Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015 doi: 10.1111/2041-210X.12442

Occupancy models for data with false positive and false negative errors and heterogeneity across sites and surveys

Paige F.B. Ferguson[1,2]*, Michael J. Conroy[1] and Jeffrey Hepinstall-Cymerman[1]

[1]*Warnell School of Forestry and Natural Resources, University of Georgia, 180 E. Green Street, Athens, GA 30602, USA; and* [2]*Cary Institute of Ecosystem Studies, 2801 Cty Road 44A, Millbrook, NY 12545, USA*

Summary

1. False positive detections, such as species misidentifications, occur in ecological data, although many models do not account for them. Consequently, these models are expected to generate biased inference.
2. The main challenge in an analysis of data with false positives is to distinguish false positive and false negative processes while modelling realistic levels of heterogeneity in occupancy and detection probabilities without restrictive assumptions about parameter spaces.
3. Building on previous attempts to account for false positive and false negative detections in occupancy models, we present hierarchical Bayesian models that utilize a subset of data with either confirmed detections of a species' presence (CP model) or both confirmed presences and confirmed absences (CACP model). We demonstrate that our models overcome the challenges associated with false positive data by evaluating model performance in Monte Carlo simulations of a variety of scenarios. Our models also have the ability to improve inference by incorporating previous knowledge through informative priors.
4. We describe an example application of the CP model to quantify the relationship between songbird occupancy and residential development, plus we provide instructions for ecologists to use the CACP and CP models in their own research.
5. Monte Carlo simulation results indicated that, when data contained false positive detections, the CACP and CP models generated more accurate and precise posterior probability distributions than a model that assumed data did not have false positive errors. For the scenarios we expect to be most generally applicable, those with heterogeneity in occupancy and detection, the CACP and CP models generated essentially unbiased posterior occupancy probabilities. The CACP model with vague priors generated unbiased posterior distributions for covariate coefficients. The CP model generated unbiased posterior distributions for covariate coefficients with vague or informative priors, depending on the function relating covariates to occupancy probabilities. We conclude that the CACP and CP models generate accurate inference in situations with false positive data for which previous models were not suitable.

Key-words: confirmed detection, detection probability, hierarchical Bayesian model, imperfect detection, informative prior, Monte Carlo simulation, observation, occupancy model, phantom species, scenario

Introduction

Many methods of data collection are subject to imperfect detections, including false negative detections (not detecting a species when it is present) and false positive detections (detecting a species when it is not present, which can occur through species misidentification). Unbiased inference requires accounting for imperfect detection. Methods to account for false negative detections have been widely adopted in occupancy models (MacKenzie *et al.* 2002) and other ecological models (e.g. mark–recapture, distance estimation and band recovery models). However, many methods that estimate a false negative detection probability assume that false positives do not occur. But false positive errors can occur in data collected by citizen scientists (Miller *et al.* 2013) and non-experts (Fitzpatrick *et al.* 2009), as well as by scientists of all experience levels (Simons *et al.* 2007; Alldredge *et al.* 2008; McClintock *et al.* 2010a; Miller *et al.* 2012). Further, Monte Carlo simulations have demonstrated that if data contain false positive errors but analyses do not account for them, inference about occupancy probability and covariate coefficients will be biased (Royle & Link 2006; McClintock *et al.* 2010b). Because false positive detections are unlikely to be eliminated through study design and because analyses that fail to account for them generate biased results, it is important to develop and employ methods that account for both types of imperfect detection.

*Correspondence author. E-mail: barlowp@caryinstitute.org

Occupancy models accounting for both types of imperfect detection have been developed in Royle & Link (2006) and Miller *et al.* (2011), but, as we will show, the model in Royle & Link (2006) has limited application by its nature, while the ability of the model in Miller *et al.* (2011) to estimate highly heterogeneous occupancy and detection probabilities has not been demonstrated yet. There is symmetry in the likelihood of the model in Royle & Link (2006) (hereafter, the Royle-Link model) so that there is not a unique set of solutions for parameter values. For example, the following two sets of parameter values have identical likelihoods under the Royle-Link model: (i) 75% of sites occupied, 30% true-positive detection rate and 10% false positive detection rate; and (ii) 25% of sites occupied, 10% true-positive detection rate and 30% false positive detection rate (Royle & Link 2006). To address this problem, Royle & Link (2006) restricted the parameter space so that the true-positive detection probability (i – false negative detection probability) is greater than the false positive detection probability. However, this assumption has received criticism (McClintock *et al.* 2010b). For example, if there is a 'phantom' species, a species that is not present in the study area but is detected, results from the Royle-Link model would suggest that the 'phantom' species is actually present (McClintock *et al.* 2010b). The Royle-Link model cannot be used to determine that the 'phantom' species is only detected because of a false positive error. Additionally, false positive and true-positive detections cannot be distinguished if there is heterogeneity in detection probabilities (McClintock *et al.* 2010b).

