

Model selection and assessment for multi-species occupancy models

KRISTIN M. BROMS,^{1,5} MEVIN B. HOOTEN,^{1,2,3} AND RYAN M. FITZPATRICK⁴

¹*Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, Colorado 80523 USA*

²*Colorado Cooperative Fish and Wildlife Unit, U.S. Geological Survey, Fort Collins, Colorado 80523 USA*

³*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA*

⁴*Aquatic Wildlife Research Group, Colorado Parks and Wildlife, Fort Collins, Colorado 80523 USA*

Abstract. While multi-species occupancy models (MSOMs) are emerging as a popular method for analyzing biodiversity data, formal checking and validation approaches for this class of models have lagged behind. Concurrent with the rise in application of MSOMs among ecologists, a quiet regime shift is occurring in Bayesian statistics where predictive model comparison approaches are experiencing a resurgence. Unlike single-species occupancy models that use integrated likelihoods, MSOMs are usually couched in a Bayesian framework and contain multiple levels. Standard model checking and selection methods are often unreliable in this setting and there is only limited guidance in the ecological literature for this class of models. We examined several different contemporary Bayesian hierarchical approaches for checking and validating MSOMs and applied these methods to a freshwater aquatic study system in Colorado, USA, to better understand the diversity and distributions of plains fishes. Our findings indicated distinct differences among model selection approaches, with cross-validation techniques performing the best in terms of prediction.

Key words: *Bayesian hierarchical models; biodiversity; cross-validation; plains fish; South Platte River Basin; species distribution maps.*

INTRODUCTION

Understanding how to measure biodiversity and what drives it are major quests for ecologists. Multi-species occupancy models (MSOMs) have been gaining in popularity as a means to analyze community-level ecology data while accounting for imperfect detection. They primarily model detection-nondetection data, often referred to as presence-absence data, to quantify species' ranges. MSOMs are fit to data on diverse taxonomic groups such as birds (Dorazio et al. 2006, Kéry and Royle 2009, Russell et al. 2009, Zipkin et al. 2009, Ruiz-Gutiérrez et al. 2010, Carrillo-Rubio et al. 2014), amphibians (Mazerolle et al. 2007, Grant et al. 2013), reef fish (MacNeil et al. 2008), freshwater fish (Holtrop et al. 2010, Kirsch and Peterson 2014, Midway et al. 2014), mammals (Burton et al. 2012), plants (Gelfand et al. 2006), and ants (Dorazio et al. 2011). The output from MSOMs is used to test a variety of hypotheses, such as comparing biodiversity among areas with different management regimes (Zipkin et al. 2010, Carrillo-Rubio et al. 2014), and areas with and without disease outbreaks (Giovanni et al. 2013). Biodiversity metrics (Dorazio et al. 2011, Broms et al. 2015), point-level species richness (Kalies and Rosenstock 2013), metacommunity structure (Mihaljevic et al. 2015), and statistics associated with the differences between species' groupings

(Zipkin et al. 2009, Ruiz-Gutiérrez et al. 2010, Tingley and Beissinger 2013) have also been estimated using the model output. MSOMs may aid landscape-level resource management and monitoring (Noon et al. 2012, Midway et al. 2014). In addition, extensions to the MSOMs have incorporated multiple states (Kéry et al. 2009), dynamic processes (Dorazio et al. 2010), and detection heterogeneity due to abundance (Yamaura et al. 2011, 2012, Tobler et al. 2015).

The MSOM provides inference for occupancy and detection probabilities and parameter estimates for all species in the system. Because the parameters for each of the species arise from community-level distributions, the data are used efficiently and fewer parameters are required than if each species was fit independently. This “borrowing strength” technique is an advantage of random effects models (Gelman and Hill 2007, Hobbs and Hooten 2015), and it has been an allure of Bayesian models since Tukey coined the phrase in 1963 (Scott and Smith 1969).

In addition, occupancy models allow for heterogeneous measurement error that accounts for species-specific, site-specific, and survey-specific differences in detection. Accounting for this type of measurement error leads to more accurate predictions than models that ignore imperfect detection (MacKenzie et al. 2006, Broms et al. 2015).

MSOMs require an understanding of Bayesian models and Bayesian software (e.g., BUGS or JAGS; Iknayan et al. 2014). In addition, model checking and model validation for Bayesian hierarchical models remain areas

Manuscript received 7 August 2015; revised 19 December 2015; accepted 5 January 2016; final version received 22 February 2016. Corresponding Editor: E. G. Cooch.

⁵E-mail: kristin.broms@rams.colostate.edu

of active research, thereby adding further complexity to the implementation of MSOMs (Hooten and Hobbs 2015). In what follows, we explain model checking procedures for MSOMs and provide suggestions for model selection in these settings. As a motivating example throughout, we analyze multi-species freshwater fish data collected from 113 sites within the South Platte River Basin in Colorado. The system has approximately 26 native species, of which 24 were detected.

THE MULTI-SPECIES OCCUPANCY MODEL

Two primary versions of the MSOM exist: one with known species richness and one with unknown species richness. First we describe the model with known species richness and then we add a hierarchical level, resulting in a new model with unknown species richness.

Known species richness

In well-studied systems, the species richness of the study area may be known. For example, in our study area, there are 26 native species (Propst and Carlson 1986, Nesler et al. 1997). For this system, we can model the native species assemblage using the model with known species richness.

Let y_{ijk} represent the detection of species i , $i = 1, \dots, N$, at site j , $j = 1, \dots, J$, on survey (visit) k , $k = 1, \dots, K_j$; N is the number of species in the community, also known as the species richness or gamma diversity, J is the number of sites that were sampled, and K_j is the number of sampling occasions or surveys of site j . The data point y_{ijk} equals 1 if the species truly occurs at the site and was detected, and is zero otherwise. The true occupancy of species i at site j , z_{ij} , is a latent indicator variable. If the species was detected at the site, then we know that it occurs at the site and $z_{ij} = 1$; otherwise its value is inferred from fitting the model. The probability of species i occurring at site j is ψ_{ij} , which can be a function of covariates. The probability of detecting the species is p_{ijk} , which also may be a function of one or more covariates. Assuming no false positives (i.e., that a species never gets recorded as present when it was absent) then the complete model statement is

$$y_{ijk} \sim \text{Bernoulli}(p_{ijk} \cdot z_{ij}) \tag{1}$$

$$\text{logit}(p_{ijk}) = \mathbf{v}'_{jk} \boldsymbol{\alpha}_i \tag{2}$$

$$z_{ij} \sim \text{Bernoulli}(\psi_{ij}) \tag{3}$$

$$\text{logit}(\psi_{ij}) = \mathbf{x}'_j \boldsymbol{\beta}_i \tag{4}$$

$$\boldsymbol{\alpha}_i \sim N(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha) \tag{5}$$

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \tag{6}$$

where the \mathbf{v}_{jk} are the covariates pertaining to survey k of site j that affect detection probabilities and the \mathbf{x}_j are the covariates pertaining to site j that affect occupancy probabilities.

