Statistical Reports

Identifying occupancy model inadequacies: can residuals separately assess detection and presence?

Wilson J. Wright [ID],[1,4] Kathryn M. Irvine [ID],[2] and Megan D. Higgs[3]

[1]*Department of Ecology, Montana State University, Bozeman, Montana 59717 USA*
[2]*U.S. Geological Survey, Northern Rocky Mountain Science Center, 2327 University Way Suite 2, Bozeman, Montana 59717 USA*
[3]*Department of Mathematical Sciences, Montana State University, Bozeman, Montana 59717 USA*

Abstract. Occupancy models are widely applied to estimate species distributions, but few methods exist for model checking. Thorough model assessments can uncover inadequacies and allow for deeper ecological insight by exploring structure in the observed data not accounted for by a model. We introduce occupancy model residual definitions that utilize the posterior distribution of the partially latent occupancy states. Residual-based assessments are valuable because they can target specific assumptions and identify ways to improve a model, such as adding spatial correlation or meaningful covariates. Our approach defines separate residuals for occupancy and detection, and we use simulation to examine whether missing structure for modeling detection probabilities can be distinguished from that for occupancy probabilities. In many scenarios, our residual diagnostics were able to separate inadequacies at the different model levels successfully, but we describe other situations when this may not be the case. Applying Moran's I residual diagnostics to assess models for silver-haired (*Lasionycteris noctivagans*) and little brown (*Myotis lucifugus*) bats only provided evidence of residual spatial correlation among detections. Targeting specific model assumptions using carefully chosen residual diagnostics is valuable for any analysis, and we remove previous barriers for occupancy analyses—lack of examples and practical advice.

Key words: hierarchical model; imperfect detection; model assessment; occupancy models; posterior predictive checks; residual diagnostics.

Introduction

Occupancy models explicitly account for imperfect detection while modeling the occurrence of a species (MacKenzie et al. 2018). Extensions to the original framework have made this approach widely applicable (Guillera-Arroita 2017), yet there is still relatively little work on assessing occupancy model fit. Thorough model assessments are a fundamental aspect of any data analysis because they help check underlying assumptions and identify inadequacies of a model for describing patterns in empirical data sets (Conn et al. 2018). Checking

occupancy models is difficult, however, because there are distinct model components for detection and occurrence. Although this hierarchical structure provides valuable ecological interpretations, it necessitates a more nuanced investigation of fit, because both levels of the model have assumptions that should be checked. To add to the challenge, it is unclear how well diagnostics are able to assess the two components separately, because they are inherently linked in the model.

The approaches originally proposed for assessing occupancy models focused on summaries of overall model fit. An omnibus goodness-of-fit test based on a chi-square statistic for unique detection histories was developed (MacKenzie and Bailey 2004), and similar test statistics have been used in posterior predictive checks (PPCs) for Bayesian analyses (e.g., Broms et al. 2016). Summaries of overall model fit, however, can be less

[4] E-mail: wilson.wright@msu.montana.edu

sensitive for detecting model inadequacies than focused evaluations that target assumptions of concern (e.g., nonindependent observations; Wright et al. 2016). Additionally, targeted diagnostics can help uncover how or why a model may be lacking compared to the observed data. Recently, Warton et al. (2017) and MacKenzie et al. (2018) presented residual diagnostics for occupancy models along with visualization strategies for diagnosing unexplained relationships between covariates and the probabilities of occupancy and detection. We explore a different approach for defining occupancy model residuals and investigate diagnostics for assessing spatial independence using Moran's I statistic (Moran 1950).

We were motivated to consider residual diagnostics for occupancy models by our work studying bat species in Montana, USA. Occupancy analyses were conducted to estimate where species susceptible to white-nose syndrome (WNS), a deadly fungal disease in bats, have high probability of occurrence in the state. This information can help focus WNS surveillance efforts on areas where at-risk species likely occur and establish baseline (pre-WNS) occupancy estimates. Based on the spatial configuration of sampled locations, we suspected a need for incorporating spatial correlation in our analyses but wanted to assess this assumption before fitting additional models. Although applications of spatial occupancy models generally focus on correlation among occupancy probabilities (e.g., Johnson et al. 2013), spatial correlation among detection probabilities is also possible and can negatively impact inferences when unmodeled. We were unaware, however, of approaches for investigating evidence of spatial correlation in residuals from a fitted occupancy model and whether we could distinguish which component(s), occupancy and/or detection, needed modification. Finally, we also wanted to assess our expanded models, which included spatial correlation, to confirm we adequately accounted for spatial patterns in the residuals.

Here, we appeal to Bayesian methods to define separate residuals for occupancy and detection to allow for different assessments at each level of the model. These residuals can be useful for assessing various model assumptions, but we focus on investigating residual spatial correlation because of our bat examples. We implemented residual diagnostics for occupancy analyses of silver-haired (*Lasionycteris noctivagans*; LANO) and little brown (*Myotis lucifugus*; MYLU) bat data from Montana. The Moran's I assessments only provided evidence of residual spatial correlation among detections, and our strategic exploration of model fit helped guide how our models could be expanded to improve inferences. We conducted a simulation study to evaluate the effectiveness of residual diagnostics for identifying inadequate spatial structure. Further, we explored when the separate residuals for detection and occupancy were able to distinguish which model component contained the assumption violation.