It is not possible to distinguish false negative and false positive detections using standard occupancy data and models that make few assumptions about parameter values (McClintock *et al.* 2010b), so Miller *et al.* (2011) proposed occupancy models (hereafter, the Miller models) that estimate false negative and false positive detection probabilities using additional information about the detection process. In Miller's multiple detection state (MDS) model, one detection method is used, but detections are classified as either detections in which false positive errors are possible (unconfirmed detections) or detections in which false positive errors are not possible (confirmed detections). Through simulations, Miller *et al.* (2011) demonstrated that their models generated occupancy probability estimates that were more accurate and precise than the Royle-Link model and the MacKenzie *et al.* (2002) model, which assumes there are no false positives.

While the Miller models appear to successfully address the problem of symmetry in the Royle-Link model and avoid assumptions about the magnitude of the true-positive detection probability relative to the false positive detection probability, their ability to identify 'phantom' species and explicitly model heterogeneity among sites and surveys has not been fully evaluated. Miller *et al.* (2011) investigated whether estimation of the false positive detection probability would be affected if data had heterogeneity in the true-positive detection probability, but none of their models estimated site- or survey-specific probabilities as functions of covariates. In Miller *et al.* (2013), the Miller model was used to model a small amount of heterogeneity in occupancy probabilities among sites, but there were no simulations to evaluate and describe model performance in this application.

The goals of our work were to develop and evaluate an occupancy model that (i) accounts for false negative and false positive detections and (ii) models the large amount of heterogeneity in occupancy and detection that may be observed in complex ecological data. Since distinguishing false negative and false positive detections may be challenging when there is heterogeneity in occupancy and detection and few assumptions about parameter values, we built our models in a Bayesian framework (unlike the Royle-Link and Miller models) so that we could investigate the potential for informative priors to improve inference. Further, complex models with latent variables have been highlighted as well suited to a hierarchical Bayesian modelling approach in which explicit state and detection model components are developed (Link & Barker 2010).

After developing the basic model structure, we used Monte Carlo simulations with realistic data scenarios to evaluate model performance using vague or informative priors. We simulated data using a range of parameter values, and we simulated scenarios where the false positive probability was greater than the true-positive probability and where there was a 'phantom' species. We describe how our models might improve estimator accuracy and inference about the relationship between environmental or anthropogenic factors and occupancy or detection probabilities. Several examples of possible applications are presented, and we describe in detail an application of our model to study the relationship between migratory songbird occupancy and residential development.

Model description

Ecological data are often not free from false positive errors (McClintock *et al.* 2010a; Miller *et al.* 2012), so a model that can make inference about occupancy and its relationship with covariates when there are false positive and false negative errors is needed. Previous work has highlighted some of the challenges with building such a model: being able to distinguish false positive and false negative processes, modelling heterogeneity in occupancy and detection, and determining 'phantom species'. In this study, we describe a model that can handle these challenges and test its performance in a variety of simulated scenarios. We developed a hierarchical Bayesian model based on the Miller MDS model that accounts for heterogeneity in occupancy, false positive detection and false negative detection probabilities.

DATA

False positive errors are not rare in ecological data, and to distinguish false positives and false negatives without restrictive assumptions about parameter values, additional data in the form of confirmed observations are needed (McClintock *et al.* 2010a,b; Miller *et al.* 2011, 2012). So the data input for the model are a detection/non-detection matrix and a confirmed/unconfirmed matrix, where the matrices are indexed by the number of sites and the number of surveys per site. In

Miller's MDS model, non-detections were always unconfirmed (false negative errors were possible), while detections could be unconfirmed (false positive errors were possible) or confirmed (false positive errors were not possible). Data of this variety could be collected in many ecological applications, for example when species are sampled indirectly (i.e. by vocalization, hair, faeces) and/or the study organism is cryptic, mobile or resembles other species.