For the multi-species occupancy model, we do not independently estimate each coefficient for each species. Instead they are influenced by the other species through a community-level set of parameters ($\boldsymbol{\mu}_\alpha, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\alpha$, and $\boldsymbol{\Sigma}_\beta$). Allowing all species' coefficients to arise from common distributions leads to more powerful models but still allows for the same range of detection probabilities and/or occupancy probabilities as single-species occupancy models (SSOMs). The estimates of the coefficients may now be shifted toward the community-level means, which implies that the models may understate how strongly a covariate affects the response variable. As we discuss in PRIORS and demonstrate through our example, we believe this type of borrowing of strength can be an asset and that the trade-off of slightly biased estimates for greater precision is a worthwhile compromise.

To complete the MSOM, we used the following weakly informative hyperpriors

$$\boldsymbol{\mu}_\alpha \sim N(\mathbf{0}, 2.25^2 \mathbf{I}) \tag{7}$$

$$\boldsymbol{\Sigma}_\alpha \equiv \sigma_\alpha^2 \mathbf{I} \tag{8}$$

$$\sigma_\alpha \sim \text{half-Cauchy}(2.25^2) \tag{9}$$

$$\boldsymbol{\mu}_\beta \sim N(\mathbf{0}, 2.25^2 \mathbf{I}) \tag{10}$$

$$\boldsymbol{\Sigma}_\beta \equiv \sigma_\beta^2 \mathbf{I} \tag{11}$$

$$\sigma_\beta \sim \text{half-Cauchy}(2.25^2). \tag{12}$$

The covariates should be standardized so that they have a mean equal to zero and variance equal to one. Standardizing allows for consistent default priors for all parameters (such as those specified above), allows for comparisons among covariate effects, and alleviates computational challenges.

Unknown species richness

In many community-level studies, the total species richness is unknown and the researcher would like to derive this and other biodiversity metrics for the study area. In this scenario, the researcher would include and estimate the number of species that were never detected. If species richness is unknown, one augments the model with a Bernoulli variable, w_i , that indicates whether or not a species exists in the assemblage of interest

$$w_i \sim \text{Bernoulli}(\pi). \quad (13)$$

If species i was ever detected during the study, then $w_i = 1$; otherwise it is inferred from the model similar to how the unknown, true occurrences, z_{ij} , are estimated. Species richness is then a derived quantity, $N = \sum_{i=1}^M w_i$, where M is an upper limit for species richness that is specified a priori. In the Bayesian framework, a prior for π is also included in the model; generally $\pi \sim \text{Beta}(\alpha_\pi, \beta_\pi)$ is used. The uniform distribution is often used for the prior distribution in this model ($\alpha_\pi = \beta_\pi = 1$, e.g., Dorazio et al. 2006, Zipkin et al. 2010). However, Link (2013) suggested that $\pi \sim \text{Beta}(0.001, 1)$, is a better choice.

If the estimate of species richness is near M , even when M is unreasonably large for the assemblage, it may be evidence of an ill-posed model. If many covariates are included in the model or there are many species that were rarely detected, the number of unknown parameters increases and the Markov chain Monte Carlo (MCMC) algorithm may have trouble converging.

In addition to species richness, other community and biodiversity metrics such as Shannon's diversity index, Hill numbers, and turnover rates have been derived using the MSOM with unknown species richness (e.g., Tingley and Beissinger 2013, Iknayan et al. 2014, Broms et al. 2015). However if a heterogeneous landscape was sampled, the addition of covariates in the occupancy probability formula (Eq. 4) may make community-level biodiversity metrics difficult to calculate.

PRIORS

Priors are an important, but often overlooked, component in Bayesian models. For example, the prior for a logit-scale parameter (i.e., a coefficient in an occupancy model) should have low probabilities outside of $(-5, 5)$ because adding 5 on the logit-scale is equivalent to shifting the predicted probability from 0.01 to 0.50, or 0.50 to 0.99 (Gelman et al. 2014a). Alternatively, if one computes the inverse-logit of a normal variable with $\sigma^2 = 2.25^2$, it is relatively flat for most of $(0,1)$ for the associated probability; the inverse logit transformation of a variable with $\sigma^2 = 10^2$ leads to a U-shaped prior distribution for the associated probability, with most of its density on values close to 0 or 1. To account for these realistic, a priori restrictions, Gelman et al. (2008) recommend a Cauchy distribution with scale $\sigma = 2.5$ for logistic coefficients and $\sigma = 10$ for the intercept, or a scaled t -distribution with scale $\sigma = 2.5$ (and $\sigma = 10$ for the intercept) and seven degrees of freedom, $\nu = 7$. (Note that a t -distribution with one degree of freedom, $\nu = 1$, results in a Cauchy distribution, and a t -distribution with infinite degrees of freedom, $\nu = \infty$, results in a normal distribution.) Gelman et al. (2008) recommend a wider prior (i.e., one with a higher variance) for the intercept because they are often working with data where the mean probability is very close to 0–1. They recommend a scaled

t -distribution because it has wider tails than a normal distribution and is therefore more robust to the occasional large value (Gelman et al. 2008, 2014a). Dorazio et al. (2011) used $\mu_\beta \sim t(0, 1.566\mathbf{I}, \nu = 7.763)$ priors, which are narrower than those recommended by Gelman et al. (2008), to encourage a preference for values < 5 . Finding little empirical difference among these options, we used normal priors (Eqs. 7 and 10), because the associated algorithms were more stable. A weakly informative prior is a proper prior that is specified to provide some information based on the realities of the data and model, but provides less information than the actual knowledge that is available (Gelman et al. 2014a).

For variance parameters in process models, Gelman et al. (2014a) recommend a uniform distribution as a vague, (i.e., diffuse or flat prior) prior for standard deviations. They prefer a uniform over a gamma distribution because it may be less sensitive to the choice of prior parameters, which are the shape and scale parameters for the gamma distribution and the upper bound associated with the uniform distribution. Gelman et al. (2014a) also recommend a half-Cauchy distribution (which is a Cauchy distribution truncated below by 0) for the standard deviation because it is a weakly informative prior that may help with convergence in small sample size situations. Community data used with MSOMs are often sparse, thus we used half-Cauchy distributions but with $\sigma = 2.25$ to reflect the same range for the standard deviations that the mean parameters have on the logit-scale.

Sensitivity to priors was tested through comparisons of model results under various prior parameterizations; we found that the normal and half-Cauchy priors led to better convergence of our algorithms than the t -distribution or uniform distribution, respectively. Across all priors, posterior distributions were similar, although the weakly informative priors led to slightly more shrinkage.

Priors for shrinkage

Our selection of weakly informative priors is a form of ridge regression, which relates to penalized likelihood methods (Hooten and Hobbs 2015, Hutchinson et al. 2015); all represent types of shrinkage, alternatively called regularization. Regularization, its relationship to model selection, and its role in the occupancy model framework are emerging topics in statistical ecology (Moreno and Lele 2010, Hooten and Hobbs 2015, Hutchinson et al. 2015). While the concept of regularization has existed for decades within the statistics community, it is slowly being adapted as a method to improve ecological models. For the MSOM model, the weakly informative priors lead to estimates that are closer to 0 than the maximum likelihood estimates (MLEs). This regularization may improve computational stability and lead to more reasonable parameter estimates than the MLEs, as we demonstrate with the South Platte data in

what follows. The regularization may also help with collinearity among covariates, over-parameterization, and the separation that can occur when a covariate perfectly predicts the outcomes associated with binary data (Gelman et al. 2008, 2014a, Hefley and Hooten 2015).

