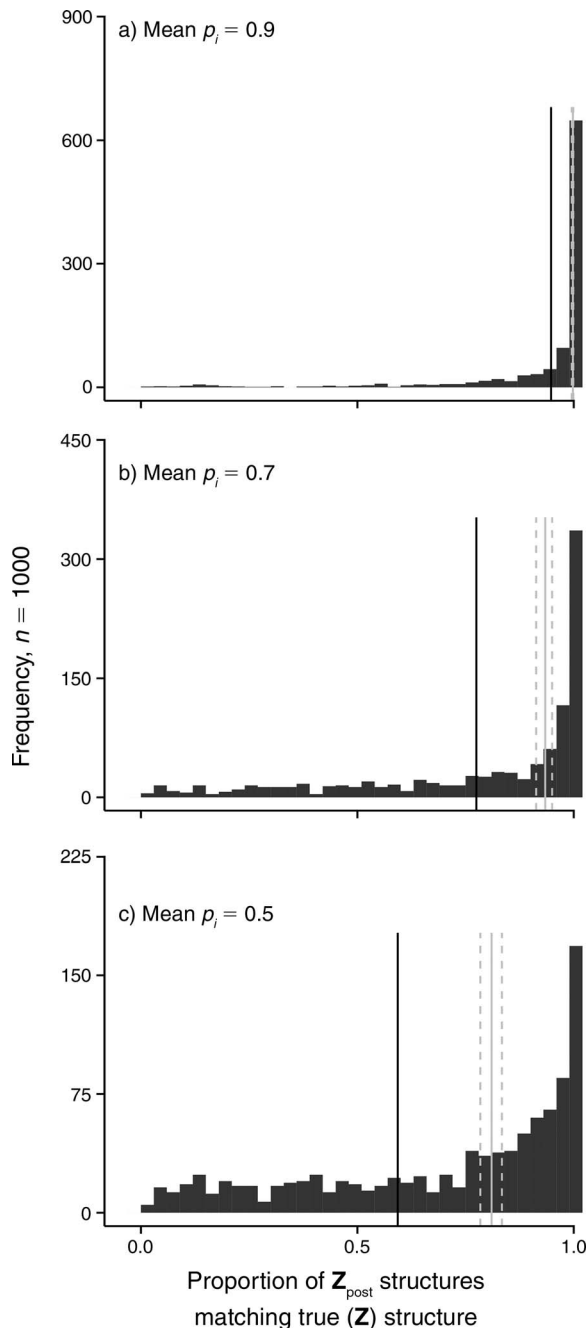
FIG. 5. Determining if an occupancy model reduces bias in assigning metacommunity structure. Black vertical lines represent the point estimate of the proportion of 1000 simulations for which the observed metacommunity, **Y**, structure matched the known metacommunity, **Z**, structure. Solid and dashed 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, **Z**, structure.

observed structure of **Y** matched the known structure of **Z**. This point estimate represents the performance of the standard EMS method. To compare this performance to that of DECEMS, for each **Z**–**Y** pair, we calculated the

percentage of cases in which the structure of $\mathbf{Z}_{post}$ matched **Z** (out of 500). Thus, we had a percentage performance for each **Z**–**Y** pair ($n = 1000$). We then calculated the median (and its bootstrapped 95% CI) of this distribution and compared this to the percentage of times the structure of **Y** matched **Z**. Effectively, this tested whether correcting for imperfect detection significantly reduced the misclassification rate of the 1000 simulated metacommunities compared to the standard EMS approach.

For all three values of mean p_i, the occupancy model significantly reduced bias in characterizing metacommunity structure (Fig. 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' (**Y**) structures matching the true occurrence matrices' (**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 **Z** structure was 99.8% (99.5–99.9%), showing a significant improvement and reducing the misclassification rate to <1% (Fig. 5a). Furthermore, the benefit of using the occupancy model increased as the mean probability of detection in the community decreased (Fig. 5b, c). Specifically, for mean $p_i = 0.7$, the structure of **Y** matched that of **Z** 77.5% of the time, while the median for $\mathbf{Z}_{post}$ matching **Z** was 93.4% (91.3–95.0%); for mean $p_i = 0.5$, the structure of **Y** matched that of **Z** 59.3% of the time, while the median for $\mathbf{Z}_{post}$ matching **Z** was 81.0% (78.4–83.4%).