An auditory survey of frog calls or a visual survey of freshwater mussels (Shea *et al.* 2011) could results in data with false positive and false negative detections. A method to obtain confirmed presences either could be applied to all surveys or to a subset of surveys. When the primary sampling method is an auditory survey, researchers could also perform visual inspections, either through binoculars or by catching individuals. Species that were detected by sound and sight would have confirmed presences, and species that were only detected by sound would have unconfirmed presences. If a species was detected by sight only, researchers could determine whether this constituted a confirmed or unconfirmed presence, depending on how distinguishable the species is and the quality (duration, distance) of the sighting. Alternatively, researchers could rely exclusively on auditory detections but could have multiple independent observers in the field or could record the survey for inspection by observers in the laboratory. A minimum proportion of observers detecting a species could be treated as a threshold to confirm a presence. This threshold could be identified by having observers identify the species in the playback of an artificial survey where all the species are known to the researcher who created the recording plus simulating how the observed false positive and false negative detection rates affect model results. To obtain confirmed presences when the primary sampling method is a visual survey, researchers could use multiple field observers as discussed above. For all survey methods, DNA analysis can also provide confirmed presences.

In other applications, a proportion of the data may have confirmed absences (false negative errors are not possible) in addition to confirmed presences. For example, if a distinctive plant was being studied and a subset of all sites were completely surveyed (Falster, Murray & Lepschi 2001) or if a test that effectively determines the presence or absence of a disease with certainty was used with a subset of the study population (Feigelson *et al.* 1994), data with both confirmed absences and confirmed presences could be obtained. Therefore, we constructed occupancy models for circumstances in which there are confirmed and unconfirmed absences and confirmed and unconfirmed presences (hereafter, the confirmed absences and presences model or CACP model) and for applications where there are only unconfirmed absences but confirmed and unconfirmed presences (hereafter, the confirmed presences model or CP model).

MODEL STRUCTURE

As with other occupancy models, our models assume that the occupancy state does not change within a season and that detections at each site and at multiple visits to a site are independent. Our models also assume that observation confirmation is independent across sites and surveys, that is whether an observation is confirmed or unconfirmed during a survey is independent of the confirmation state during previous survey(s).

Each of the $i = 1, 2,\ldots, R$ sites is occupied ($z_i = 1$) or not ($z_i = 0$). Whether a site is occupied can be considered the realization of a Bernoulli trial with probability of occupancy, ψ ($z_i \sim \text{Bern}(\psi)$). The occupancy probability can be constant across sites or vary depending on site-specific covariates, which are incorporated in the model through a logit-linear model.

At an occupied site, a true-positive detection may occur on sampling occasion $t = 1, 2,\ldots,T$ with probability P_{11}, or a false negative detection may occur with probability ($1–P_{11}$). At an unoccupied site, a false positive detection may occur with probability P_{10}, or a true negative detection may occur with probability ($1–P_{10}$). The true-positive detection probability or false positive detection probability may be constant across sites and sampling occasions or may vary depending on covariates.

We simulated data for the CACP model assuming that the observation confirmation probability was the same for confirmed absences and confirmed presences. This is not required but served as the starting point for model development. A site is known to be occupied when there is a confirmed detection. However, when a detection is unconfirmed, the site's occupancy state is unknown. To infer the latent occupancy state, the CP model required a parameter, b, to describe the probability of a confirmed detection at an occupied site. Then, $1–b$ is the probability of an unconfirmed detection at an occupied site. At an unoccupied site, the probability of a confirmed detection is 0. Whether a detection was confirmed was modelled as the realization of a Bernoulli trial with confirmation probability, $z_i{*}b$. Detections (y_{it}) were modelled as outcomes of Bernoulli trials (Fig. 1). More details are discussed in Appendix S1 (Supporting information), and code is provided in Appendix S2.

In summary, the unknown latent state is whether the site is occupied ($z_i = 1$) or not ($z_i = 0$), and unknown parameters are the occupancy probability (ψ), true-positive detection probability (P_{11}), false positive detection probability (P_{10}), observation confirmation probability (b), and intercepts and coefficients for any logit-linear models incorporating covariates. The data are whether the species was detected ($y_{it} = 1$)

(a)

	$z_i = 0$	$z_i = 1$
$c_i = 0$	p_{10}	p_{11}
$c_i = 1$	0	1

(b)

	$z_i = 0$	$z_i = 1$
$c_i = 0$	p_{10}	p_{11}
$c_i = 1$	Undefined	1

Fig. 1. Probabilities of detections in (a) the confirmed absences and confirmed presences (CACP) model and (b) the confirmed presences (CP) model given occupancy (z) and observation confirmation (c) states at site i and false positive (P_{10}) and true-positive (P_{11}) detection probabilities.

or not ($y_{it} = 0$), whether observations were confirmed ($c_{it} = 1$) or not ($c_{it} = 0$), and the values of any covariates in the model.