METHODS

Model formulation and residual definitions

Occupancy models are frequently expressed in a hierarchical form capturing both the ecological state process of occupancy (or presence) and the observation process of detection (Kéry and Royle 2016, MacKenzie et al. 2018). In this formulation, for a single-season model, the partially latent occupancy states (Z_i) for sites $i = 1, \ldots, n$ are modeled as

$$Z_i \sim \text{Bernoulli}(\psi_i), \tag{1}$$

where ψ_i represents the probability of occupancy. Conditional on site occupancy ($Z_i = 1$), the species is detected during visits $j = 1, \ldots, J_i$ with probabilities p_{ij} such that

$$[Y_{ij}|Z_i = 1] \sim \text{Bernoulli}(p_{ij}), \tag{2}$$

where J_i is the total number of visits to site i. Heterogeneity in the occupancy probabilities among sites can be incorporated using the generalized linear model framework, typically by connecting the probabilities to site-level covariates through a logit link function, $\text{logit}(\psi_i) = \mathbf{X}_i \boldsymbol{\beta}$. Similarly, the probability of detection across sites and/or visits is often modeled by $\text{logit}(p_{ij}) = \mathbf{V}_{ij} \boldsymbol{\alpha}$. Here, $\mathbf{X}_i$ is a row vector of site-level covariates for site i and $\mathbf{V}_{ij}$ is a row vector combining site-level and visit-level covariates for visit j at site i. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ represent the vectors of coefficients for the occupancy and detection components of the model respectively.

Basic occupancy models assume no false positive detections occur, implying sites with at least one detection ($(\sum_{j=1}^{J_i} y_{ij}) > 0$) are truly occupied and Z_i is one. However, ambiguity in occupancy status occurs at sites without any detections, and we use a Bayesian approach to describe the uncertainty in occupancy states. Consider the case where Markov chain Monte Carlo (MCMC) methods are used to approximate the posterior distribution of interest. Conditional on the vector of observed detections from site i, $\mathbf{y}_i$, Bayes' Theorem provides the probability it is occupied as

$$\Pr(Z_i^{[t]} = 1|\mathbf{y}_i) = \begin{cases} 1, & \left(\sum_{j=1}^{J_i} y_{ij}\right) > 0, \\ \frac{\psi_i^{[t]} \prod_{j=1}^{J_i} \left(1 - p_{ij}^{[t]}\right)}{\left(1 - \psi_i^{[t]}\right) + \psi_i^{[t]} \prod_{j=1}^{J_i} \left(1 - p_{ij}^{[t]}\right)}, & \left(\sum_{j=1}^{J_i} y_{ij}\right) = 0, \end{cases} \tag{3}$$

given the current iteration's probabilities of occupancy and detection (denoted $\psi_i^{[t]}$ and $p_{ij}^{[t]}$ for iteration t). Samples from the posterior distribution of the occupancy states are generated as random draws from Bernoulli distributions with these probabilities (Eq. 3) at each

MCMC iteration. After convergence, the collection of draws approximates the posterior distribution of the latent occupancy states and can be used to make inference about other quantities of interest (e.g., total number of sites occupied; Kéry and Royle 2016) or to evaluate a model's predictive performance (Zipkin et al. 2012). For sites with unknown occupancy status, the posterior draws for the latent occupancy state depend on ψ_i and p_{ij}, meaning model assessments may not be able to untangle the two processes completely. However, posterior distributions of latent variables still provide useful information that can be used to construct PPCs (Gelman 2004, Gelman et al. 2005) or pivotal discrepancy measures (Yuan and Johnson 2012). Here, we use the posterior distribution of the occupancy states to define occupancy model residuals.

Similar to the approaches used by Warton et al. (2017) and MacKenzie et al. (2018), we define residuals for the occupancy component (Eq. 1) and separate residuals for the detection component (Eq. 2). Our definitions differ, however, because the residuals are calculated conditional on the posterior distribution of latent occupancy states. For binary logistic regression, raw residuals are often defined as the observed value minus the probability of "success", and we use this residual definition. The occupancy residual for site i and iteration t is expressed as

$$o_i^{[t]} = z_i^{[t]} - \psi_i^{[t]}, \tag{4}$$

where $z_i^{[t]}$ is a draw from the posterior distribution of the occupancy state at site i. For this component, the term "discrepancy" may be more appropriate, because the occupancy states are partially latent, but we use "residuals" to connect our suggestions to more familiar residual diagnostics.

