

# Estimating Abundances of Interacting Species Using Morphological Traits, Foraging Guilds, and Habitat

Robert M. Dorazio<sup>1\*</sup>, Edward F. Connor<sup>2</sup>

**1** U.S. Geological Survey, Southeast Ecological Science Center, Gainesville, Florida, United States of America, **2** Department of Biology, San Francisco State University, San Francisco, California, United States of America

## Abstract

We developed a statistical model to estimate the abundances of potentially interacting species encountered while conducting point-count surveys at a set of ecologically relevant locations – as in a metacommunity of species. In the model we assume that abundances of species with similar traits (e.g., body size) are potentially correlated and that these correlations, when present, may exist among all species or only among functionally related species (such as members of the same foraging guild). We also assume that species-specific abundances vary among locations owing to systematic and stochastic sources of heterogeneity. For example, if abundances differ among locations due to differences in habitat, then measures of habitat may be included in the model as covariates. Naturally, the quantitative effects of these covariates are assumed to differ among species. Our model also accounts for the effects of detectability on the observed counts of each species. This aspect of the model is especially important for rare or uncommon species that may be difficult to detect in community-level surveys. Estimating the detectability of each species requires sampling locations to be surveyed repeatedly using different observers or different visits of a single observer. As an illustration, we fitted models to species-specific counts of birds obtained while sampling an avian community during the breeding season. In the analysis we examined whether species abundances appeared to be correlated due to similarities in morphological measures (body mass, beak length, tarsus length, wing length, tail length) and whether these correlations existed among all species or only among species of the same foraging guild. We also used the model to estimate the effects of forested area on species abundances and the effects of sound power output (as measured by body size) on species detection probabilities.

**Citation:** Dorazio RM, Connor EF (2014) Estimating Abundances of Interacting Species Using Morphological Traits, Foraging Guilds, and Habitat. PLoS ONE 9(4): e94323. doi:10.1371/journal.pone.0094323

**Editor:** K. Ullas Karanth, Wildlife Conservation Society, India

**Received:** December 12, 2013; **Accepted:** March 15, 2014; **Published:** April 11, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was conducted as a part of the Binary Matrices Working Group at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award #EF-0832858, with additional support from The University of Tennessee, Knoxville. One author (RD) also received financial support from the U.S. Geological Survey's Status and Trends program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: bdorazio@usgs.gov

## Introduction

Much of ecological research is driven by the desire to understand observed patterns of variation in the abundance or occurrence of individual plants or animals. While this variation may be associated with the environmental requirements, dispersal ability, and biotic interactions of each species, the relative importance of these components can be difficult to quantify or assess. Part of the difficulty is methodological – that is, induced by limitations or deficiencies in the methods that are used to analyze data from community-level surveys.

For example, inferences about species interactions or associations traditionally have focused on examining patterns of co-occurrence between pairs of species observed at several locations [1]. Such data have been examined using so-called “null model” analyses aimed primarily at testing hypotheses about the importance of competitive interactions within a community of species [2–4]. In these analyses species co-occurrences are often quantified as a function of a community's incidence matrix [5,6], which includes the binary occupancy state (presence or absence) of each species at each sample location. More recently, parametric statistical models have been developed to estimate the effects of

one species on another. In some of these models the effects of imperfect detectability of a species are ignored [7,8]; in others the errors in detection of individuals are explicitly accounted for in the model's underlying assumptions [9–12]. An advantage of using parametric modeling is that the strengths of interspecific interactions are specified in terms of estimable model parameters, allowing the null hypothesis (no interactions) to be tested and specific alternatives to be quantified.

While these recent advances are useful, species occupancy state provides a relatively coarse summary of a species' local population size  $N$  – specifically that  $N > 0$  (species is present) or  $N = 0$  (species is absent). Interactions between species are likely to be inferred more easily by analyzing a matrix of species- and location-specific abundances. For example, the abundances of strongly interacting species are likely to be positively or negatively correlated; however, these correlations may be difficult to detect if analyses are limited to frequencies of co-occurrence between species. Furthermore, analyses of incidence matrices may be more sensitive to sampling errors because a species of low abundance may incorrectly be regarded as absent even though  $N > 0$  for this species. For these reasons, models of species counts are expected to be more useful

than models of species occurrences in estimating the magnitude of interspecific interactions.

Several approaches have been used to analyze species-specific counts of individuals (plant or animals) encountered while surveying a set of ecologically relevant locations. Many of these approaches assume pairwise independence among species and focus on examining the effects of environmental or habitat covariates on species abundance. In this approach the counts of individuals of different species are aggregated to obtain a total count of individuals that belong to a community, a guild, or a functional group [13–19]. Linear models are used to relate these totals to location-specific measurements of habitat. While statistical analyses of such aggregated counts obviously suffer from a loss of information, the results of the analyses also can be misleading. For example, estimates of covariate effects on abundance may primarily reflect the effects on the most abundant species. The counts of rare and potentially strongly interacting species will be swamped by the counts of the abundant species. Also, significant effects of covariates on abundance may be difficult to detect if species of nearly equal abundance have opposite responses to the same covariates.

Another approach often used in the analysis of species-specific counts involves multivariate ordination [18,20–22]. In these analyses the distance-based metric selected for ordination generally fails to specify the mean-variance relationship in the counts correctly, leading to conclusions (such as statistical significance of covariate effects) that confound differences in mean with differences in variance [23].

In other studies species-specific counts are analyzed to detect potential associations between species that may suggest the presence of competitive interactions or habitat associations. Not surprisingly, some of these analyses are based on null models for testing pairwise interactions between species [24–27]. These analyses differ in underlying assumptions and in the indices used to summarize patterns of species co-occurrence and spatial aggregation. Importantly, however, rejection of the null hypothesis (no interactions) is not accompanied by estimation of the magnitude of interspecific interactions – primarily because null model analyses usually are conducted nonparametrically (see [28] for an exception).

Another approach to the analysis of species-specific counts is based on univariate, linear-regression models wherein counts of one species are regressed on habitat measurements and on standardized counts of other species to estimate pairwise species interaction strengths (“competition coefficients”) [29–31]. In this approach counts of individuals are regarded as surrogates of abundance; however, these regression models do not honor the discrete nature of the counts (which are assumed to be normally distributed), and the counts are treated as both stochastic response variables and fixed predictors, even though both are clearly outcomes of sampling.