Simulation study: methods

EVALUATING MODEL PERFORMANCE

To evaluate the performance of the CACP and CP models, we simulated data under a variety of scenarios (Table 1). For each scenario, 100 data sets, each with three visits to 250 sites, were simulated using the same parameter values. For each data set, we fit an occupancy model, assessed a parameter's bias by calculating the absolute error as the difference between the mean of the parameter's posterior distribution and the parameter value used to simulate data and calculated the width of 95% Bayesian credible intervals (BCIs). For each scenario, we calculated the number of model runs that converged and the per cent of converged model runs in which the BCI contained the parameter value used to simulate data.

Data were simulated in R version 2.15.3 (R Development Core Team 2013), and models were run in OPENBUGS version 3.2.2 using the R2OPENBUGS package (Sturtz, Ligges & Gelman 2005; Lunn *et al.* 2009). We ran three Markov Chain Monte Carlo (MCMC) chains with 100 000 iterations, a burn-in of 50 000 and thinning of one (Brooks & Gelman 1998; Link & Eaton 2011). Convergence was assessed with the Gelman–Rubin potential scale reduction factor (R-hat), and chains were considered converged if R-hat ≤ 1·04 (Brooks & Gelman 1998).

BASIC SCENARIOS WITHOUT HETEROGENEITY

First, we evaluated the CACP and CP models without heterogeneity in occupancy or detection (Table 1). We simulated data in 12 scenarios with various probabilities (Table 2). The observation confirmation probability used to generate data was always 0·03. It may be difficult to confirm observations in the field, so it is desirable for the models to generate accurate and precise posterior probabilities with a low rate of observation confirmation.

Priors for the occupancy probability and true-positive detection probability were always vague: U(0, 1) or Beta(0·5, 0·5), but we investigated model performance with vague (U(0, 0·5)) and informative (Beta(1, 9)) priors for the false positive detection probability (Appendix S1). Using an informative prior may improve the precision of posterior distributions and avoid parameter identifiability problems. We constrained the false positive probability to be <0·5, which suggests that if a site is unoccupied, an observer is more likely to make a true negative detection than a false positive detection. This constraint was consistent with the probability of false positive detections seen in controlled experiments (Farmer, Leonard & Horn 2012; Miller *et al.* 2012) and could aid model convergence. We also considered CP model performance with vague (U(0, 1) or Beta(0·5, 0·5)) or informative (U(0·01, 0·05) or Beta(10, 300)) priors for the observation confirmation probability.

SPECIAL CASES WITHOUT HETEROGENEITY

We simulated data with errors in confirmed observations to evaluate model robustness (Table 1). For the CACP model, there was a 0·006 probability of a confirmed observation error (presence or absence), which corresponded to a 0·1 probability of an error given a confirmed observation. For the CP model, there was a 0·005 probability of a confirmed presence when the site was actually unoccupied (Appendix S1). We also evaluated the CACP and CP models in a scenario with a 'phantom' species, which had zero probability of occupancy and true-positive detection but a positive probability of false positive detection (Table 1).

Table 1. Scenarios that were simulated to evaluate the confirmed absences and confirmed presences (CACP) and confirmed presences (CP) model, where P_{10} is the false positive detection probability and b is the observation confirmation probability. CACP and CP model performance was compared to that from a model assuming there were no false positive errors although simulated data always contained false positive errors. Columns in the table describe the main types of simulated scenarios. Rows in the table describe features of the simulated scenarios. Entries in the table indicate whether the CACP and/or CP model was applied to the condition described by the respective column and row. For example, the first column indicates that the CACP model was applied to scenarios without heterogeneity that were simulated using 12 parameter value combinations and the model was evaluated with a vague prior for P_{10} and with an informative prior for P_{10}