MODEL CHECKING AND VALIDATION

Goodness-of-fit evaluates model fit against the within-sample data; it is one form of model assessment, which quantifies how consistent the data are with the chosen model and whether the model assumptions are appropriate. Within-sample data refer to the data that are used to fit the model. In ecological studies, it is common to treat all the data as within-sample data. Model selection is the process of selecting a best-fitting model from a set of models; it may be done using within-sample data as with AIC or with out-of-sample data, using predictive performance scoring rules. Out-of-sample data refer to a subset of the data that is withheld when fitting a model and is instead used afterward to assess how well the model predicts. Predictive performance measures a model's ability to predict out-of-sample data and is important for model assessment.

Goodness-of-fit

Goodness-of-fit tests compare the observed response variables to the values predicted by the model. Evaluating goodness-of-fit is difficult for occupancy models because we would like to know how well we are estimating the true occurrences, z_{ij} . For most of our sites, z_{ij} is unknown and we have no true values against which to compare our estimates. Instead we compare our detections, y_{ijk} , to the detections predicted by the model, \tilde{y}_{ijk} . Another difficulty with occupancy models is that the goodness-of-fit tests commonly used for generalized linear models, based either upon Pearson residuals or deviance residuals, do not follow a χ^2 -distribution when calculated for binary data and therefore cannot be used to assess model fit without an adjustment (McCullagh and Nelder 1989, Kéry and Schaub 2012). Typically, one bins or groups the data, either by a categorical covariate or based on the predicted probabilities, so that the response is binomial rather than binary and then one can use the χ^2 -statistic (Gelman and Hill 2007, Hosmer et al. 2013). In the occupancy model framework, one might bin their data by summing over the surveys for each site. However, a statistic based on these sums will still not follow a χ^2 distribution because of the latent occupancy process. White et al. (1999) and MacKenzie and Bailey (2004) recommend using parametric bootstrapping for capture–recapture and occupancy models, respectively, to create a new distribution for the sum of squared residuals for the given data set, and then check if the observed test statistic value falls near the middle of the bootstrapped distribution.

We derived a goodness-of-fit test statistic (i.e., a Bayesian P -value) for the MSOM that is similar to the parametric bootstrapping used by White et al. (1999) and MacKenzie and Bailey (2004). The primary difference between the test statistics is that our P -value is naturally Bayesian while theirs is non-Bayesian. Our P -value, also known as a posterior predictive check, followed the same process as Carrillo-Rubio et al. (2014), Kroll et al. (2014), and Tobler et al. (2015), but was based on the deviance residuals rather than the Pearson's residuals because of its relationship to information criterion theory (Spiegelhalter et al. 1998). We describe the approach in detail in Appendix S1.

In addition to calculating the Bayesian P -values, we looked at diagnostic plots to visually examine the model fits. We plotted the deviance residuals for each species and site and then plotted the residuals against the covariates to highlight outliers and areas of lack of fit.

Model selection

If the goal is prediction, then one could avoid model selection by including all the covariates in the model and using regularization to restrict the parameter space through strong priors (e.g., Gelman and Hill 2007:67, Gelman et al. 2014a:367, Barker and Link 2015, Hooten and Hobbs 2015). However in ecology, often the goal of an analysis is to infer how the covariates relate to the response variables and thus we selected among competing models. We implemented and compared several model selection options because there is no consensus on an optimal model selection criterion for Bayesian occupancy models.

Within-sample model selection.—Out-of-sample validation and k -fold cross-validation, described in the *Predictive Performance*, are useful for choosing a best-predicting model (Hooten and Hobbs 2015). However, these options may be too time-consuming or may not be available in sparse data situations (Gelman et al. 2014b). Therefore it is helpful to have a selection criterion that involves fitting each model only once. Criteria based on information theory use in-sample calculations to estimate the expected out-of-sample prediction error (Gelman et al. 2014b). Unfortunately, the two most common methods of this type of model selection, AIC (Akaike information criterion; Akaike 1973) and DIC (deviance information criterion; Spiegelhalter et al. 2002), are not ideal for Bayesian hierarchical occupancy models because of the models' latent parameters (Hooten and Hobbs 2015), Gelfand et al. (2006) point out that DIC is also sensitive to parameterization.

WAIC (Watanabe-Akaike information criterion; Watanabe 2010) has recently been promoted as a within-sample model selection criterion that is suitable for Bayesian hierarchical models as long as the data are conditionally independent (Gelman et al. 2014b, Hooten and Hobbs 2015). Following these suggestions, we used

WAIC to compare models for the South Platte fish data. The conditional predictive ordinate criterion (CPO) has been touted as another potential model selection criterion (Petit 1990, Geisser 1993, Held et al. 2010), and therefore we also compared models using this criterion. Their formulas are provided in Appendix S1.

Predictive performance.—To select a model based on its true predictive performance, one needs out-of-sample data. To use out-of-sample data, one can collect more data or completely withhold a subset of their data, then fit the models to the within-sample data and compare score statistics based on the hold-out data to pick a best model. Alternatively, one can perform k -fold cross-validation. In cross-validation, one splits their data into k^* sets. (We use k^* to represent the cross-validation folds so as not to confuse it with the k surveys that are conducted at each site.) Withholding one of the k^* subsets, one fits the models to the rest of the data, and calculates a scoring rule to measure how close the models' predictions are to the hold-out data. This process is repeated for each subset of data, then the scoring function is summed or averaged over the k^* folds. The model with the best score is identified as having the best predictive performance. For most scoring rules, low values indicate best performance, but for some, such as AUC, high values indicate the best performance. After the final model is chosen, it is refit using the entire data set so that inference is based on all the data.

There are many options for the scoring rule. One popular choice for binary data models is AUC, area under the receiver operator characteristic curve (Hanley and McNeil 1982). AUC is a rank-based rule, similar to the Mann-Whitney test, that gives a high score when the predicted probabilities related to the successes ($y_i = 1$) are consistently higher than the probabilities associated with the failures ($y_i = 0$); it does not matter how far these probabilities are from 0 and 1, only their relative rankings matter. AUC has been used to compare and pick models in the species distribution modeling literature (Elith and Leathwick 2007, Elith and Graham 2009) and is popular in the machine learning community (Hastie et al. 2009). Within the MSOM literature, AUC has been used by Zipkin et al. (2012) and Mattsson et al. (2013), although they did not use AUC in the context of cross-validation. Despite being a popular scoring rule, it is not local and proper (Hernández-Orallo et al. 2012). A scoring rule is proper if its expected value is maximized by the true, generating model, and is local if bad predictions lead to worse scores (Vehtari and Ojanen 2012). Therefore, we also compared models using three other scoring rules that are proper: the Brier, Logarithmic (to be consistent with rest of paper), and 0–1 scores (Gneiting and Raftery 2007; formulas in Appendix S1). Brier and 0–1 scores have been traditionally used in binary data models such as logistic regression, but may not have the same properties for occupancy models. Thus, we calculated these scores mainly to compare and contrast with the other scoring rules.