An alternative approach is based on fitting statistical models that assume the counts of individuals are distributed as mixtures of Poisson and lognormal distributions. Similar to the regression models, the counts are regarded as surrogates of abundance; however, Poisson-lognormal mixtures account for discreteness of the counts and for the increase in variance with mean that is induced when individuals are distributed randomly or in spatial aggregations [32,33]. A univariate Poisson-lognormal mixture for modeling species-specific counts of individuals observed at a single location was proposed by [34]. This model was extended by [35] for pairs of locations but considered only stochastic sources of variation in abundance among locations (as opposed to specifying the effects of differences in habitat on species abundances). In both

models species-specific abundances were assumed to vary exchangeably among species (via the lognormal distribution) and the abundances of different species were not assumed to be correlated (say, as a consequence of interactions among species). The multivariate Poisson-lognormal distribution [36] was developed to allow abundances of different species to be correlated; however, heterogeneity in species abundances among locations was not specified as a function of habitat, which may explain why the model has not attracted much attention by ecologists. This model was extended recently to include the effects of location-specific covariates on species abundances [37].

An implicit assumption of all of the count models described above is that each observed count is assumed to equal the actual abundance of individuals. In other words, every individual of every species is assumed to be detected with the same probability (one) at every sample location. This assumption is seldom satisfied in surveys of natural communities because individuals which are present and available for detection are routinely missed during sampling [38,39]. Furthermore, the probability of detection generally varies considerably among species. When an entire community is sampled, many species may be represented by only one or a few individuals at each location, so it is crucial that analyses account for the effects of imperfect detectability if species abundances are to be estimated accurately [40,41].

$N$ -mixture models [42] allow abundances to be estimated from samples of repeated point counts while accounting for imperfect detections of individuals. In these models the number of individuals detected during a survey (i.e., the observed count) is assumed to equal the sum of  $N$  independent Bernoulli outcomes each with identical success (detection) probability. However, since the abundance of animals at each sample location is unknown,  $N$  must be estimated by coupling the Bernoulli model of the observations with a model that specifies how  $N$  varies among locations. This model of abundance includes a mixing distribution for  $N$ , and hence the name  $N$ -mixture models. Although extensions of these models have been proposed for multiple species [43–45], to our knowledge [46] provide the only published study in which  $N$ -mixtures have been used to analyze counts from multiple, interacting species. In this study, though, the effects of habitat on species abundances were not included in the analysis.

Our study was motivated by a desire to determine whether the effects of environmental (habitat) covariates or interactions between species were more influential in determining the abundances of individual species. Specifically, we sought to develop a statistical model of counts from community-level surveys that overcomes many of the limitations of existing models. In the paper we propose a model to estimate the abundances of potentially interacting species using repeated point counts of each species observed while sampling a set of ecologically relevant locations. This model includes two extensions of the multivariate Poisson-lognormal mixture: it allows abundances of species with similar traits to be correlated and it allows the effects of habitat on abundance to be estimated. Applying the  $N$ -mixture idea, we combine this model of species abundances with a binomial model of the observed counts that allows the probability of detection to be estimated for each species in conjunction with its abundance. To illustrate the potential benefits of this approach, we analyzed point counts obtained while sampling a community of forest birds during the breeding season. These data are far from ideal for our analysis, but they are typical of the kind of information obtained in avian community-level surveys. Our objective here is to provide a proof-of-concept of our statistical approach, not an exhaustive analysis of the avian point count data.

### Statistical Analysis

In this section we describe a statistical model for estimating the abundances of potentially interacting species encountered while surveying a set of distinct locations. The sample locations are assumed to be representative of some collection of locations that is considered to be ecologically relevant – as in a metacommunity of species. In the model we assume that abundances of species with similar traits are potentially correlated and that these correlations, when present, may exist among all species or only among functionally related species (such as members of the same foraging guild). We also assume that abundances vary among locations owing to systematic and stochastic sources of heterogeneity. For example, if abundances differ among locations due to differences in habitat, then measures of habitat can be included in the model as covariates. Naturally, the quantitative effects of these covariates are assumed to differ among species.

Our model also accounts for the effects of detectability on the observed counts of each species. This aspect of the model is especially important for rare or uncommon species that may be difficult to detect in community-level surveys. Estimating each species’ detectability requires some sampling locations to be surveyed repeatedly using different observers or different visits of a single observer. These within-location replicates provide the information needed to estimate both abundance and detectability of each species.

### Modeling species abundances

We begin by describing a model of species- and location-specific abundances, which are not directly observable but are the quantities of primary scientific interest. Let  $N_{ik}$  denote a random variable for the number of individuals of species  $i$  that are present and available to be observed at sample location  $k$  ( $i=1, \dots, I$ ;  $k=1, \dots, K$ ).  $I$  is the number of distinct species observed among all  $K$  sample locations. We assume that the expected abundance of individuals of species  $i$  at location  $k$ , say  $\lambda_{ik}$ , is constant during the surveys of each location and that  $N_{ik} \sim \text{Poisson}(\lambda_{ik})$ .

We specify location-specific differences in the abundance of individuals using a log-linear regression model of  $\lambda_{ik}$ :  $\log(\lambda_{ik}) = \mathbf{b}'_i \mathbf{x}_k$ . (A prime symbol indicates the transpose of a vector or matrix.) In this model the vector  $\mathbf{x}_k = (1, x_{1k}, \dots, x_{qk})'$  includes the observed values of  $q$  covariates thought to be informative of abundance at location  $k$ . For example,  $x_k$  may include measures of habitat at location  $k$ . The parameter vector  $\mathbf{b}_i = (b_{0i}, b_{1i}, \dots, b_{qi})'$  includes an intercept and the effects of the  $q$  covariates. If the values of each covariate are centered at zero, the intercept parameter  $b_{0i}$  denotes the log-scale, expected abundance of species  $i$  at the average value of the covariates.

To allow for potential correlations in abundance among species, we specify dependence among the  $I$  intercepts using the multivariate normal distribution:

$$(b_{01}, b_{02}, \dots, b_{0I})' \sim \text{Normal}(\beta_0 \mathbf{1}, \sigma_{b_0}^2 \mathbf{R}) \tag{1}$$

where  $\mathbf{1}$  is an  $I$ -vector of ones,  $\beta_0$  and  $\sigma_{b_0}^2$  parameterize the mean and variance among intercepts, and  $\mathbf{R}$  is their  $I \times I$  correlation matrix. This specification follows the approach taken by [36] but without an unstructured correlation matrix. To ensure that  $\mathbf{R}$  is positive definite, we borrow an idea from geostatistical modeling [47,48] and use an exponential function,  $R_{ij} = \exp(-d_{ij}/\phi)$ , to specify the correlation between abundances of species  $i$  and  $j$  as a function of a positive, scalar-valued parameter  $\phi$  and a measure of

dissimilarity  $d_{ij}$  in traits of species  $i$  and  $j$ . For simplicity, we propose the Euclidean distance  $\|\mathbf{v}_i - \mathbf{v}_j\|$  between trait vectors  $\mathbf{v}_i$  and  $\mathbf{v}_j$  as a measure of  $d_{ij}$ , but alternative measures of dissimilarity are possible (provided  $d_{ij} > 0, \forall i \neq j$ ). Thus, if traits of species  $i$  and  $j$  are similar (low  $d_{ij}$ ), their abundances are positively correlated under this model.