Conditions	No heterogeneity	Observation confirmation errors	'Phantom' species	Heterogeneity	No false positives
12 parameter value combinations	CACP, CP	CACP, CP			CACP, CP
Informative prior for P_{10}	CACP, CP	Best prior from no heterogeneity scenarios– CACP, CP	Best prior from no heterogeneity scenarios– CACP, CP	CACP, CP	
Vague prior for P_{10}	CACP, CP	Best prior from no heterogeneity scenarios– CACP, CP	Best prior from no heterogeneity scenarios– CACP, CP	CACP, CP	
Informative prior for b	CP			No observation confirmation errors– CP	
Vague prior for b	CP	CP	CP	CP	
Observation confirmations errors		CACP, CP	CACP, CP	CACP, Vague prior for b– CP	
'Phantom' species			CACP, CP		CACP, CP
Three sets of intercept and coefficient values				CACP, CP	CACP, CP

Table 2. Scenarios with no heterogeneity among sites or surveys. Data with various probabilities of occupancy (ψ), true-positive detection (P_{11}) and false positive detection (P_{10}) were simulated to evaluate occupancy models

Scenario	ψ	P_{11}	P_{10}
1	0·7	0·6	0·05
2	0·7	0·6	0·15
3	0·7	0·2	0·05
4	0·7	0·2	0·15
5	0·3	0·6	0·05
6	0·3	0·6	0·15
7	0·3	0·2	0·05
8	0·3	0·2	0·15
9	0·7	0·02	0·05
10	0·7	0·1	0·15
11	0·3	0·02	0·05
12	0·3	0·1	0·15

SCENARIOS WITH HETEROGENEITY

We evaluated the CACP and CP models when there was heterogeneity in occupancy, true-positive detection and false positive detection probabilities (Table 1). We simulated data with one explanatory variable (per cent forest cover) affecting occupancy probabilities and one explanatory variable (temperature) affecting true-positive detection probabilities. We simulated data using three sets of intercept and coefficient values to represent strong quadratic, weak quadratic and linear effects of the covariate on occupancy probabilities (Table 3, Fig. S1). In all three scenarios, there was a linear function relating temperature to true-positive detection probabilities. We also simulated false positive detection probabilities that decreased over three time periods, representing an observer gaining experience.

Covariate data were standardized to have a mean of zero and variance of one to aid convergence. Since we modelled the effects of covariates through a logit-linear equation, all covariate coefficients had a N (0, 0·368) prior, which is a vague Jeffrey's prior for a parameter on the logit scale (Lunn *et al.* 2012). Intercept terms had vague priors (U(0, 1) or Beta (0·5, 0·5)) and were logit-transformed before inclusion in the logit-linear equation.

Table 3. Scenarios where occupancy and detection probabilities exhibited heterogeneity among sites and surveys. Occupancy probabilities were simulated through a logit-linear model with intercept (psi_{int} before undergoing a logit transformation), coefficient for the linear term (a_1) and coefficient for the quadratic term (a_2). True-positive detection probabilities were simulated through a logit-linear model with intercept ($P_{11\text{int}}$ before undergoing a logit transformation) and coefficient for the linear term (c_1). False positive detection probabilities were simulated for three time periods (P_{10t1}, P_{10t2}, P_{10t3}), and the observation confirmation probability was b

	Strong quadratic	Weak quadratic	Linear
psi_{int}	0·7	0·7	0·7
a_1	1·5	2·6	3·2
a_2	−1·8	−0·4	NA
$p_{11\text{int}}$	0·5	0·5	0·5
c_1	−2	−1·2	−2·5
P_{10t1}	0·1	0·1	0·1
P_{10t2}	0·07	0·07	0·07
P_{10t3}	0·04	0·04	0·04
b	0·03	0·03	0·03

MODEL ASSUMING NO FALSE POSITIVE DETECTIONS

We compared the performance of the CACP and CP models to a Bayesian parameterization of the MacKenzie *et al.* (2002) model that assumed false positive detections did not occur (no false positive model) and assessed improvement in accuracy and preci-