EXAMPLE: SOUTH PLATTE RIVER BASIN

We applied MSOMs to the fish populations of the warm-water reaches of the South Platte River Basin in Colorado (Fig. 1). Within our study region, there are 26 native species of which 24 were detected, plus at least 21 introduced species. Of the 26 native species, eight are currently listed as special status by the State of Colorado. These include the endangered lake chub (*Couesius plumbeus*), plains minnow (*Hybognathus placitus*), suckermouth minnow (*Hybognathus mirabilis*), and northern redbelly dace (*Chrosomus eos*); the threatened brassy minnow (*Hybognathus hankinsoni*) and common shiner (*Luxilus cornutus*); and the species of special concern Iowa darter (*Etheostoma exile*) and stonecat (*Noturus flavus*). The reduction in population size of these native fishes may be due to anthropogenic changes including stream barriers, altered flow regime, siltation, channelization, changes in water quality, or introduced species (Fausch and Bestgen 1997, Perkin et al. 2015). Understanding the distributions of native plains fish species is essential to promoting conservation and potential expansion of remaining populations.

The South Platte Basin is fairly homogeneous within our study region. High gradient, mountainous streams from the Rocky Mountains flow east into the lower gradient transition and high plains zone streams where our study took place (Fig. 1). Occupancy probabilities were functions of the site-level covariates. We included elevation as a covariate in our models (ELEV; data *available online*)⁶ because of its potential role in ecological processes and because it was highly correlated with reach catchment area ($\rho = -0.74$), latitude ($\rho = -0.68$), and longitude ($\rho = -0.83$). The reach catchment area is the accumulated area that contributes to a stream segment's flow and is a continuous-valued version of stream order. We used an indicator variable to represent whether the stream segment was in the transition zone (WEST = 1) or high plains zone (WEST = 0). Because much of the state's urban development (including the metropolitan area of Denver) occurs along the Front Range corridor where the transition zone becomes the high plains zone, and that development is likely to affect fish distributions, we also included an interaction between elevation and the transition zone indicator variable (ELEV·WEST). The other site-level covariates were indicator variables representing whether a stream segment was part of the perennial, connected stream network (POND = 0), whether it was a pond (POND-Y = 1), or whether it was an intermittent stream with large areas of dry land between the site and the connected stream network (POND-Int = 1), and land cover covariates that were thought to influence fish distributions: the percent of land area that was cropland within a 2000-m buffer (CROPS), the percent of land area that was urban development within a 1000-m buffer (DVLDPD), and the percent of land area that was wetlands within a 1000-m buffer (WTLNDS; Jin et al. 2013).

⁶ <http://ned.usgs.gov>.

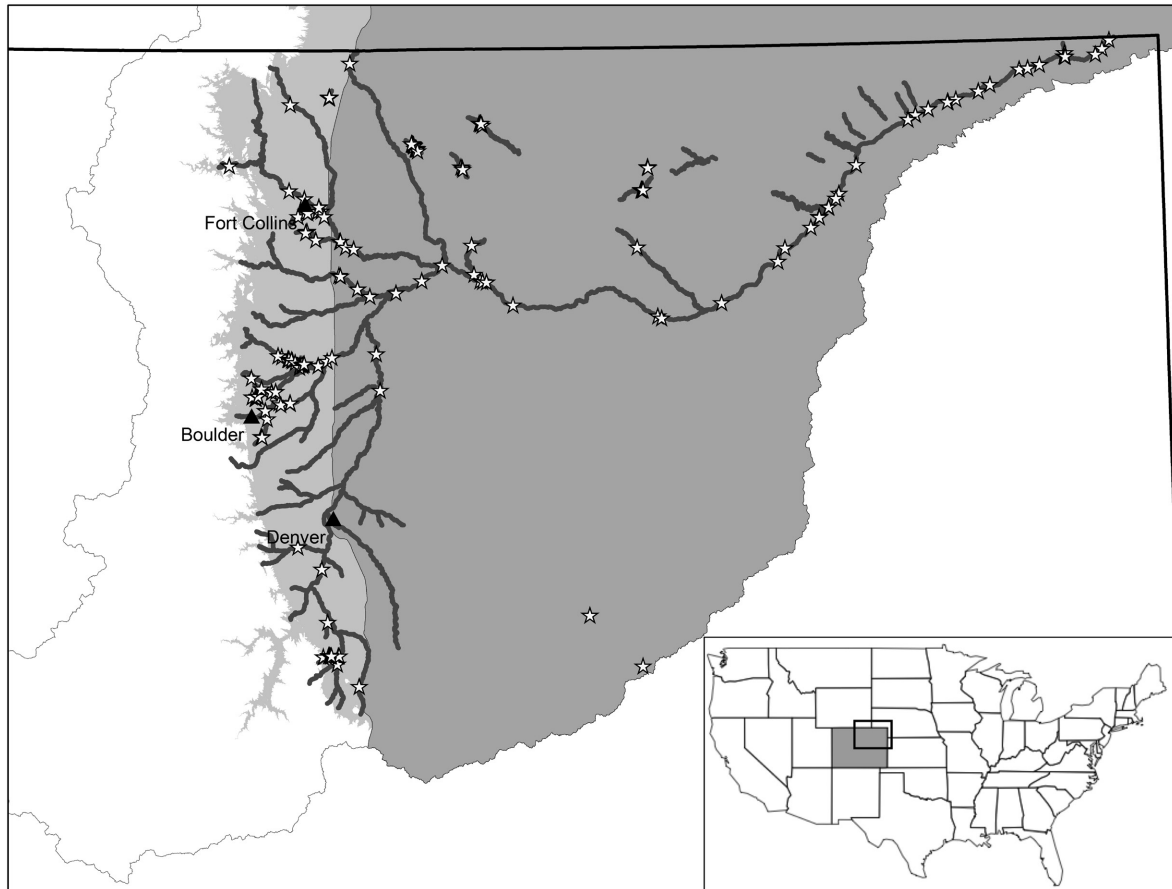


FIG. 1. Our study area within the South Platte River Basin, USA. The sites where sampling took place are represented by stars. The gray edging outlines the South Platte River Basin and the black edging is the Colorado State boundary. The white polygon of the river basin is the headwaters zone and was excluded from our sampling and analyses. The light gray polygon represents the transition zone and the darker gray area represents the high plains zone of the basin. Stream segments that are not connected to the main stem of the South Platte were considered to be intermittent sites. Ponds do not have length and therefore are represented on the map as the starred points not on a stream segment.