Our decision to specify correlations in abundance as a function of similarity of morphological traits was motivated by the need to ensure that  $\mathbf{R}$  is positive definite; however, the idea that species of similar morphology may have positively correlated abundances also has ecological support. For example, two partially opposing views exist about how species interactions might affect correlations in abundance between species. One view is based on the concept of limiting similarity in which species of similar morphology compete for the same resources [49], and coexistence of these species requires them to “partition” the environment to minimize overlap in resource use [50]. Under this view one might expect that species of similar morphology – and therefore similar resource use – might not co-occur, or if they do co-occur, their abundances might be negatively correlated. However, the spatial scale at which resource partitioning occurs is not known. The shared-niche model suggests that at larger spatial scales the occurrences and abundances of competing species may be positively correlated even though resources are partitioned at a smaller scale [51]. In addition, most species of birds do not maintain interspecific territories. Finally, [52]’s classic study of warblers suggests that resource partitioning among avian species – if it occurs – might be at a very fine scale (e.g., different parts of the forest canopy). Therefore, estimated abundances of individuals vulnerable to detection in a point-count survey may not reflect competitively mediated resource partitioning because the region of detection may – and often does – include a wide variety of microhabitats and individual species territories.

Another view is based on the idea that species which are morphologically similar and consume the same resources are likely to aggregate at the same sites where shared resources or habitat is abundant. There is evidence that closely related species, which tend to be morphologically similar, tend to co-occur more than expected [53,54] or to have their occurrences or abundances positively correlated [8,55], at least at larger spatial scales. Species co-occurrence and positive correlations in abundance probably reflect associations caused by shared habitat preferences or shared resource requirements. In light of these two views, we suggest that at the scale of our point-count surveys the decision to specify correlations in abundance between species as a positive function of their morphological similarity is not unreasonable.

To complete the model of species abundances, we assume that species-specific effects of the  $m$ th covariate of mean abundance ( $m=1, \dots, q$ ) vary exchangeably among species as follows:

$$(b_{m1}, b_{m2}, \dots, b_{mI})' \sim \text{Normal}(\beta_m \mathbf{1}, \sigma_{b_m}^2 \mathbf{I})$$

where  $\beta_m$  and  $\sigma_{b_m}^2$  are mean and variance parameters, respectively, and  $\mathbf{I}$  is a  $I \times I$  identity matrix. Thus, while the effects of any particular habitat covariate may differ among species, these effects are assumed to vary independently among species. In addition, we assume mutual independence between the vector of intercept parameters and the vectors of covariate effects.

**Submodels of species abundances.** Two special cases of our abundance model are noteworthy. First, we may consider a model wherein correlations in abundance are assumed only among related species. For example, we may assume that abundances are positively correlated among species of the same

foraging guild but not correlated with species in different guilds. If the species are ordered by guild, this model is essentially the same as that specified in Eq. 1 except that the correlation matrix  $\mathbf{R}$  is block-diagonal:

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_2 & & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_G \end{pmatrix}$$

where each block corresponds to the correlation matrix of a distinct foraging guild  $\mathbf{R}_g$  ( $g=1, \dots, G$ ). Elements of each submatrix  $\mathbf{R}_g$  depend on the traits of species in guild  $g$  and on the parameter  $\phi$  as described earlier.

Another special case of our abundance model corresponds to absence of correlations in abundance wherein  $\mathbf{R} = \mathbf{I}$ . In this model we assume that all species are ecologically similar so that their intercepts (i.e., log-scale mean abundances) are exchangeable. In this case differences in species traits are assumed to be uninformative of differences in species abundances, and the abundance of one species is assumed to be independent of the abundance of another species.

### Modeling species counts

Species- and location-specific abundances usually cannot be observed directly in multi-species surveys. Instead, counts of individuals are typically observed, but the counts are subject to errors in detection of individuals. An observed count generally underrepresents the number of individuals of each species present, and this bias is particularly true for uncommon species that are easily missed. Therefore, a model is needed to relate the observed counts of each species to the actual abundance of that species.

We assume that  $J_k$  independent surveys are conducted at the  $k$ th sample location such that each species' abundance remains constant during the period of sampling. These surveys may be conducted by independent observers or by repeated visits of a single observer. As proposed by [42], the idea is to treat the individuals at each site as a closed population of size  $n_{ik}$  (a realization of the random variable  $N_{ik}$ ) and to use the replicate counts of these individuals to estimate the probability of detection per individual.

Given these assumptions, we describe a model of the observed counts that allows detection probabilities to be estimated for each species. Let  $Y_{ikj}$  denote a random variable for the number of individuals of species  $i$  detected during the  $j$ th survey of sample location  $k$  ( $j=1, \dots, J_k$ ). We assume that  $Y_{ikj}|n_{ik} \sim \text{Binomial}(n_{ik}, p_{ikj})$ , where  $p_{ikj}$  denotes the conditional probability of detecting an individual of species  $i$  during the  $j$ th survey of site  $k$ . In addition, because surveys are conducted independently at each site, we assume that each site's observations are conditionally independent, that is,

$$\Pr(Y_{ik1} = y_{ik1}, \dots, Y_{ikJ_k} = y_{ikJ_k} | n_{ik}, p_{ikj}) = \prod_{j=1}^{J_k} \Pr(Y_{ikj} = y_{ikj} | n_{ik}, p_{ikj})$$

Various models of  $p_{ikj}$  may be constructed depending on the availability of species- or site-specific covariates thought to be informative of detectability. For example, suppose a vector of  $r$  covariate measurements  $\mathbf{w}_i = (w_{1i}, \dots, w_{ri})'$  is thought to be informative of the detectability of species  $i$ . The effects of these covariates on  $p_{ikj}$  can be specified using a logit-linear regression model with random intercepts as follows:

$$\text{logit}(p_{ikj}) = a_{0i} + \alpha' \mathbf{w}_i$$

**Table 1.** Posterior means and 95% credible intervals for the parameters of three models.

Parameter	Model 1			Model 2			Model 3		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
$\beta_0$	-0.741 (0.008)	-1.165 (0.008)	-0.302 (0.008)	-0.786 (0.009)	-1.270 (0.012)	-0.300 (0.009)	-0.774 (0.007)	-1.186 (0.008)	-0.381 (0.007)
$\sigma_{\beta_0}$	1.472 (0.004)	1.202 (0.004)	1.798 (0.006)	1.641 (0.007)	1.308 (0.006)	2.060 (0.010)	1.475 (0.005)	1.193 (0.005)	1.806 (0.007)
$\phi$	-	-	-	0.069 (0.002)	0.003 (0.0005)	0.240 (0.006)	0.016 (0.0003)	0.001 (0.0001)	0.050 (0.001)
$\beta_1$	-0.112 (0.001)	-0.291 (0.002)	0.066 (0.001)	-0.112 (0.001)	-0.292 (0.002)	0.070 (0.002)	-0.108 (0.001)	-0.286 (0.002)	0.072 (0.002)
$\sigma_{\beta_1}$	0.696 (0.002)	0.559 (0.002)	0.860 (0.003)	0.698 (0.002)	0.564 (0.002)	0.862 (0.003)	0.693 (0.002)	0.558 (0.002)	0.854 (0.003)
$\alpha_0$	-1.696 (0.009)	-2.105 (0.012)	-1.399 (0.009)	-1.649 (0.009)	-2.019 (0.011)	-1.331 (0.009)	-1.637 (0.007)	-1.955 (0.010)	-1.347 (0.007)
$\sigma_{\alpha_0}$	0.883 (0.008)	0.620 (0.007)	1.242 (0.010)	0.849 (0.007)	0.612 (0.006)	1.161 (0.009)	0.835 (0.007)	0.595 (0.006)	1.133 (0.008)
$\alpha_1$	-0.669 (0.011)	-1.031 (0.011)	-0.288 (0.011)	-0.631 (0.013)	-1.036 (0.013)	-0.170 (0.013)	-0.631 (0.010)	-1.002 (0.011)	-0.283 (0.011)

Model 1 assumes uncorrelated abundances. Model 2 assumes correlated abundances among species of the same foraging guild. Model 3 assumes correlated abundances among all species. Monte Carlo standard errors are given in parentheses.  
doi:10.1371/journal.pone.0094323.t001

$$a_{0i} \stackrel{iid}{\sim} \text{Normal}(\alpha_0, \sigma_{a_0}^2)$$

In this formulation latent (unobserved) sources of heterogeneity in detection among species are specified exchangeably using the normal distribution, whereas systematic (observed) sources of heterogeneity in detection are specified by the regression parameters  $\alpha = (\alpha_1, \dots, \alpha_q)'$ . We used this example because it corresponds to the model used in our analysis of point counts of avian species. More generally, however, the observation model – that is, the model of the observed counts – depends on the availability of auxiliary data, which could include species-, location- or even survey-specific covariates. Therefore, this approach to modeling building is extremely versatile. Adopting a hierarchical approach to model building [41], wherein separate models are used to describe the ecological and sampling processes, allows us to specify an explicit relationship between the observed counts and the latent abundances of each species.

### Estimating model parameters

The hierarchical model described in the previous sections would be difficult to fit using classical methods owing to the high-dimensional and analytically intractable integrations involved in evaluating a marginal (integrated) likelihood function of upper-level parameters ( $\beta, \alpha, \phi$  and variance parameters). We therefore adopted a Bayesian approach to inference and used Markov chain Monte Carlo methods [56] to fit the model and to estimate its parameters. In Appendix S1 we describe the algorithm used to calculate summaries of the posterior distribution and other ecologically relevant functionals of the Markov chain.

We used a posterior-predictive loss criterion [57] to compare the three models of species-specific abundances. Specifically, we used the following criterion:

$$D = \sum_i \sum_j \sum_k \text{Var}(\tilde{y}_{ikj} | \text{data}) + \{E(\tilde{y}_{ikj} | \text{data}) - y_{ikj}\}^2$$

which minimizes the expected squared-error loss between the observed counts ( $y_{ikj}$ ) and the counts predicted under a model ( $\tilde{y}_{ikj}$ ). Models with lower values of this criterion are preferred because they have lower predictive variance (first term in  $D$ ) and lower lack of fit (second term in  $D$ ).

### Description of Data Sets

We analyzed species- and location-specific counts of 73 avian species observed while sampling 46 tracts of forest in southeastern Connecticut [58]. Tracts were defined as areas of forest not interrupted by powerlines, highways, or unforested areas wider than 10 m. [58] provide a detailed description of the species, sampling methods, and environmental covariates. Briefly, birds were detected aurally (and sometimes visually) within each of 89 sample locations during three 20-min, point-count surveys (100 m radius). During each survey the locations of birds detected were mapped, and birds that were spatially separated and singing were treated as distinct individuals. Repeated surveys of the same location were conducted during early morning hours (530 h to 1000 h) on three separate days during the breeding season (21 May to 11 July). The sample locations were sufficiently far apart that birds were unlikely to have been observed at more than one location. The movements of birds were limited during the sampling period owing to mating and nesting behaviors.

Of the various environmental covariates measured by [58], tract-level measures of forested area appeared to be the best predictors of counts of avian species. We therefore used one of these measures – forested area (ha) within 2 km of a tract's center – to describe differences in avian habitat among sample locations and to predict the abundance of individuals of each species at these locations. Appendix S2 contains the species-specific point counts and measurements of forested area at each sample location.

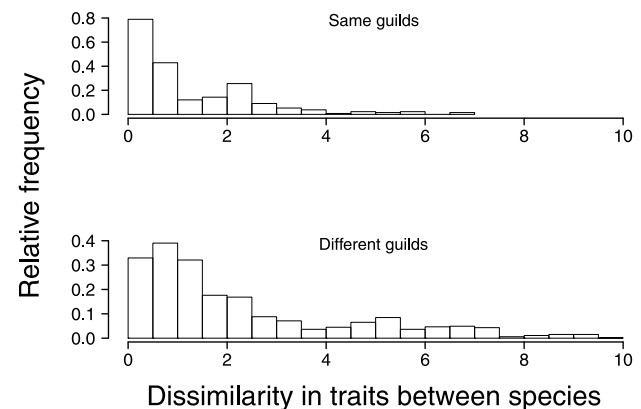
[58] also used habitat to classify each species into one of three categories: (1) interior species, whose territories are normally restricted to the interior of a forest, (2) edge species, whose territories are primarily concentrated on a forest's edge, and (3) interior-edge species, whose territories may include both forest interior and edge. We did not use these categories to model species abundances; however, we did examine whether our model-based estimates of species-specific abundances appeared to agree with [58]'s classification of species.

We used several kinds of information about individual species to inform different components of our model. For example, we used body mass data [59] to estimate the sound power output (mW) of singing males according to an allometric relationship reported by [60]. We reasoned that birds with greater sound power output (i.e., higher body mass) would, on average, be more detectable than birds with lower sound power output.

We also used species-specific morphological information to model correlations in abundance between species. We obtained measures of morphological traits (body mass, beak length, tarsus length, wing length, and tail length) for each species (primarily from [61]). We then reduced the dimensionality of the trait measurements by performing a principal components analysis on their correlation matrix and by computing principal-component scores from the two eigenvectors associated with the highest two eigenvalues.