We also use the posterior distribution of occupancy states in our definition of detection residuals. For posterior draw t, the detection residual for site i and visit j is expressed as

$$\left[d_{ij}^{[t]} \middle| z_i^{[t]} = 1 \right] = y_{ij} - p_{ij}^{[t]}. \tag{5}$$

These detection residuals are only defined for "occupied" sites based on the latent state $z_i^{[t]}$ for a given posterior draw t. By conditioning on a posterior draw of the occupancy state $\left(z_i^{[t]} \right)$, sites without any detections ($\mathbf{y}_i$ containing all zeros) still contribute detection residuals for some posterior draws because these sites might actually be occupied $\left(z_i^{[t]} = 1 \text{ for some iterations} \right)$. Additionally, the total number of detection residuals will vary across posterior draws because of the uncertainty in true occupancy at sites with no detections.

Residual-based model assessments

Residual diagnostics are a familiar tool for checking linear model assumptions and could be equally useful for assessing occupancy models. Residuals can be used graphically or form the basis for PPCs, making them useful for evaluating various aspects of a model depending on the chosen assessment (Conn et al. 2018). For occupancy models, a key aspect for any assessment is that it can be performed using both occupancy and detection residuals, with the goal of separately evaluating the two levels of the model. Perhaps the most familiar and straightforward residual assessment is to check for patterns in the plot of residuals vs. values of a covariate. With binary residuals, as we have for occupancy and detection (Eqs. 4 and 5), this can be achieved by looking at binned residual plots (e.g., Gelman et al. 2000). In this case, the values of a covariate are used to form approximately equally sized groups. Plotting the averages of the residuals in each group vs. the corresponding average covariate values can reveal unexplained structure in fitted probabilities. Examining such residual plots from multiple posterior draws characterizes variation over the posterior distribution.

Besides graphical assessments, residuals can be the basis for calculating discrepancy measures for PPCs. Systematic difference between discrepancy measures calculated from residuals based on observed data and those expected under the assumed model structure provide evidence of a mismatch between the data and fitted model. For each posterior draw, the observed discrepancy is compared to the corresponding value from the posterior predicted replicates. The proportion of posterior draws where the observed value is larger than the posterior predictive value is termed the posterior predictive probability (PPP) associated with that discrepancy measure. Very large, or small, PPPs indicate predictions from the model are not consistent with the observed data relative to the chosen measure. Assessments of overall model fit can be performed using a discrepancy measure like the chi-square statistic based on total detections at a site (e.g., Broms et al. 2016). Alternatively, the discrepancy measure is constructed to assess an assumption of interest more specifically. Again, the benefit of separate residuals for detection and occupancy is that analysts can perform assessments to consider model inadequacies one component at a time. Although residuals can be used to assess many aspects of a model, we focus on evaluating spatial correlation in occupancy model residuals as motivated by our bat examples.

Investigating unaccounted for spatial correlation

To investigate spatial structure unaccounted for by a fitted occupancy model, we use Moran's I statistic calculated with residuals to create correlograms and perform PPCs. The Moran's I statistic is commonly used for areal data to explore spatial correlation in continuous

Statistical Reports

response variables (Moran 1950, Schabenberger and Gotway 2005). Based on a defined set of neighbors, this statistic summarizes the magnitude of spatial correlation in the residuals. Let $w_{ii'}$ be an indicator for whether sites i and i' are neighbors (1) or not (0) and $w_{..} = \sum_{i=1}^{n} \sum_{i'=1}^{n} w_{ii'}$. With the occupancy residuals $\left(o_i^{[t]}\right)$, Moran's I for posterior draw t is calculated as

$$I^{[t]} = \frac{n}{(n-1)\mathrm{Var}(o^{[t]})w_{..}} \sum_{i=1}^{n} \sum_{i'=1}^{n} w_{ii'} \left(o_i^{[t]} - \bar{o}^{[t]}\right)\left(o_{i'}^{[t]} - \bar{o}^{[t]}\right),$$

(6)

where $\bar{o}^{[t]}$ and $\mathrm{Var}\left(o^{[t]}\right)$ denote the sample mean and variance, respectively, of the n occupancy residuals. The equation can be used to calculate a Moran's I statistic with the detection residuals by replacing the o terms with d terms. For the Moran's I empirical correlogram, residuals are considered neighbors when the distance between their corresponding sites (centers of areal units) is within a given distance class, where different sets of neighbors can be created across classes defined by increasing distances. Our empirical correlogram is the plot of the Moran's I statistic calculated with neighbors from each distance class vs. the corresponding average distance between neighbors. The plot indicates how residual spatial correlation changes across distance classes, where larger Moran's I statistics indicate greater correlation. Because the residuals are defined for each posterior draw, there is variability in these statistics over the posterior distribution. We simultaneously plot empirical correlograms from a random collection of posterior draws to summarize the posterior variability in the degree of remaining spatial correlation.