Detection probabilities were modeled as functions of the covariates described above and the following survey-level covariates: year as a factor (y2009, y2010, y2011, y2012, y2013) to account for the fact that the fish distributions and abundances may change from year to year due to natural fluctuations and unrecorded changes in flow; day-of-year (YDAY) and day-of-year squared (YDAY²) to account for potential within-season fluctuations; an indicator variable (SEINE) that equals 0 if the survey method was electrofishing and equals 1 if the survey method was seining; and a variable of the logarithm of the maximum count (MAX_CT) associated with the sampling occasion.

Our data set consisted of 113 sites each surveyed 1–24 times, with a median of 4 surveys per site. The original data consisted of counts of each species of fish. The counts were highly variable and not amenable to an N-mixture model (Royle 2004) so we aggregated them into the detection-nondetection data. All continuous variables were scaled to have a mean of 0 and a standard

deviation of 1. A full description of the South Platte River Basin and the sampling details as related to our analysis may be found in Broms et al. (*unpublished data*).

We fit the MSOM with known species richness to the native fish populations of the stream network (fish species are listed in Appendix S2). We included the two undetected species by augmenting our detection array with two rows of zero detections. All model selection and assessment procedures were applied to this model. We fit five models that included different covariates based on previous single-species results (Table 1; Broms et al. *unpublished data*). One model included all possible covariates (FULL; eight variables to model occupancy and seventeen to model detection probabilities), one model included only those covariates that were found to significantly correlate with the five representative species in Broms et al. (*unpublished data*) (LIMITED; one variable to model occupancy and six to model detection probabilities), and one model included only those covariates that were found to significantly correlate with at least

two of the representative species (MIDDLE-1; one variable to model occupancy and twelve to model detection probabilities. We fit a fourth model that included all covariates except for the land cover variables (MIDDLE-2; five variables to model occupancy and fourteen to model detection probabilities). To assess whether the model selection criteria could discriminate against superfluous variables, we fit a fifth model (EXTRA) that was the same as the LIMITED model but had an additional, randomly generated standard normal variable, to model occupancy probabilities (Table 1).

We used MCMC as implemented in JAGS v3.4.0 (Plummer 2003) to fit the models. We obtained three chains with different starting values for each model for 150 000 iterations with a burn-in of 50,000 iterations and a thinning of 100, leaving a total of 3,000 samples. In some settings, thinning does not help with convergence and could be inefficient (Link and Eaton 2012), but we thinned because of data storage limits. We assessed convergence through visual examination of the trace, density, and autocorrelation plots. We performed a five-fold cross-validation on each of the five models and calculated AUC, Brier, Logarithmic, and 0–1 scoring rules. We assessed the model fit for all five models but only present the diagnostics for the model that was picked as best-predicting. We calculated the *P*-value associated with the model, plotted the deviance residuals for each species and for each site, plotted the deviance residuals against each covariate, and created predicted occupancy maps.

Finally, we fit two additional MSOMs with unknown species richness to better understand how inference would change under the model alteration. These models included the same covariates as the previously chosen, best-predicting model but were fitted to different data

sets. One model was fit to the native fish data, plus 50 augmented rows of zero detections to allow for the undetected species, and one model was fit to all of the fish detection data, both native and introduced, plus the 50 augmented rows.

RESULTS

The within-sample model selection criteria, WAIC and CPO, followed similar patterns and ranked the models based on more parsimonious models given larger, and therefore worse, scores (Table 2). WAIC was able to discriminate against the model with the superfluous variable, while the CPO criteria gave the model with the superfluous variable (EXTRA model) a slightly better ranking than the same model without it (LIMITED model, Table 2). The 0–1 score also suggested the FULL model, but the other cross-validation scores ranked the MIDDLE-2 model as best-predicting (Table 2). The 0–1 score was lowest for the FULL model for each hold-out data set, but the Brier score, Log score, and AUC suggested different models for each hold-out data set, usually agreeing with each other but sometimes indicating different models (Appendix S3: Table S1). All scores were able to discriminate against the model with the superfluous variable.

Based on the cross-validation scores, we identified the MIDDLE-2 model, the model without the land cover variables, as best-predicting. The posterior predictive *P*-value associated with this model was 0.94, suggesting possible lack of fit. However, the *P*-value associated with the LIMITED model was 0.0003 and the *P*-value associated with the FULL model was 0.993, indicating that the statistic is sensitive to model choice and one should investigate potential causes of lack of fit further if using certain models for inference. Neither the plots of the deviance residuals versus the covariates nor the plots of the residuals associated with each species revealed where the lack of fit occurred (Appendix S3: Fig. S1). Boxplots of the residuals associated with each site highlighted some areas for improvement in the fit (Appendix S3: Fig. S2). Locations with large deviances (sites 7, 36, 89, 102, and 154, Appendix S3: Fig. S2) were ponds, intermittent streams, or backwaters associated with a larger stream. The larger deviances were therefore expected; fewer ponds and intermittent streams were surveyed and the backwaters were not given a separate category within the model although they are expected to provide different habitat and have different assemblages than the main stream segments.

The plots of the predicted occupancy probabilities were more informative (Fig. 2, all species in Appendix S4). In general, the predicted occupancies matched expert opinion with only a few exceptions. The Green Sunfish (*Lepomis cyanellus*, SNF) is not especially common or widespread in the higher elevation, transition zone although the model has fairly high predictions for that region. Other minor exceptions are that the common shiner (CSH) was historically in the northwest part

TABLE 1. The five models fit to the South Platte fish data and the covariates that they included in the occupancy and detection probability functions.

Model	Occupancy variables	Detection variables
LIMITED	ELEV	YEAR, ELEV
MIDDLE-1	ELEV	YEAR, ELEV, POND, MAX_CT, YDAY, YDAY2, CROPS
MIDDLE-2	ELEV, WEST, POND, ELEV, WEST	YEAR, ELEV, WEST, POND, ELEV, WEST, MAX_CT, YDAY, YDAY2, SEINE
FULL	ELEV, WEST, POND, ELEV, WEST, CROPS, DVLDP, WTLNDS	YEAR, ELEV, WEST, POND, ELEV, WEST, MAX_CT, YDAY, YDAY2, SEINE, CROPS, DVLDP, WTLNDS
EXTRA	ELEV, RANDOM	YEAR, ELEV

Note: See Example: South Platte River Basin for definition of variables.

TABLE 2. Model selection values for the five models that we fit to the South Platte fish data. Deviance, WAIC, and CPO are within-sample statistics, calculated from fitting the models to the full data set. The 0–1 score, Log score, AUC, and Brier score are the averaged values from the 5-fold cross-validation statistics. The model that each statistic suggests as best-fitting or predicting is in boldface.