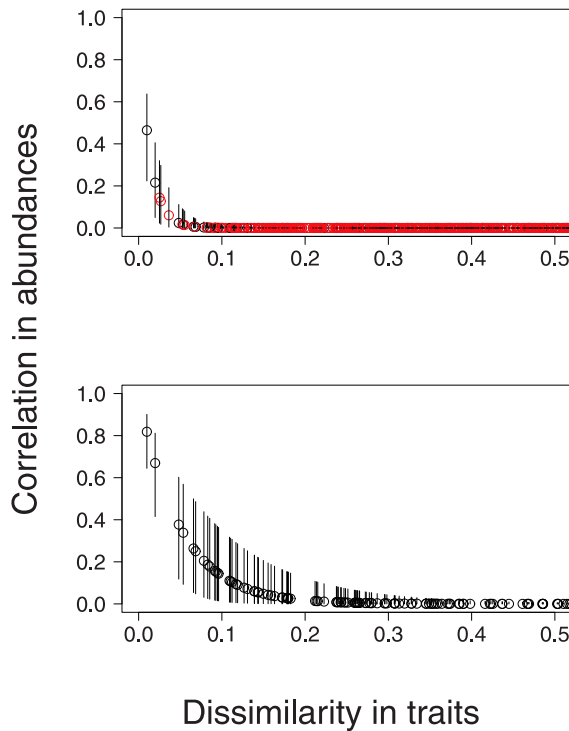
We examined whether abundances of all species or only functionally related species were correlated by using foraging guilds to define inter-relatedness among species. We used foraging guilds assigned by [62] for the breeding season if a species used different foraging strategies seasonally; if not, we used the year-round guild assignment of a species. Appendix S3 contains a species list with guild assignments, morphological traits, and sound power estimates.

We fitted three models to the avian point counts. Each model included the effects of forested area on species-specific abundances



**Figure 1. Dissimilarity in morphological traits between species of same foraging guild (upper panel) and between species of different foraging guilds (lower panel).**

doi:10.1371/journal.pone.0094323.g001



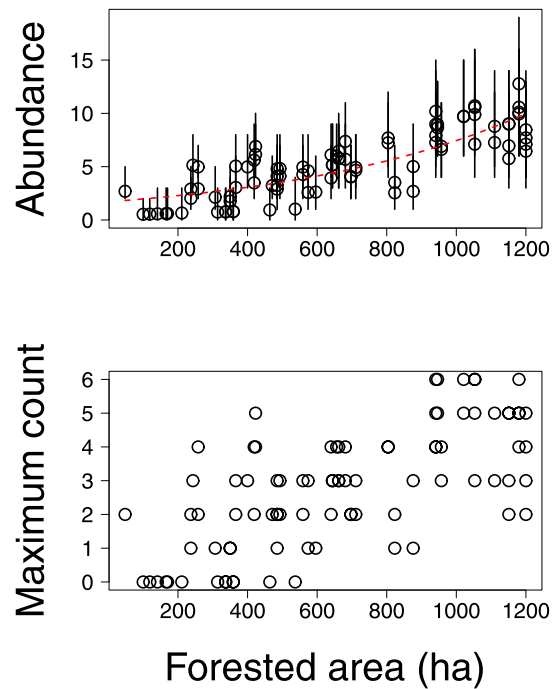
**Figure 2. Estimated correlations in abundance between species (posterior medians and 95% credible intervals) as a function of dissimilarity in their morphological traits.** Black indicates species of same foraging guild; red indicates species of different foraging guilds. Upper panel corresponds to estimates obtained by assuming correlated abundances among all species; lower panel corresponds to estimates obtained by assuming correlated abundances among species of same foraging guild. Estimated correlations are equal to zero for species with trait dissimilarities greater than 0.5. doi:10.1371/journal.pone.0094323.g002

and the effects of sound power output on species-specific detection probabilities. The three models differed in the assumed pattern of correlation between abundances of different species – that is, we assumed no correlation, correlation among species of the same foraging guild, or correlation among all species.

**Results**

Based on our model-selection criterion, the model in which abundances were assumed to be correlated among all species was favored over the other, less-complex models, but the differences between models were not that great ( $D=7129$  (correlated abundances among all species),  $D=7133$  (uncorrelated abundances),  $D=7143$  (correlated abundances among species of the same foraging guild)). The similarity in these model fits was evident also in estimates of their parameters (Table 1). For example, estimates of  $\phi$ , which determine the magnitude of correlations in abundance, were close to zero even though the dissimilarities in morphological traits ranged from 0 to 10 (Figure 1). This result was evident also in the estimates of correlation between abundances of different species. Abundances of species of very similar morphology were positively correlated; however, the abundances of most species were not strongly correlated (Figure 2).

In contrast to morphological traits, habitat – as measured by forested area – appeared to have a strong influence on avian abundances. The effects of forested area were significantly positive for 15 species and significantly negative for 23 species (Appendix



**Figure 3. Maximum observed count of ovenbirds and estimates of their abundance (posterior means and 95% credible intervals) at sample locations.** Dashed line indicates the estimated relationship between mean abundance of ovenbirds and forested area. doi:10.1371/journal.pone.0094323.g003

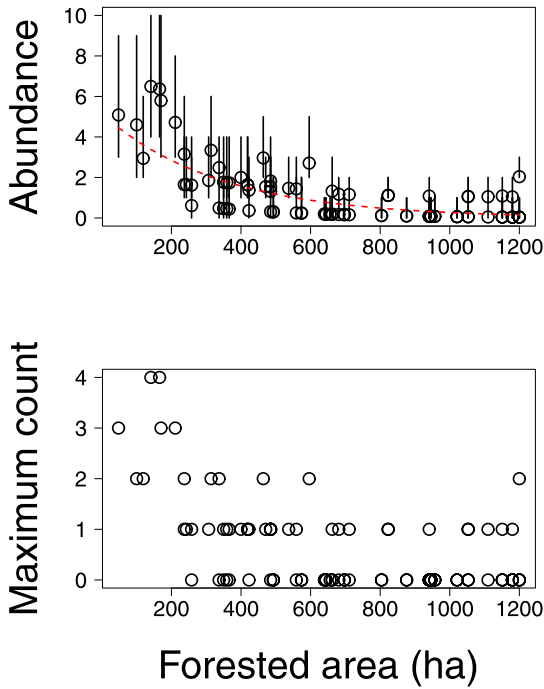
S4). As an illustration Figure 3 shows that estimates of ovenbird abundance were generally higher at locations with greater forested area. This trend was evident also in the maximum counts of ovenbirds detected at sample locations. Estimates of catbird abundance provide an example of the opposite trend, i.e., lower abundances and lower counts at sample locations with greater forested area (Figure 4).