For a related PPC, we use the "queen" neighbor definition (i.e., adjacent sites based on moving a queen chess piece; Schabenberger and Gotway 2005) to calculate Moran's I. In other words, the pair of occupancy residuals $o_i^{[t]}$ and $o_{i'}^{[t]}$ are neighbors if sites i and i' are queen neighbors. A pair of detection residuals $\left(d_{ij}^{[t]}$ and $d_{i'j'}^{[t]}\right)$ are classified as neighbors if they are from the same site ($i = i'$) or from neighboring sites based on the queen definition (sites i and i' neighbors). With the Moran's I statistic as the discrepancy measure, this PPC can be summarized by calculating the PPP of more observed correlation within the queen's neighborhood than expected under a fitted model. This assessment is potentially less descriptive than the Moran's I empirical correlogram because it focuses on one distance class instead of simultaneously displaying multiple classes.

Bat data examples

Mist netting and acoustic surveys for silver-haired (LANO) and little brown (MYLU) bats were conducted in Montana during the summers (defined as June to September) of 2008, 2009, and 2010. We briefly describe the methods here and include further details in Appendix S1. A single nightly survey using either mist netting or an acoustic detector was considered a single visit to build detection histories. Mist netting captures were considered unambiguous detections, and we assumed all captured bats were correctly identified to species. Acoustic files were analyzed using Sonobat (version 3.0; https://sonobat.com/) and manual confirmation of species identifications by an expert unequivocally indicates a species was detected during an acoustic survey (Reichert et al. 2018). All surveys were associated with the 10×10-km grid developed for the North American Bat Monitoring Program (NABat; Loeb et al. 2015) for analyses. Each grid cell was considered a site where species presence (or not) was of interest. Throughout we often use the term "occupancy," but for highly mobile species like bats we prefer the terms "presence" or "occurrence." Sites across Montana were surveyed, but surveys were often spatially and temporally clustered because of logistical constraints (see Appendix S1), suggesting spatial correlation (potentially because of temporal correlation) could be an important consideration in our analyses.

Forest percent cover (log-transformed), elevation, ruggedness, and average degree days were included as site-level covariates based on similar analyses for Oregon and Washington (Rodhouse et al. 2015). For each nightly visit, we included the method (mist netting or acoustic), survey date within the year, and daily weather measurements (precipitation and minimum temperature) from nearby weather stations as potential detection-level explanatory variables. For each species, we fit initial models assuming independence for both model components after accounting for covariates.

We used binned residual plots to assess the assumed functional form of the covariates used to model the probabilities. For each covariate, we plotted the binned residuals from 5 to 10 random posterior draws. To explore unaccounted-for spatial correlation, we examined Moran's I empirical correlograms (distance classes for every 15 km) and PPCs, as described previously. Based on these assessments, we expanded the models to include, as needed, spatially correlated effects using restricted spatial regression (Hughes and Haran 2013, Johnson et al. 2013). This method assumes an intrinsic conditionally autoregressive (ICAR) structure on site-level random effects added to the logit-transformed probabilities for occupancy and/or detection. The restricted structure of the spatial effects can make model fitting more computationally efficient and help reduce confounding with covariates that have spatial structure. We reexamined the Moran's I residual assessments after fitting the spatial occupancy models. We conducted these analyses using R (version 3.4.1; R Core Team 2017). All models were fit using Stan (Carpenter et al. 2017) called via R with the rstan package (version 2.17.3; Stan Development Team 2018). All data and code for these analyses are provided in Data S1.

Simulation study

Motivated by our bat data analyses, we conducted a simulation study to evaluate the effectiveness of Moran's I PPP for identifying inadequate spatial structure in an occupancy model. For ease of presenting results over many simulated data sets, our comparisons focused on the PPP summary instead of graphical assessment. All simulations were performed using R (version 3.4.1; R Core Team 2017) and models were fit with the rstan package (version 2.17.3; Stan Development Team 2018). The code for all simulations is available in Data S2.

We simulated data sets for sites defined as cells on a 15×15 grid ($n = 225$) with $J = 5$ visits to each. The data-generating occupancy probabilities were defined as $\text{logit}(\psi_i) = \beta_0 + \beta_1 x_i + \eta_i$ with $\beta_0 = 0$ and $\beta_1 = 1$. Detection probabilities were defined as $\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 v_{ij} + \theta_i$ with $\alpha_0 = 0$ and $\alpha_1 = 1$. Both of the covariates, x_i and v_{ij}, were random draws from independent Uniform$(-1, 1)$ distributions. Spatial correlation was incorporated in the data-generating process by adding spatially correlated, site-level effects to the logit-transformed probabilities of occupancy and/or detection (denoted by η_i and θ_i, respectively). These spatial effects approximately followed an ICAR structure with scaling factor σ representing the strength of spatial correlation (described for occupancy models by Johnson et al. 2013). We generated data sets under four scenarios, which varied based on which model component the spatially correlated effects were added to: detection only (all $\eta_i = 0$), occupancy only (all $\theta_i = 0$), both with independent spatial effects ($\eta_i \neq \theta_i$, but with equal σ), or both with identical spatial effects ($\eta_i = \theta_i$). For each scenario, we varied the strength of spatial correlation using $\sigma = \{0, 0.5, 1, 1.5, 2\}$ and generated 100 different data sets for each value. Note that when $\sigma = 0$, all simulated spatial effects are zero, which represents no added spatial correlation.