Model	Deviance	WAIC	CPO	0–1 score	Log score	AUC	Brier
LIMITED	7471.1	7780.6	3906.0	807.87	1507.8	0.884	309.63
MIDDLE-1	6904.7	7301.6	3676.6	746.32	1489.2	0.896	295.31
MIDDLE-2	6378.4	6863.5	3470.5	682.32	1436.5	0.908	274.89
FULL	6179.4	6728.4	3405.7	661.21	1602.3	0.877	324.41
EXTRA	7469.2	7782.9	3905.9	807.60	1508.9	0.884	311.33

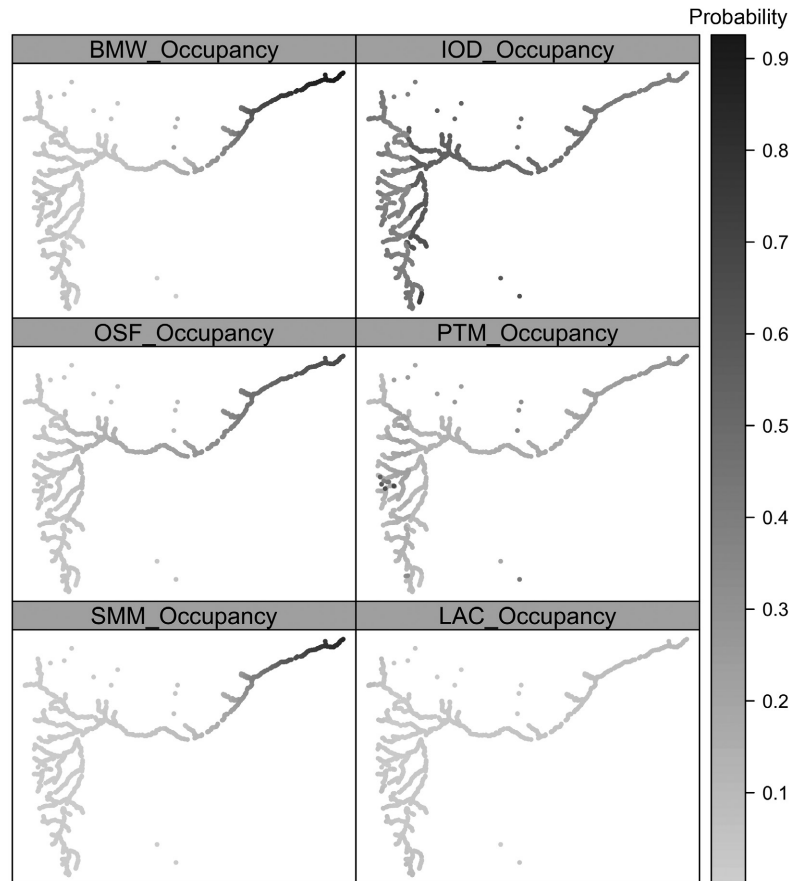


FIG. 2. Predicted occupancy probabilities for five representative species, the brassy minnow (BMW), Iowa darter (IOD), orangespotted sunfish (OSF), plains topminnow (PTM), and suckermouth minnow (SMM), plus one of the species that went undetected during the study, the lake chub (LAC). Scientific and common names for each species may be found in Appendix S2. Colored maps of the predicted occupancy and detection probabilities for all native species in the basin may be found in Appendix S4.

of the basin although now it is only found in the west and southwest; the Orangespotted Sunfish (*Lepomis humilis*, OSF) was historically common near the border of the transition and high plains zones but the model did not predict them to occur there; and the Stonecat (STP) has only been found near Longmont, Colorado, along the Front Range corridor, but based on its life history traits, one would expect the stonecat to also occur in the lower basin as predicted by the model.

Figure 2 shows the predicted occupancy maps for five species and for one of the two native species that went undetected during the data collection, the lake chub. Tables of the associated coefficient values from the MSOM fits and from previous SSOM fits are provided in Appendix S4. The standard deviations were still high relative to the mean values for many coefficients but had realistic ranges. The lake chub (LAC) went undetected during the surveys. It was predicted to have very low

occupancy probabilities throughout the basin but with slightly higher occupancy probabilities at higher elevations and near the border with Nebraska (Fig. 2). The lake chub is a glacial relict species that should be found in only a few locations at high elevations but not lower in the basin, while the plains minnow, which also went undetected during the surveys, is only found lower in the basin. The model was unable to differentiate the distributions of the two undetected species, but it did correctly predict where either of them may occur.

Summaries of the posterior distributions for the mean and standard deviations associated with each coefficient are provided in Appendix S3: Tables S2 and S3. The means and standard deviations associated with the Pond indicator variable had wide credible intervals for both the occupancy and detection components of the model; this is due to the fact that only seven of our sites were Ponds and about half of the native species were never detected in any of the ponds. The credible intervals associated with the standard deviations of the elevation-transition zone interaction were also wide; suggesting high variability among species or that there were few detections to inform this variable (Appendix S3: Table S3). The means and standard deviations associated with the years had small values, implying that on average, detection probabilities changed very little from year to year.

The species richness estimated by the MSOM with unknown species richness had a median value of 26 species (95% credible interval: 24–31), which was the actual value, and inference on the coefficients was nearly identical to the MSOM with known species richness. The MSOM with unknown species richness fitted to all detection data, including both the native and introduced species always predicted species richness at M , its maximum allowable value, and therefore we did not draw any inference from that model.

DISCUSSION

The SSOM is often fit to ecological, presence-absence data, and is more commonly used than the MSOM. The benefits of the SSOM are that they can quickly be fit using maximum likelihood in either MARK (White and Burnham 1999), PRESENCE (MacKenzie et al. 2002) or the “unmarked” package in R (Fiske and Chandler 2011), and then a best-predicting model can be selected via AIC, or models can be averaged (but see Cade 2015). However, it is advantageous to fit MSOMs, when applicable, to analyze multi-species data. For example, our use of the MSOM led to community-level inference and parameter estimates for all species in the community, including rarer species, objectives which were not obtainable through SSOM analyses.

Two main versions of the MSOM exist and they can be used to obtain inference for different ecological questions. By fitting a model with unknown species richness to homogeneous areas, one can calculate and

compare biodiversity metrics in pursuit of answering macroecology questions. Biodiversity metrics are traditionally calculated from count data, but there is a growing body of literature that computes analogous metrics from detection-nondetection data for when count data is unavailable or unreliable (Dorazio et al. 2011, Chao et al. 2014, Iknayan et al. 2014, Broms et al. 2015). Fitting the MSOM using a known species richness, as we did here, can also be valuable. One can fit the model to data from a heterogeneous region to infer how different species respond to the characteristics of their environment by examining the covariate estimates. In this way, the MSOM can be used for the same purposes as a SSOM but by efficiently using the data, one gains inference for all species of the system.

Another advantage of the model is that the Bayesian framework can help with sparse data situations. Multi-species data are likely to be sparse because there are often several species that were not detected or rarely detected. One way to improve the estimates for small sample sizes is through a penalized likelihood, where a small amount of bias is introduced to greatly reduce the variances (e.g., Moreno and Lele 2010, Hutchinson et al. 2015). In a similar vein, we used weakly informative priors as a regularization on the model. The weakly informative priors helped with collinearity between covariates and separation related to indicator variables. For example, few surveys were conducted in 2011 and the Brassy Minnow was detected at only one site that year. Because of the small sample size, the SSOM standard error associated with the related coefficient was >1000 (Appendix S4: Table S1), a value much higher than could reasonably be expected. The weakly informative priors associated with the MSOM shrank the standard error to the range we expect for scaled variables on the logit-scale. The MSOM also allowed us to include all covariates in fitting our models; the SSOMs were over-parameterized for some species and led to poor estimation for some covariates.