Our estimates of abundances appear to agree with the classification of species proposed by [58]. The estimated abundances of forest interior species increased with the forested area of sample locations (Figure 5), whereas edge species were more abundant at sample locations with less forested area. The estimated abundances of interior-edge species were highest at sample locations with either low or high forested area, suggesting that this category may contain a mixture of species that prefer interior or edge habitats.

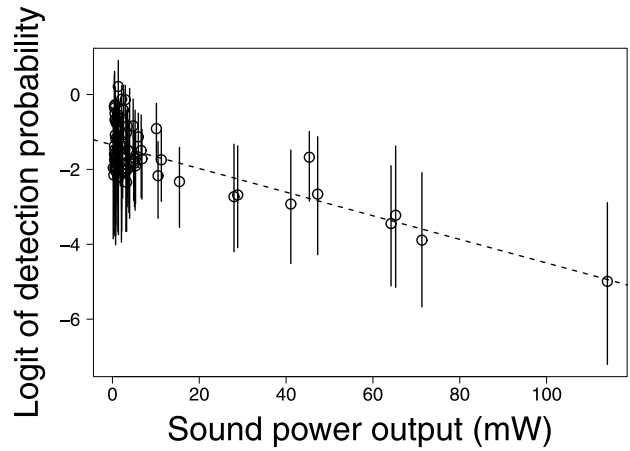
Contrary to our prior beliefs about detectability of birds, species-specific estimates of detection probability decreased with increases in sound power output (Figure 6); however, most of the decrease was associated with larger species. The estimated detection probabilities of smaller species were highly variable ( $\hat{\sigma}_{a_0}=0.835$ ). Although larger birds have higher sound power outputs, they also vocalize less frequently than smaller birds, and calls of larger birds tend to be lower in frequency and more easily masked by background noises [63]. A more complete model of heterogeneity in species-specific detection probabilities might include measures of average acoustic frequency and how often each species calls.

**Discussion**

Our model extends the multivariate Poisson-lognormal mixture proposed by [36] to include effects of habitat covariates on

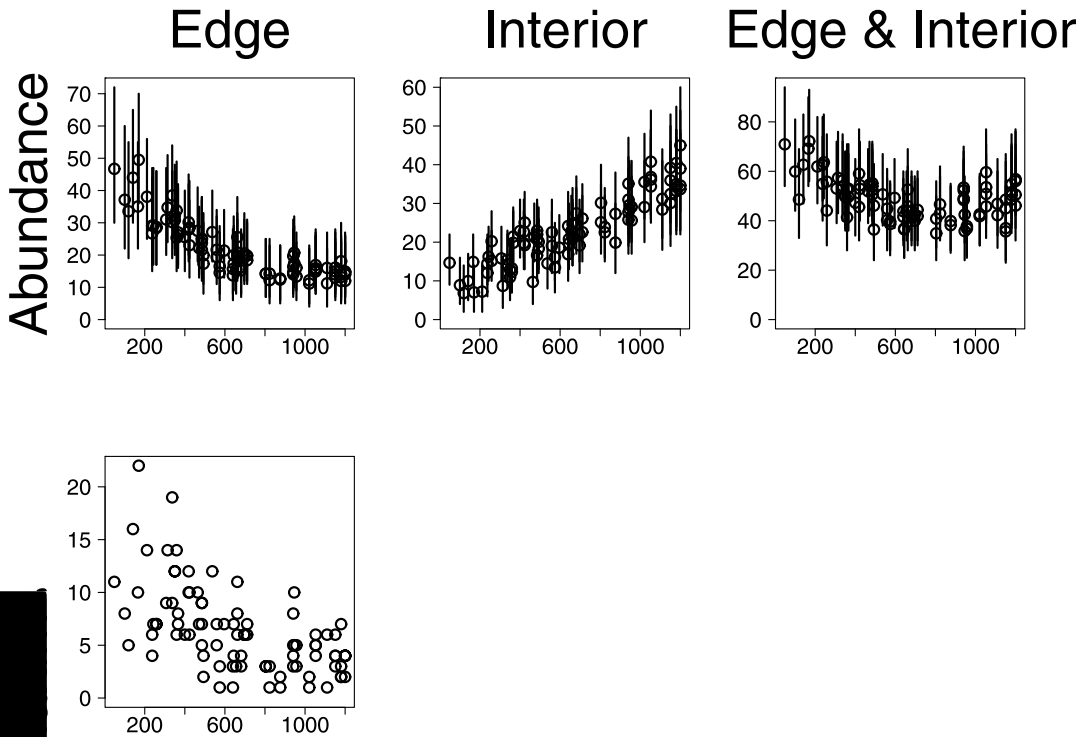


**Figure 4. Maximum observed count of catbirds and estimates of their abundance (posterior means and 95% credible intervals) at sample locations.** Dashed line indicates the estimated relationship between mean abundance of catbirds and forested area.  
doi:10.1371/journal.pone.0094323.g004



**Figure 6. Estimated probability of detection (on logit scale with 95% credible interval) vs. sound power output (mW) of each species.**  
doi:10.1371/journal.pone.0094323.g006

abundance, to specify pairwise correlations in abundance between species in terms of morphological similarity, and to account for errors in detection of individuals. The model is a multispecies,  $N$ -mixture model because the observed counts of each species are modeled conditional on a set of latent abundance parameters and a separate component of the model is used to specify heterogeneity



in abundance of individuals among species and locations. In our model information from replicated surveys within locations is used to inform species detection probabilities. In this regard our approach is similar to that of [43], [44] and [45]; however, in those models the abundances of different species are not assumed to be correlated. Our approach also differs from that of [37], who extended the multivariate Poisson-lognormal mixture for the effects of habitat but did not formally address the impact of detection errors on estimates of abundance.

In our analysis of avian point counts, habitat exerted strong effects on the abundances of most species. These effects were evident in estimates of abundance of individual species (Figures 3 and 4) and in the estimated abundances of groups of species classified by [58] (Figure 5). An important benefit of our modeling approach is that we can estimate the residual correlations in abundance between species having accounted for the effects of habitat on abundances. In our analysis the abundances of different species did not appear to be strongly correlated unless these species were very similar morphologically (Figure 2).