For each simulated data set, three occupancy models were fit with each including the correct covariate structure (i.e., x_i and v_{ij} included in all fitted models). The "nonspatial" model assumed independence and did not include spatial correlation for either component. The additional models fit added an ICAR spatial structure among the detection or occupancy probabilities ("spatial detection" and "spatial occupancy," respectively). We calculated a PPP based on Moran's I with a queen neighborhood definition for every model fit, with separate assessments using the detection and occupancy residuals. For each scenario and strength of correlation, we calculated the proportion of simulated data sets identified as having more residual spatial correlation than expected under the fitted models using these assessments. To summarize over all the simulated data sets, we considered a Moran's I PPP greater than 0.9 as an indication of spatial correlation. In practice, however, a strict cutoff is not necessary and the probabilities should be interpreted on a continuous scale, as we did in our empirical analyses.

Bat data examples

For both species, the Moran's I empirical correlograms using the detection residuals decreased with increasing distance, but there were no strong patterns apparent using the occupancy residuals (Fig. 1a). Additionally, the Moran's I PPP provided evidence for residual spatial correlation in the detection probabilities for both species (LANO: >0.999; MYLU: 0.993) but not the occupancy probabilities (LANO: 0.574; MYLU: 0.680). Plotting binned residuals vs. the available covariates did not reveal patterns that would indicate an important covariate was excluded or an alternative functional form of a covariate was needed in our initial models (e.g., Fig. 1b for elevation). Therefore, we opted to expand our initial models by incorporating spatial correlation among the detection probabilities.

After explicitly modeling spatial correlation among detection probabilities, assessments of the updated models no longer indicated unexplained spatial correlation. The Moran's I empirical correlograms no longer showed strong patterns (Appendix S1) and the PPPs using the detection residuals were lower for both species (LANO: 0.674; MYLU: 0.659). There was still the question, however, of how well these residual diagnostics could assess the detection component separately from the occupancy component.

Simulation study

Spatial correlation in detection only.—For data generated with spatial correlation in the detection component only, Moran's I PPP with detection residuals identified model inadequacy only when the fitted model incorrectly assumed independent detection probabilities (Fig. 2a, orange lines fitting non-spatial and spatial occupancy models). As expected, the proportion of simulated data sets with PPP > 0.90 increased as the strength of spatial correlation among detection probabilities increased. When detection probabilities were adequately modeled with an ICAR spatial structure, assessments did not indicate remaining spatial correlation in detection residuals, as desired (Fig. 2a, orange line fitting spatial detection model). For the Moran's I PPC based on occupancy residuals, the proportion of simulated data sets with PPP > 0.90 remained low regardless of the fitted model (Fig. 2a, purple lines for all models).

Spatial correlation in occupancy only.—When data sets were simulated with spatially correlated occupancy probabilities, the Moran's I PPP using detection residuals indicated model inadequacy in a small proportion of simulated data sets (Fig. 2b, orange lines fitting all models). The proportion of simulated data sets where the occupancy residual assessments correctly identified the fitted model was misspecified increased as the strength
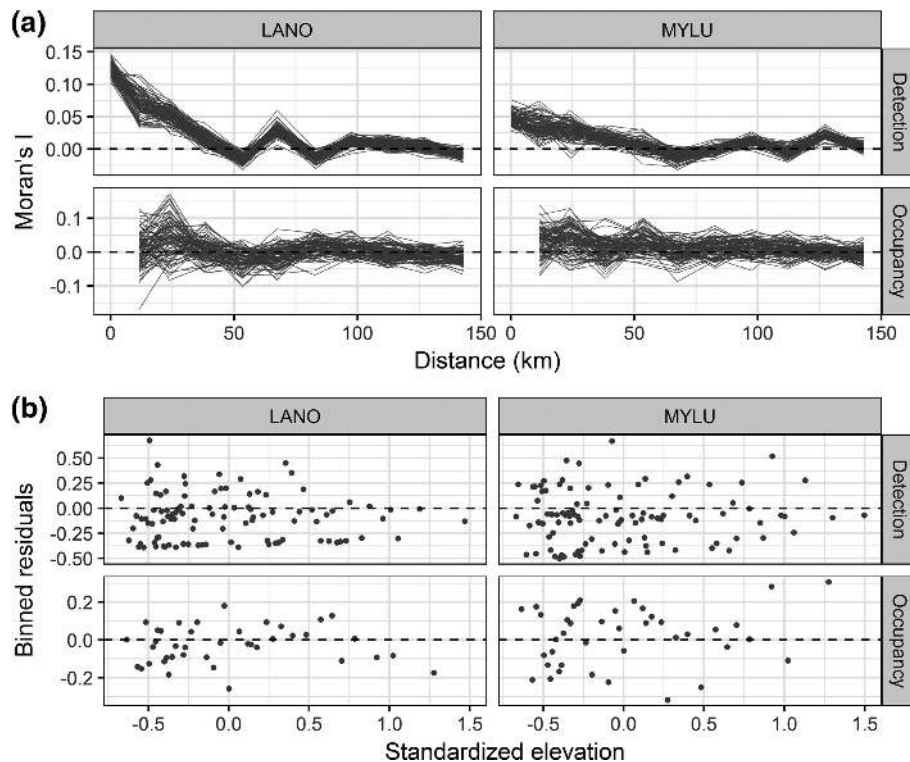
FIG. 1. Residual-based model assessments for the silver-haired (LANO) and little brown (MYLU) bat analyses that assumed spatial independence. The Moran's I correlograms from 100 random posterior draws consistently show a decreasing pattern in the detection residuals but no strong patterns in the occupancy residuals (a). For both species, there were no strong patterns in the binned residual plots using the elevation covariate (b). We show the residuals for a single random posterior draw, but for each covariate we examined plots from 5 to 10 random posterior draws.