Choice of priors, goodness-of-fit tests, and model selection remain areas of active research for Bayesian statisticians and occupancy modelers, and there is no consensus among statisticians on many of these topics (Hooten and Hobbs 2015). To assess goodness-of-fit, we used Bayesian P -values and diagnostic plots of the goodness-of-fit residuals for model checking. In our application, the plots had larger deviance residuals for sites with uncommon habitat variable combinations and were therefore valuable in highlighting sites with sparse data. We recommend that practitioners perform model checking but believe there is room for improvement in finding additional, appropriate methods. For example, the visual diagnostics could be improved upon by using Albert and Chib's residuals and plotting their posterior distributions (Albert and Chib 1995) rather than plotting the mean values of the residuals.

For model selection, we used cross-validation with Log scores to select a best-fitting model based on its

predictive performance. Cross-validation is already a popular model selection criterion in machine learning (Hastie et al. 2009), and its popularity is increasing in ecology. Within-sample criteria WAIC and CPO have been proposed for model selection for Bayesian hierarchical models. These criteria are similar in spirit to AIC. Our analyses suggested that they may be susceptible to over-fitting, particularly with the detection probability covariates. Because the information criteria treated each data point equally, each survey contributed equally to the likelihood, thereby causing a few sites with extra surveys to be influential. The form of cross-validation we used reduced the influence of outlier sites because it withheld data by site rather than survey. Finally, we note that other options for the model and variable selection exist. For example, indicator variable selection (e.g., Dorazio et al. 2011), reversible-jump MCMC, and Bayesian model averaging may be used for Bayesian multimodel inference although we did not investigate these techniques.

Ecosystem-based management often depends on taking a multi-species approach to set policy standards. MSOMs are poised to be a crucial tool for biological conservation administered through this type of management, and therefore it is important for ecologists to be familiar with their methods, comfortable implementing the models, and confident in the results that the models infer.

ACKNOWLEDGMENTS

Funding for this research was provided by Colorado Parks and Wildlife. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

LITERATURE CITED

- Akaike, H. 1973. Second International Symposium on Information Theory, Chapter information theory as an extension of the maximum likelihood principle. pages 267–281. Akademiai Kiado, Budapest, Hungary.
- Albert, J., and S. Chib. 1995. Bayesian residual analysis for binary response regression models. *Biometrika* 82:747–759.
- Barker, R. J., and W. A. Link. 2015. Truth, models, model sets, AIC, and multimodel inference: a Bayesian perspective. *The Journal of Wildlife Management* 79:730–738. Doi: 10.1002/jwmg.890.
- Broms, K. M., M. B. Hooten, and R. M. Fitzpatrick. 2015. Accounting for imperfect detection in Hill numbers for biodiversity studies. *Methods in Ecology and Evolution* 6:99–108.
- Burton, A. C., M. K. Sam, C. Balangtaa, and J. S. Brashares. 2012. Hierarchical multi-species modeling of carnivore responses to hunting, habitat and prey in a West African protected area. *PLoS ONE* 7:1–14.
- Cade, B. S. 2015. Model averaging and muddled multimodel inference. *Ecology* 96:2370–2382.
- Carrillo-Rubio, E., M. Kéry, S. J. Morreale, P. J. Sullivan, B. Gardner, E. G. Cooch, and J. P. Lassoie. 2014. Use of multispecies occupancy models to evaluate the response of bird communities to forest degradation associated with logging. *Conservation Biology* 28:1034–1044.
- Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs* 84:45–67.
- Dorazio, R. M., J. A. Royle, B. Söderström, and A. Glimskär. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* 87:842–854.
- Dorazio, R. M., M. Kéry, J. A. Royle, and M. Plattner. 2010. Models for inference in dynamic metacommunity systems. *Ecology* 91:2466–2475.
- Dorazio, R. M., N. J. Gotelli, and A. M. Ellison. 2011. Biodiversity loss in a changing planet, Chapter Modern methods of estimating biodiversity from presence-absence surveys. InTech, Rijeka, Croatia.
- Elith, J., and C. H. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32:66–77.
- Elith, J., and J. Leathwick. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13:265–275.
- Fausch, K. D., and K. R. Bestgen. 1997. Ecology and Conservation of Great Plains Vertebrates, volume *Ecological Studies* 125, Chapter ecology of fishes indigenous to the central and southwestern Great Plains. pages 131–166. Springer-Verlag, New York, USA.
- Fiske, I., and R. Chandler. 2011. unmarked: An R Package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* 43:1–23.
- Geisser, S. 1993. Predictive inference: an introduction. Chapman & Hall/CRC, London, UK.
- Gelfand, A. E., J. A. Silander Jr., S. Wu, A. M. Latimer, P. O. Lewis, A. G. Rebelo, and M. Holder. 2006. Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis* 1:41–92.
- Gelman, A., and J. Hill. 2007. Data Analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.
- Gelman, A., A. Jakulin, M. G. Pittau, and Y. Sung Su. 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* 2:1360–1383.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014a. Bayesian data analysis, Third edition. CRC Press, Taylor and Francis Group, Boca Raton, Florida, USA.
- Gelman, A., J. Hwang, and A. Vehtari. 2014b. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24:997–1016.
- Giovanni, J., A. J. Kroll, J. E. Jones, B. Altman, and E. B. Arnett. 2013. Effects of management intervention on post-disturbance community composition: an experimental analysis using Bayesian hierarchical models. *PLoS ONE* 8:e59900.
- Gneiting, T., and A. E. Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102:359–378.
- Grant, E. H. C., E. F. Zipkin, J. D. Nichols, and J. P. Campbell. 2013. A strategy for monitoring and managing declines in an amphibian community. *Conservation Biology* 27:1245–1253.
- Hanley, J., and B. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction, Second edition. Springer Science+Business Media, LLC, New York, New York, USA.