These results indicate that, somewhat surprisingly, incorporating an occupancy model increases the accuracy in assigning metacommunity structure, even when there are relatively high detection probabilities (~90%) among species. 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 cryptic due to small size or aggregation patterns among hosts (e.g., negative binomial distribution).

DISCUSSION

Integrating multispecies occupancy models into the EMS framework can effectively 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). We found that integrating occupancy modeling with EMS (i.e., DECEMS) can lead to striking reductions in the rates of metacommunity structure misclassification that results from imperfect detection, even when detection error was quite low. This is an important improvement that should ensure the best possible classification of metacommunity structure from community data sets. Given that an occupancy model estimates species-specific covariate effects, this method can also be used to better inform how species-level responses can scale up to affect metacommunity-wide

patterns of occurrence. This helps to address a key gap within EMS by improving our ability to link metacommunity patterns to species-level processes.

Based on these findings, we suggest that metacommunity ecologists will often benefit from using occupancy models in their assessment of metacommunity structure and in determining the environmental covariates that might lead to structure. This method should prove particularly useful in cases where species detection is known to be problematic, although our results show that DECEMS improves EMS performance even for data sets with little detection error (more discussion in the Appendix). Occupancy modeling requires repeated surveys over a time period in which it can be safely assumed that community member composition is not changing. This design allows for the estimation of species-specific detection probabilities. The design of repeated 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 independent surveys on the same day. Thus, collecting data to accommodate an occupancy model does not necessarily have to increase effort, and the end result is improved accuracy.

Future extensions to the occupancy-modeling framework presented here could further improve how we assess the influence of environmental covariates on metacommunity structure. For instance, Jackson et al. (2012) presented a maximum likelihood method using multilevel models, for which occupancy models are a specific example, to determine how environmental variation leads to changes in community composition. These authors showed 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), although we did not test this with our particular model herein. 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 probabilities. 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 theoretical 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 strengthen 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). We have shown that integrating occupancy modeling and EMS (i.e., DECEMS) can improve the accuracy of metacommunity structure classification, an important step towards understanding where and why certain structures emerge. We propose that a continued merger of the fields of occupancy modeling and metacommunity ecology should enrich and deepen our study of populations and communities across spatial scales.

Literature Cited

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(5):e38007.

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. Carnegie Institution of Washington, Washington, D.C., USA.

Dallas, T. 2014. metacom: an R package for the analysis of metacommunity structure. Ecography 37:402–405.

Dallas, T., and S. J. Presley. 2014. Relative importance of host environment, transmission potential and host phylogeny to the structure of parasite metacommunities. Oikos 123:866–874.

de la Sancha, N. U., 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 20:1058–1070.

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:art55.

Gelman, A. 1996. Inference and monitoring convergence. Pages 131–143 in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. Markov chain Monte Carlo in practice. Chapman & Hall, New York, New York, USA.

Gilpin, M., and J. Diamond. 1982. Factors contributing to non-randomness in species co-occurrences on islands. Oecologia 52:75–84.

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., USA.

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. University of Chicago Press, Chicago, Illinois, USA.

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 and 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., et al. 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. Journal of Biogeography 40:1560–1571.

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

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2013. vegan: community ecology package. http://CRAN.R-project.org/package=vegan

Olito, C., and J. W. Fox. 2014. Species traits and abundances predict metrics of plant-pollinator network structure, but not pairwise interactions. Oikos. http://dx.doi.org/10.1111/oik.01439

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

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.

R Development Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/

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., 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.

Royle, J., and R. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic Press, Waltham, Massachusetts, USA.

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

CONCEPTS & SYNTHESIS

SUPPLEMENTAL MATERIAL

Ecological Archives

The Appendix and a Supplement are available online: http://dx.doi.org/10.1890/14-1580.1.sm