of simulated spatial correlation among occupancy probabilities increased (Fig. 2b, purple lines fitting nonspatial and spatial detection models). The Moran's I PPC with the occupancy residuals rarely resulted in PPP > 0.90 when the fitted model correctly included an ICAR structure on the occupancy probabilities (Fig. 2b, purple line fitting spatial occupancy model).

Spatial correlation in both components.—When spatial correlation was simulated in both occupancy and detection probabilities, regardless of whether the spatial effects were independent or identical, the Moran's I PPP using detection residuals successfully identified models that incorrectly assumed a detection process without spatial correlation (Fig. 2c, d, orange lines fitting nonspatial and spatial occupancy models). Moran's I PPP using occupancy residuals, however, had less clear-cut abilities to identify unaccounted-for spatial correlation. Specifically, the proportion of Moran's I PPP > 0.90 using occupancy residuals was larger if the fitted model assumed independence compared to assessing a fitted model that included spatially correlated detections (Fig. 2c, d, comparing purple lines for nonspatial and spatial detection model fits). Accounting for spatial correlation among detection probabilities increases

uncertainty in the posterior distribution of occupancy states, and may have contributed to the assessments with occupancy residuals having a decreased ability to discern spatial correlation compared to the same assessments for the nonspatial model.

DISCUSSION

We introduced definitions for occupancy model residuals that utilize the posterior distribution of the partially latent occupancy states. When residual diagnostics were applied in our data analyses, the Moran's I empirical correlograms and PPPs for our initial models provided evidence of residual spatial correlation among detection probabilities for silver-haired and little brown bats. Conversely, we found no evidence that occupancy probabilities had residual spatial correlation. In general, many different underlying processes (e.g., abundance) could explain our results. For example, temporal patterns in bat activity could manifest as spatial correlation in our data because nearby grid cells were often sampled within days of each other. Other work has shown temporal correlation in detection probabilities for bats (Wright et al. 2016) likely due to runs in bat activity over consecutive nights (Hayes 1997). Future monitoring efforts could
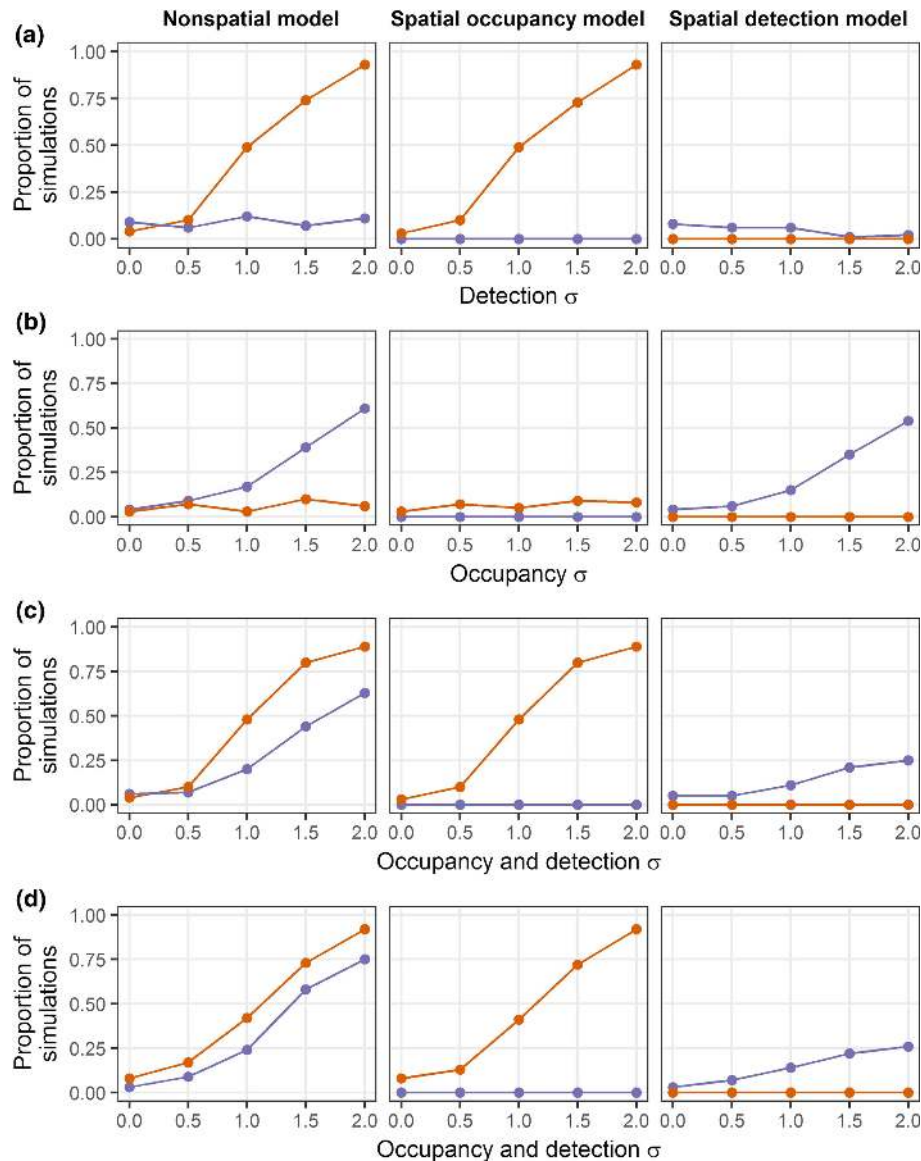
FIG. 2. For the detection residuals (orange) and occupancy residuals (purple), proportion of simulated data sets where the Moran's I posterior predictive probability was larger than 0.9 vs. the level of spatial correlation (σ). Scenarios explored incorporating spatial correlation for (a) detection only, (b) occupancy only, (c) both detection and occupancy with separate effects, and (d) both detection and occupancy with identical effects. For all scenarios and generating coefficient values, 100 data sets were simulated. As indicated by the columns of the plot, each data set was analyzed using a nonspatial model, a model including spatial correlation among occupancy probabilities only, and a model including spatial correlation among detection probabilities only.