- Hefley, T. J., and M. B. Hooten. 2015. On the existence of maximum likelihood estimates for presence-only data. *Methods in Ecology and Evolution* 6:648–655.
- Held, L., B. Schrödle, and H. Rue. 2010. Statistical modelling and regression structures, Chapter posterior and cross-validated predictive checks: a comparison of MCMC and INLA. pages 91–110. Springer-Verlag, Berlin, Germany.
- Hernández-Orallo, J., P. Flach, and C. Ferri. 2012. Threshold choice methods: the missing link. arXiv page arXiv:1112.2640v2.
- Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, New Jersey, USA.
- Holtrop, A. M., Y. Cao, and C. R. Dolan. 2010. Estimating sampling effort required for characterizing species richness and site-to-site similarity in fish assemblage surveys of wadeable Illinois streams. *Transactions of the American Fisheries Society* 139:1421–1435.
- Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. *Ecological Monographs* 85: 3–28.
- Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. Applied logistic regression, Third edition. Wiley, Hoboken, New Jersey, USA.
- Hutchinson, R. A., J. J. Valente, S. C. Emerson, M. G. Betts, and T. G. Dietherich. 2015. Penalized likelihood methods improve parameter estimates in occupancy models. *Methods in Ecology and Evolution* 6:949–959.
- Ikayan, K. J., M. W. Tingley, B. J. Furnas, and S. R. Beissinger. 2014. Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution* 29:97–106.
- Jin, S., L. Yang, P. Danielson, C. Homer, J. Fry, and G. Xian. 2013. A comprehensive change detection methods for updating the National Land Cover Data to circa 2011. *Remote Sensing of Environment* 132. http://www.mrlc.gov/nlcd11_data.php
- Kalies, E. L., and S. S. Rosenstock. 2013. Stand structure and breeding birds: implications for restoring ponderosa pine forests. *The Journal of Wildlife Management* 77:1157–1165.
- Kéry, M., and J. A. Royle. 2009. Modeling demographic processes in marked populations, Chapter inference about species richness and community structure using species-specific occupancy models in the National Swiss Breeding Bird Survey MHB. pages 639–656. Springer Science+Business Media, LLC, New York, New York, USA.
- Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS. A hierarchical perspective. Academic Press, Waltham, Massachusetts, USA.
- Kéry, M., J. A. Royle, M. Plattner, and R. M. Dorazio. 2009. Species richness and occupancy estimation in communities subject to temporary emigration. *Ecology* 90:1279–1290.
- Kirsch, J. E., and J. T. Peterson. 2014. A multi-scaled approach to evaluating the fish assemblage structure within Southern Appalachian streams. *Transactions of the American Fisheries Society* 143:1358–1371.
- Kroll, A. J., Y. Ren, J. E. Jones, J. Giovanini, R. W. Perry, R. E. Thill, J. Don White, and T. B. Wigley. 2014. Avian community composition associated with interactions between local and landscape habitat attributes. *Forest Ecology and Management* 326:46–57.
- Link, W. A. 2013. A cautionary note on the discrete uniform prior for the Binomial N. *Ecology* 94:2173–2179.
- Link, W. A., and M. J. Eaton. 2012. On thinning of chains in MCMC. *Methods in Ecology and Evolution* 3:112–115.
- MacKenzie, D. I., and L. L. Bailey. 2004. Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics* 9:300–318.
- 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* 83:248–255.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Elsevier Academic Press, San Diego, California, USA.
- MacNeil, M. A., E. H. Tyler, C. J. Fonnesebeck, S. P. Rushton, N. V. Polunin, and M. J. Conroy. 2008. Accounting for detectability in reef-fish biodiversity estimates. *Marine Ecology Progress Series* 367:249–260.
- Mattsson, B. J., E. F. Zipkin, B. Gardner, P. J. Blank, J. R. Sauer, and J. A. Royle. 2013. Explaining local-scale species distributions: relative contributions of spatial autocorrelation and landscape heterogeneity for an avian assemblage. *PLoS ONE* 8:1–12.
- Mazerolle, M. J., L. L. Bailey, W. L. Kendall, J. A. Royle, S. J. Converse, and J. D. Nichols. 2007. Making great leaps forward: accounting for detectability in herpetological field studies. *Journal of Herpetology* 41:672–689.
- McCullagh, P., and J. Nelder. 1989. Generalized linear models, Second edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Boca Raton, Florida, USA.
- Midway, S. R., T. Wagner, and B. H. Tracy. 2014. A hierarchical community occurrence model for North Carolina stream fish. *Transactions of the American Fisheries Society* 143:1346–1357.
- Mihaljevic, J. R., M. B. Joseph, and P. T. Johnson. 2015. Using multi-species occupancy models to improve the characterization and understanding of metacommunity structure. *Ecology* 96:1783–1792.
- Moreno, M., and S. R. Lele. 2010. Improved estimation of site occupancy using penalized likelihood. *Ecology* 91:341–346.
- Nesler, T. P., R. VanBuren, J. A. Stafford, and M. Jones. 1997. Inventory and status of south Platte River native fishes in Colorado. Technical report, Colorado Division of Wildlife, Denver, Colorado, USA.
- Noon, B. R., L. L. Bailey, T. D. Sisk, and K. S. McKelvey. 2012. Efficient species-level monitoring at the landscape scale. *Conservation Biology* 26:432–441.
- Perkin, J., K. Gibo, A. Cooper, T. Turner, M. Osborner, E. Johnson, and K. Mayes. 2015. Fragmentation and dewatering transform Great Plains stream fish communities. *Ecological Monographs* 85:73–92.
- Petit, L. I. 1990. The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society, Series B* 52:175–184.
- Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs Sampling. *Proceedings of the 3rd international workshop on distributed statistical computing* 124:125.
- Propst, D. L., and C. A. Carlson. 1986. The distribution and status of warmwater fishes in the Platte River drainage, Colorado. *The Southwestern Naturalist* 31:149–167.
- Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Ruiz-Gutiérrez, V., E. F. Zipkin, and A. A. Dhondt. 2010. Occupancy dynamics in a tropical bird community: unexpectedly high forest use by birds classified as non-forest species. *Journal of Applied Ecology* 47:621–630.
- Russell, R. E., J. A. Royle, V. A. Saab, J. F. Lehmkuhl, W. M. Block, and J. R. Sauer. 2009. Modeling the effects of environmental disturbance on wildlife communities: avian responses to prescribed fire. *Ecological Applications* 19:1253–1263.
- Scott, A., and T. M. F. Smith. 1969. Estimation in multi-stage surveys. *Journal of the American Statistical Association* 64:830–840.

- Spiegelhalter, D. J., N. G. Best, and B. P. Carlin. 1998. Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. *Research Report* 98-009.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:1–34.
- Tingley, M. W., and S. R. Beissinger. 2013. Cryptic loss of montane avian richness and high community turnover over 100 years. *Ecology* 94:598–609.
- Tobler, M. W., A. Z. Hartley, S. E. Carrillo-Perceste, and G. V. N. Powell. 2015. Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. *Journal of Applied Ecology* 52:413–421.
- Vehtari, A., and J. Ojanen. 2012. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys* 6:142–228.
- Watanabe, S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11:3571–3594.
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46(Supplement):120–138.
- White, G. C., K. P. Burnham, and D. R. Anderson. 1999. Advanced features of program MARK. *International Wildlife Management Congress* 7.5:368–377.
- Yamaura, Y., J. A. Royle, K. Kuboi, T. Tada, S. Ikeno, and S. Makino. 2011. Modelling community dynamics based on species-level abundance models from detection/nondetection data. *Journal of Applied Ecology* 48:67–75.
- Yamaura, Y., J. A. Royle, N. Shimada, S. Asanuma, T. Sato, H. Taki, and S. Makino. 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.
- Zipkin, E. F., A. DeWan, and J. A. Royle. 2009. Impacts of forest fragmentation on species richness: a hierarchical approach to community modelling. *Journal of Applied Ecology* 46: 815–822.
- Zipkin, E. F., J. A. Royle, D. K. Dawson, and S. Bates. 2010. Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation* 143:479–484.
- Zipkin, E. F., E. H. C. Grant, and W. F. Fagan. 2012. Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. *Ecological Applications* 22:1962–1972.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1890/15-1471.1/supinfo>