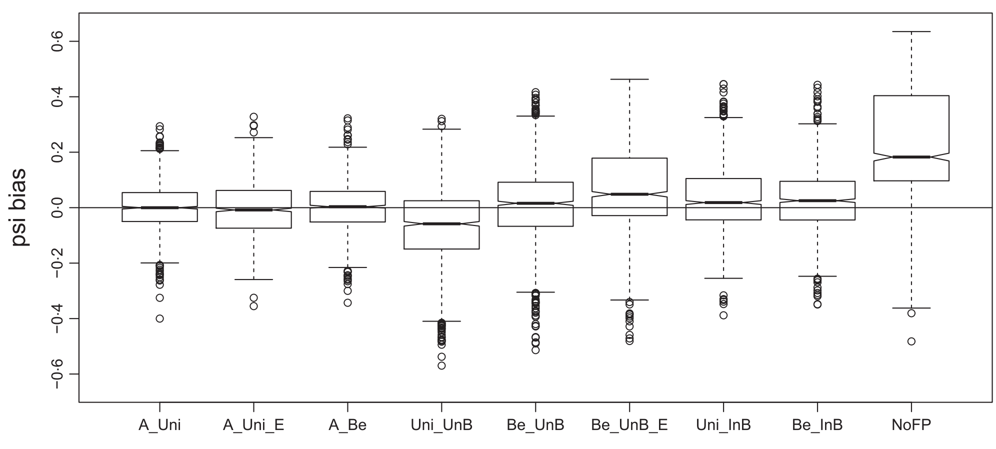

Fig. 2. Bias of posterior occupancy probabilities (ψ) from models in scenarios with no heterogeneity. Bias was calculated as the difference between the mean of the parameter's posterior distribution and the value used to simulate data. Results are presented from all converged model runs: A = data with confirmed absences in addition to confirmed presences, otherwise data only have confirmed presences; E = data with observation confirmation errors, otherwise data have no observation confirmation errors; Uni = vague prior for the false positive detection probability; Be = informative prior for the false positive detection probability; UnB = vague prior for the observation confirmation probability; InB = informative prior for the observation confirmation probability; and NoFP = model assuming there were no false positive detections although data always contained false positive errors. If the notches around the medians of two plots do not overlap, there is strong evidence that the medians differ. The box includes the first through third quartile. Whiskers extend to the most extreme data point that is no more than 1·5 times the interquartile range from the box. Small circles represent outliers.

sion that resulted from accounting for both types of imperfect detection (Table 1).

Simulation study: results

SCENARIOS WITHOUT HETEROGENEITY

The CACP and CP models generated more accurate posterior distributions than the no false positive model when data contained false positive detections (Fig. 2). Posterior occupancy probabilities from the no false positive model were biased high and imprecise. The CACP models consistently generated unbiased posterior occupancy probabilities, and the CP models with at least one informative prior generated posterior occupancy probabilities that were essentially unbiased when there were no observation confirmation errors. Posterior occupancy probabilities from the CP model were biased low when the model had only vague priors (Fig. 2) and biased high when data had observation confirmation errors (Fig. 2, Tables S1–S4 and S6).

When data contained detections of 'phantom' species, the most accurate and precise posterior occupancy probabilities resulted when there were confirmed absences and no observation confirmation errors, and posterior occupancy probabilities were biased slightly higher when there were no confirmed absences (Fig. 3). Posterior false positive detection probabilities from the CACP and CP models were accurate and precise (Tables S5 and S6).

SCENARIOS WITH HETEROGENEITY

In all scenarios with heterogeneity, the CACP and CP models generated more accurate posterior occupancy probabilities than the no false positive model. All CACP and CP models generated essentially unbiased posterior occupancy probabilities in the strong quadratic, weak quadratic and linear scenarios (Figs 4–6 and S2–7). Posterior occupancy probabilities from the no false positive model were biased high and imprecise (Figs 4–6), and occupancy probability BCIs often did not contain the probabilities used to simulate data (Tables S7–9).

The no false positive model also generated more biased covariate coefficient posterior distributions than the CACP or CP models. In the strong quadratic scenario, posterior distributions for covariate coefficients were unbiased from all CACP and CP models when there were no observation confirmation errors (Figs 7 and 8). The CACP model with vague priors generated unbiased covariate coefficient posterior distributions in the weak quadratic and linear scenarios (Figs 9–11). The CP model with an informative prior for the false positive detection probabilities and vague prior for the observation confirmation probability generated unbiased covariate coefficient posterior distributions in the weak quadratic scenario (Figs 9 and 10). With an informative prior also for the observation confirmation probability and both priors vague, the CP model generated unbiased covariate coefficient posterior distributions in the linear scenario.