However, an important limitation of our model is that species abundances can only be positively correlated owing to the assumed structure of the correlation matrix  $\mathbf{R}$ . Ideally, we also would like to be able to estimate negative correlations (e.g., those induced by competitive interactions between species). For example, [36]'s multivariate Poisson-log normal mixture assumes

$$(b_{01}, b_{02}, \dots, b_{0I})' \sim \text{Normal}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma})$$

where  $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \dots, \beta_{0I})'$  and  $\boldsymbol{\Sigma}$  denotes an unstructured  $I \times I$  matrix of variances and covariances. Estimates of  $\boldsymbol{\Sigma}$  may include positive and negative covariances. Unfortunately, our attempts to fit this model to the avian counts were unsuccessful. We suspect that limited information in the data is responsible for some of the estimation problems. For example, in our model the species- and location-specific abundances  $N_{ik}$  are latent parameters, not data. In the model of [36],  $N_{ik}$  corresponds to an *observed* count, and the replicate observations among locations provide direct information about the parameters  $\boldsymbol{\beta}_0$  and  $\boldsymbol{\Sigma}$ . Also, the number of parameters to be estimated in  $\boldsymbol{\Sigma}$  ( $I(I+1)/2 = 2701$ ) is relatively large compared to the number of sample locations ( $K = 89$ ); thus problems of parameter identifiability may have limited our ability to fit models with unstructured  $\boldsymbol{\Sigma}$ .

This topic is obviously important and requires additional research. Being able to estimate correlations in abundance between species *and* the effects of environmental covariates on those species allows us to compare their relative magnitudes. However, a word of caution is in order here. While it may be tempting to interpret the correlations in abundance as evidence of

interspecific interactions, the correlations also may be produced if the abundances of supposedly “interacting” species are influenced (positively or negatively) by unobserved – and therefore unmodeled – environmental covariates [64]. Care is therefore recommended during interpretation of results.

Despite the potential for misinterpretation, we believe that multispecies  $N$ -mixture models provide a useful conceptual framework for the analysis of community-level survey data. We anticipate that new data sets with greater numbers of sample locations and with alternative sampling protocols (e.g., double observers at the same location) will provide the information needed to estimate both positive and negative correlations in abundance between species. We also anticipate that these new data sets may allow species which are present but unobserved to be included in the model, as in multispecies occupancy models [65]. This extension would be useful for estimating species richness and other community-level measures of biodiversity [66].

## Supporting Information

**Appendix S1 MCMC algorithm used in model fitting.**  
(PDF)

**Appendix S2 Species-specific point counts and measurements of forested area at each sample location (site).** Order of species (columns) is identical to order of rows in Appendix S3.  
(CSV)

**Appendix S3 Morphological measurements, foraging guild, and sound power output of avian species.**  
(PDF)

**Appendix S4 Estimates of model parameters for avian point-count data.**  
(PDF)

## Acknowledgments

We thank Robert Askins for generously allowing us to use his avian point count data to illustrate our modeling approach. We also thank anonymous referees whose comments helped us to clarify the description of modeling assumptions. This work was conducted as a part of the Binary Matrices Working Group at the National Institute for Mathematical and Biological Synthesis. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## Author Contributions

Conceived and designed the experiments: RD EC. Analyzed the data: RD. Contributed reagents/materials/analysis tools: EC. Wrote the paper: RD EC. Obtained permission for use of data: EC.

## References

- Cole LC (1949) The measurement of interspecific association. *Ecology* 30: 411–424.
- Diamond JM (1975) Assembly of species communities. In: Cody ML, Diamond JM, editors, *Ecology and evolution of communities*, Cambridge, Massachusetts: Harvard University Press. pp. 342–444.
- Connor EF, Simberloff D (1979) The assembly of species communities: chance or competition? *Ecology* 60: 1132–1140.
- Gotelli NJ, Graves GR (1996) *Null models in ecology*. Smithsonian Institution Press.
- Gotelli NJ (2000) Null model analysis of species co-occurrence patterns. *Ecology* 81: 2606–2621.
- Colwell RK, Mao CX, Chang J (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* 85: 2717–2727.
- Ovaskainen O, Hottola J, Shitonen J (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal infections. *Ecology* 91: 2514–2521.
- Sebastián-Gonzalez E, Sánchez-Zapata JA, Botella F, Ovaskainen O (2010) Testing the heterospecific attraction hypothesis with time-series data on species co-occurrence. *Proceedings of the Royal Society B* 277: 2983–2990.
- MacKenzie DI, Bailey LL, Nichols JD (2004) Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology* 73: 546–555.
- Dorazio RM, Kéry M, Royle JA, Plattner M (2010) Models for inference in dynamic metacommunity systems. *Ecology* 91: 2466–2475.
- Richmond OMW, Hines JE, Beissinger SR (2010) Two-species occupancy models: a new parameterization applied to co-occurrence of secretive rails. *Ecological Applications* 20: 2036–2046.
- Waddle JH, Dorazio RM, Walls SC, Rice KG, Beauchamp J, et al. (2010) A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications* 20: 1467–1475.
- Clausen HD, Holbeck HB, Reddersen J (2001) Factors influencing abundance of butterflies and burnet moths in the uncultivated habitats of an organic farm in Denmark. *Biological Conservation* 98: 167–178.