avoid the need for fitting spatial occupancy models if the spatial locations and timing of surveys are modified (e.g., MacKenzie et al. 2018:303–304). Compared to the models assuming independence, the posterior distributions from our expanded models had lower overall detection probabilities and higher overall occupancy probabilities for both species (Appendix S1), consistent with patterns expected from correlated detection probabilities (Hines et al. 2010). Incorporating spatial correlation among detection probabilities also increased the uncertainty in all model parameters conveyed by the

posterior distributions. Consequently, assessing spatial structure of both model components is a crucial, and perhaps underappreciated, step in evaluating occupancy models.

Untangling whether inadequacies of a fitted model correspond to the estimated probabilities of occupancy or detection is important because it helps suggest how to appropriately expand a model. With our residual definitions, we attempt to evaluate the detection and occupancy components separately by conditioning on the posterior distribution of the latent occupancy states.

Statistical Reports

However, the two components are inherently linked in the model because the occupancy posterior distribution (Eq. 3) depends on both model parameters (ψ and p). In practice, our simulations illustrated that residual-based assessments can still be useful for separately evaluating each level of the model. Consistently across the scenarios we explored, the Moran's I assessment with the detection residuals was only sensitive to unexplained spatial correlation among detection probabilities. Similarly, assessments with the occupancy residuals indicated model misspecification only for scenarios where simulated spatial correlation was not included for the occupancy component. For occupancy residuals, however, performance of the Moran's I PPC depended on the fitted model structure for the detection probabilities. For instance, adding an ICAR spatial structure on the detection probabilities reduced the effectiveness of this PPC for identifying spatial correlation among occupancy probabilities. This may be due to more uncertainty in the fitted occupancy probabilities compared to a nonspatial model. Residual assessments for missing covariates can also separately assess the model components (simulation results not shown) and help pinpoint how a model could be improved.

Other scenarios could result in residual assessments that are less useful for separately checking the levels of occupancy models. It is more difficult to distinguish the two components when there is less information for informing the occupancy status at sites without any detections. In general, this happens when sites have a small number of visits or detection probabilities are small. As an extreme example, with only one visit to every site there is no information to assess (or even estimate) detection and occupancy separately. Consequently, lower detection probabilities and/or fewer visits per site could result in residual assessments that are less able to distinguish which model level contains the assumption violation. This makes it important to think critically about the study design and whether there are enough data to fit and evaluate models (Warton et al. 2017, MacKenzie et al. 2018). Because the number of visits and expected detection probabilities are the same criteria considered for reducing uncertainty in occupancy estimates, following general study design recommendations (MacKenzie et al. 2018) will also improve the effectiveness of assessing models with residual diagnostics.