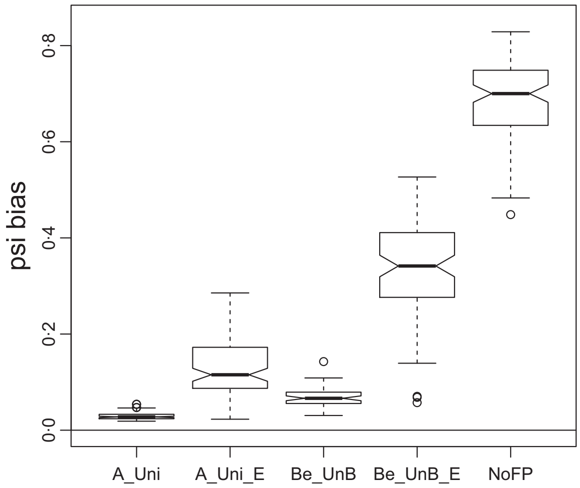

Fig. 3. Bias of posterior occupancy probabilities (ψ) from models in scenarios with a 'phantom' species and no heterogeneity. A 'phantom' species had zero probability of occupancy and true-positive detection but a positive probability of false positive detection. Additional box plot details can be found in the Fig. 2 legend.

Simulation study: discussion

Classic occupancy models have assumptions that are violated in many ecological data sets, and estimates from these models are not robust (Miller *et al.* 2011, 2012). In order to generate accurate estimates, an occupancy model needs to address both false positive and false negative detections, but additional data, such as confirmed presences, are needed to differentiate these processes (McClintock *et al.* 2010a,b). Confirmed absences may further improve estimation, but depending on the application, it may not be possible to obtain these data. In these cases, inference may be improved by incorporating existing knowledge through informative priors.

Monte Carlo simulations demonstrated that the CACP and CP models generate accurate inference while overcoming some of the limitations of existing occupancy models. Namely, our models distinguish false positive and false negative detection processes, explicitly model high levels of heterogeneity in occupancy and detection probabilities, minimize assumptions about parameter values and can identify 'phantom' species. In particular, suspected 'phantom' species may be identified if the posterior occupancy probability and the upper bound of its BCI are small and the BCI for the true-positive detection probability is very wide. Our models are also robust to errors in confirmed observations. Conducting a simulation study such as ours reveals model behaviour given data of different qualities and quantities (with confirmed absences and confirmed presences, without confirmed absences but with confirmed presences, with and without observation confirmation errors, with and without covariates), so researchers can anticipate the strengths and limitations of the model in various applications (Peck 2004).

Considering patterns in the results presented here can help researchers evaluate the CACP and CP models. Accurate posterior distributions resulted from the CACP model with vague

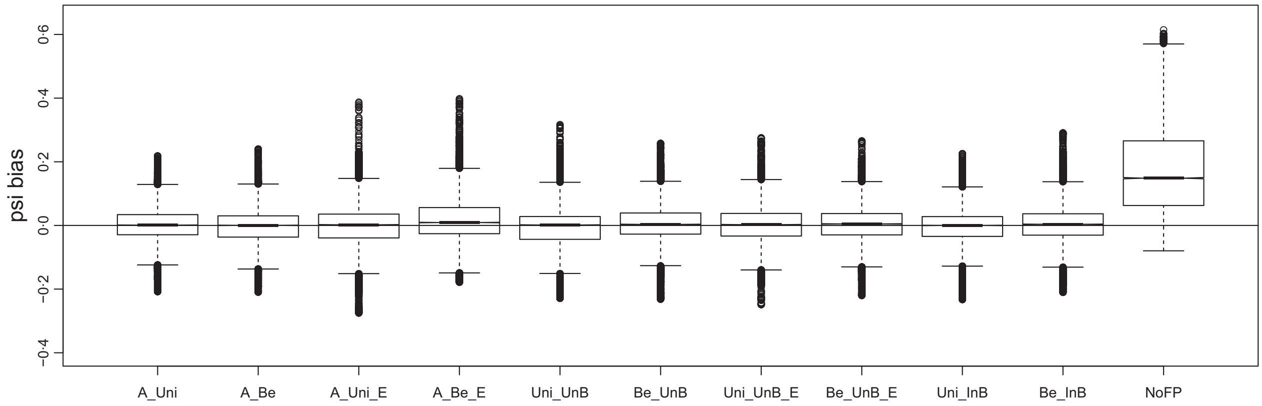

Fig. 4. Bias of posterior occupancy probabilities (ψ) from models in the scenario where a strong quadratic function related occupancy probabilities to a covariate. Additional box plot details can be found in the Fig. 2 legend.