14. Graham CH, Blake JG (2001) Influence of patch- and landscape-level factors on bird assemblages in a fragmented tropical landscape. *Ecological Applications* 11: 1709–1721.
15. Carvell C (2002) Habitat use and conservation of bumblebees (*Bombus* spp.) under different grassland management regimes. *Biological Conservation* 103: 33–49.
16. Tworek S (2002) Different bird strategies and their response to habitat changes in an agricultural landscape. *Ecological Research* 17: 339–359.
17. Rickman JK, Connor EF (2003) The effect of urbanization on the quality of remnant habitats for leaf-mining Lepidoptera on *Quercus agrifolia*. *Ecography* 26: 777–787.
18. Collinge SK, Prudic KL, Oliver JC (2003) Effects of local habitat characteristics and landscape context on grassland butterfly diversity. *Conservation Biology* 17: 178–187.
19. Dallimer M, Skinner AMJ, Davies ZG, Armsworth PR, Gaston KJ (2012) Multiple habitat associations: the role of offsite habitat in determining onsite avian density and species richness. *Ecography* 35: 134–145.
20. Gotelli NJ, Ellison AM (2004) A primer of ecological statistics. Sunderland, Massachusetts: Sinauer Associates.
21. Kirk DA, Welsh DA, Baker JA, Thompson ID, Csizy M (2012) Avian assemblages differ between old-growth and mature white pine forests of Ontario, Canada: a role for supercanopy trees? *Avian Conservation and Ecology* 7: 4.
22. Ribeiro DB, Batista R, Prado PI, Brown Jr KS, Freitas AVL (2012) The importance of small scales to the fruit-feeding butterfly assemblages in a fragmented landscape. *Biodiversity and Conservation* 21: 811–827.
23. Warton DI, Wright ST, Wang Y (2012) Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* 3: 89–101.
24. Graves GR, Gotelli NJ (1993) Assembly of avian mixed-species flocks in Amazonia. *Proceedings of the National Academy of Sciences USA* 90: 1388–1391.
25. Hausdorf B, Hennig C (2007) Null model tests of clustering of species, negative co-occurrence patterns and nestedness in meta-communities. *Oikos* 116: 818–828.
26. Lester PJ, Abbott KL, Sarty M, Burns KC (2009) Competitive assembly of south Pacific invasive ant communities. *BMC Ecology* 9: doi:10.1186/1472-6785-9-3.
27. Ulrich W, Gotelli NJ (2010) Null model analysis of species associations using abundance data. *Ecology* 91: 3384–3397.
28. Gotelli NJ, Dorazio RM, Ellison AM, Grossman GD (2010) Detecting temporal trends in species assemblages with bootstrapping procedures and hierarchical models. *Philosophical Transactions of the Royal Society of London Series B* 365: 3621–3631.
29. Fox BJ, Luo J (1996) Estimating competition coefficients from census data: a re-examination of the regression technique. *Oikos* 77: 291–310.
30. Shenbrot G, Krasnov B (2002) Can interaction coefficients be determined from census data? Testing two estimation methods with Negev desert rodents. *Oikos* 99: 47–58.
31. Ritchie EG, Martin JK, Johnson CN, Fox BJ (2009) Separating the influences of environment and species interactions on patterns of distribution and abundance: competition between large herbivores. *Journal of Animal Ecology* 78: 724–731.
32. Taylor LR (1961) Aggregation, variance and mean. *Nature* 189: 732–735.
33. He F, Gaston KJ (2003) Occupancy, spatial variance, and the abundance of species. *American Naturalist* 162: 366–375.
34. Bulmer MG (1974) On fitting the Poisson lognormal distribution to species-abundance data. *Biometrics* 30: 101–110.
35. Engen S, Lande R, Walla T, DeVries PJ (2002) Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. *American Naturalist* 160: 60–73.
36. Aitchison J, Ho CH (1989) The multivariate Poisson-lognormal distribution. *Biometrika* 76: 643–653.
37. Schmidt AM, Rodriguez MA (2011) Modelling multivariate counts varying continuously in space. In: Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, et al., editors, *Bayesian Statistics 9*, Oxford: Oxford University Press. pp. 611–638.
38. Yoccoz NG, Nichols JD, Boulinier T (2001) Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16: 446–453.
39. Chen G, Kéry M, Plattner M, Ma K, Gardner B (2013) Imperfect detection is the rule rather than the exception in plant distribution studies. *Journal of Ecology* 101: 183–191.
40. Williams BK, Nichols JD, Conroy MJ (2002) Analysis and management of animal populations. San Diego, California: Academic Press.
41. Royle JA, Dorazio RM (2008) Hierarchical modeling and inference in ecology. Amsterdam: Academic Press.
42. Royle JA (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60: 108–115.
43. Yamaura Y, Royle JA, Shimada N, Asanuma S, Sato T, et al. (2012) Biodiversity of man-made open habitats in an underused country: a class of multispecies abundance models for count data. *Biodiversity and Conservation* 21: 1365–1380.
44. Chandler RB, King DI, Raudales R, Trubey R, Chandler C, et al. (2013) A small-scale land-sparing approach to conserving biological diversity in tropical agricultural landscapes. *Conservation Biology* 27: 785–795.
45. Barnagaud J, Barbaro L, Papaix J, Deconchat M, Brockerhoff EG (2014) Habitat filtering by landscape and local forest composition in native and exotic New Zealand birds. *Ecology* 95: 78–87.
46. Mutshinda CM, O'Hara RB (2011) Integrating the niche and neutral perspectives on community structure and dynamics. *Oecologia* 166: 241–251.
47. Diggle PJ, Ribeiro Jr PJ (2007) Model-based geostatistics. New York: Springer Science+Business Media.
48. Wikle CK (2010) Hierarchical modeling with spatial data. In: Gelfand AE, Diggle PJ, Fuentes M, Guttorp P, editors, *Handbook of spatial statistics*, Boca Raton, Florida: Chapman & Hall/CRC. pp. 89–106.
49. MacArthur R, Levins R (1967) The limiting similarity, convergence and divergence of coexisting species. *American Naturalist* 101: 377–385.
50. Schoener TW (1974) Resource partitioning in ecological communities. *Science* 185: 27–39.
51. Wisheu IC (1998) How organisms partition habitats: different types of community organization can produce identical patterns. *Oikos* 83: 246–258.
52. MacArthur RH (1958) Population ecology of some warblers of northeastern coniferous forests. *Ecology* 39: 599–619.
53. Simberloff DS (1970) Taxonomic diversity of island biotas. *Evolution* 24: 23–47.
54. Gilpin ME, Diamond JM (1982) Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia* 52: 75–84.
55. Ricklefs RE (2011) Applying a regional community concept to forest birds of eastern North America. *Proceedings of the National Academy of Sciences USA* 108: 2300–2305.
56. Geyer CJ (2011) Introduction to Markov chain Monte Carlo. In: Brooks S, Gelman A, Jones GL, Meng XL, editors, *Handbook of Markov chain Monte Carlo*, Boca Raton, Florida: Chapman & Hall/CRC. pp. 3–48.
57. Gelfand AE, Ghosh SK (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika* 85: 1–11.
58. Askins RA, Philbrick MJ, Sugeno DS (1987) Relationship between the regional abundance of forest and the composition of forest bird communities. *Biological Conservation* 39: 129–152.
59. Dunning JB (1993) CRC handbook of avian body masses. Boca Raton, Florida: CRC Press.
60. Calder WA (1990) The scaling of sound output and territory size: are they matched? *Ecology* 71: 1810–1816.
61. Poole A, editor (2005) *The Birds of North America Online*: <http://bna.birds.cornell.edu/BNA/>. Ithaca, NY: Cornell Laboratory of Ornithology.
62. De Graaf RM, Tilghman NG, Anderson SH (1985) Foraging guilds of North American birds. *Environmental Management* 9: 493–536.
63. Wiley RH, Richards DG (1982) Adaptations for acoustic communication in birds: sound transmission and signal detection. In: Kroodsma DE, Miller EH, Ouellet H, editors, *Acoustic Communication in Birds Volume 1*, New York: Academic Press. pp. 131–181.
64. Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, et al. (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* 88: 15–30.
65. Dorazio RM, Gotelli NJ, Ellison AM (2011) Modern methods of estimating biodiversity from presence-absence surveys. In: Grillo O, Venora G, editors, *Biodiversity Loss in a Changing Planet*, Rijeka, Croatia: InTech. pp. 277–302.
66. Iknayan KJ, Tingley MW, Furnas BJ, Beissinger SR (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology and Evolution* 29: 97–106.