Other approaches to define residuals for occupancy models do not condition on the posterior distribution of the latent occupancy states. The residuals proposed by Warton et al. (2017), for instance, are conditional on naive occupancy and apply the "Dunn-Smyth" transformation proposed for discrete observations. The "Dunn-Smyth" transformation includes random noise that we found reduced the ability to identify model inadequacies (Appendix S2). Residuals defined conditional on naive occupancy but without the transformation performed better and, in many scenarios, were comparable to the latent occupancy residuals we described (Appendix S2). However, calculating residuals conditional on naive occupancy is more cumbersome because it requires additional expected values that are complicated and not included in standard model output. Instead, conditioning on the latent occupancy state allows for intuitive residual definitions that we found easier to implement and use in practice.

Residuals are a cornerstone for assessing many statistical models because they are a flexible tool for investigating assumptions of interest. Defining residuals conditional on the posterior distribution of the occupancy states helps account for the hierarchical structure of occupancy models when performing residual-based assessments. In our bat data examples, we focused on investigating unexplained spatial correlation and improving inferences by using residual diagnostics to guide model expansions. We also found that residuals defined with the occupancy posterior distribution can effectively explore other model assumptions, such as the functional form of covariates for the fitted probabilities. Using carefully chosen residual diagnostics to target particular model assumptions should become an integral part of evaluating occupancy models. Specific assessments for both levels of an occupancy model can help guard against potentially erroneous inferences and allow for deeper ecological insight by exploring structure not accounted for by a model.

Literature Cited

Broms, K. M., M. B. Hooten, and R. M. Fitzpatrick. 2016. Model selection and assessment for multi-species occupancy models. Ecology 97:1759–1770.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software 76:1–32.

Conn, P. B., D. S. Johnson, P. J. Williams, S. R. Melin, and M. B. Hooten. 2018. A guide to Bayesian model checking for ecologists. Ecological Monographs 88:526–542.

Gelman, A. 2004. Exploratory data analysis for complex models. Journal of Computational and Graphical Statistics 13:755–779.

Gelman, A., Y. Goegebeur, F. Tuerlinckx, and I. V. Mechelen. 2000. Diagnostic checks for discrete data regression models using posterior predictive simulations. Journal of the Royal Statistical Society Series C (Applied Statistics) 49:247–268.

Gelman, A., I. Van Mechelen, G. Verbeke, D. F. Heitjan, and M. Meulders. 2005. Multiple imputation for model checking: Completed-data plots with missing and latent data. Biometrics 61:74–85.

Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography 40:281–295.

Hayes, J. P. 1997. Temporal variation in activity of bats and the design of echolocation-monitoring studies. Journal of Mammalogy 78:514–524.

Hines, J. E., J. D. Nichols, J. A. Royle, D. I. MacKenzie, A. M. Gopalaswamy, N. S. Kumar, and K. U. Karanth. 2010. Tigers on trails: occupancy modeling for cluster sampling. Ecological Applications 20:1456–1466.

Hughes, J., and M. Haran. 2013. Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. Journal of the Royal Statistical Society Series B (Statistical Methodology) 75:139–159.

Johnson, D. S., P. B. Conn, M. B. Hooten, J. C. Ray, and B. A. Pond. 2013. Spatial occupancy models for large data sets. Ecology 94:801–808.

Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and static models. Academic Press, San Diego, California, USA.

Loeb, S. C., et al. 2015. A plan for the North American bat monitoring program (NABat). General Technical Report SRS-208. Southern Research Station, U.S. Department of Agriculture, Forest Service, Asheville, North Carolina, USA.

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, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2018. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Second edition. Academic Press, San Diego, California, USA.

Moran, P. A. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17–23.

R Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reichert, B., et al. 2018. A guide to processing bat acoustic data for the North American bat monitoring program (NABat). Open-File Report 2018–1068. U.S. Geological Survey, Reston, Virginia, USA.

Rodhouse, T. J., P. C. Ormsbee, K. M. Irvine, L. A. Vierling, J. M. Szewczak, and K. T. Vierling. 2015. Establishing conservation baselines with dynamic distribution models for bat populations facing imminent decline. Diversity and Distributions 21:1401–1413.

Schabenberger, O., and C. A. Gotway. 2005. Statistical methods for spatial data analysis. Chapman & Hall/CRC Press, Boca Raton, Florida, USA.

Stan Development Team. 2018. RStan: the R interface to Stan. Version 2.17.3. http://mc-stan.org

Warton, D. I., J. Stoklosa, G. Guillera-Arroita, D. I. MacKenzie, and A. H. Welsh. 2017. Graphical diagnostics for occupancy models with imperfect detection. Methods in Ecology and Evolution 8:408–419.

Wright, W. J., K. M. Irvine, and T. J. Rodhouse. 2016. A goodness-of-fit test for occupancy models with correlated within-season revisits. Ecology and Evolution 6:5404–5415.

Yuan, Y., and V. E. Johnson. 2012. Goodness-of-fit diagnostics for Bayesian hierarchical models. Biometrics 68:156–164.

Zipkin, E. F., E. H. Campbell 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/10.1002/ecy.2703/suppinfo

Statistical Reports