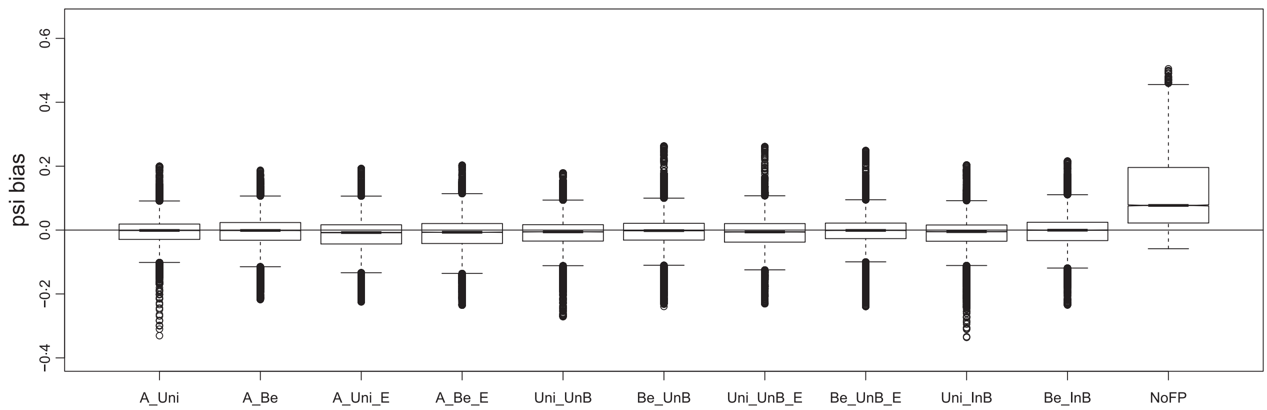

Fig. 5. Bias of posterior occupancy probabilities (ψ) from models in the scenario where a weak quadratic function related occupancy probabilities to a covariate. Additional box plot details can be found in the legend for Fig. 2.

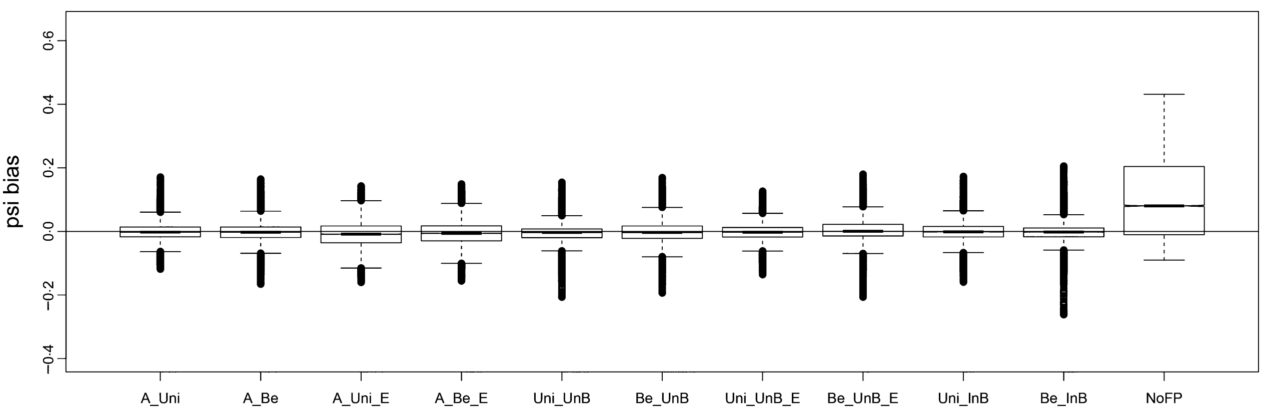

Fig. 6. Bias of posterior occupancy probabilities (ψ) from models in the scenario where a linear function related occupancy probabilities to a covariate. Additional box plot details can be found in the legend for Fig. 2.

priors because data had both confirmed presences and confirmed absences. We expect that posterior occupancy probabilities from the CP model with vague priors and no heterogeneity were biased low because the data did not contain enough information to generate accurate estimates. Additional information in the form of an informative prior or confirmed absences resulted in accurate estimates. We can consider specifically why posterior occupancy probabilities were biased low and the observation confirmation probability was biased high in the CP model with vague priors. A matrix of 0s and 1s indi-

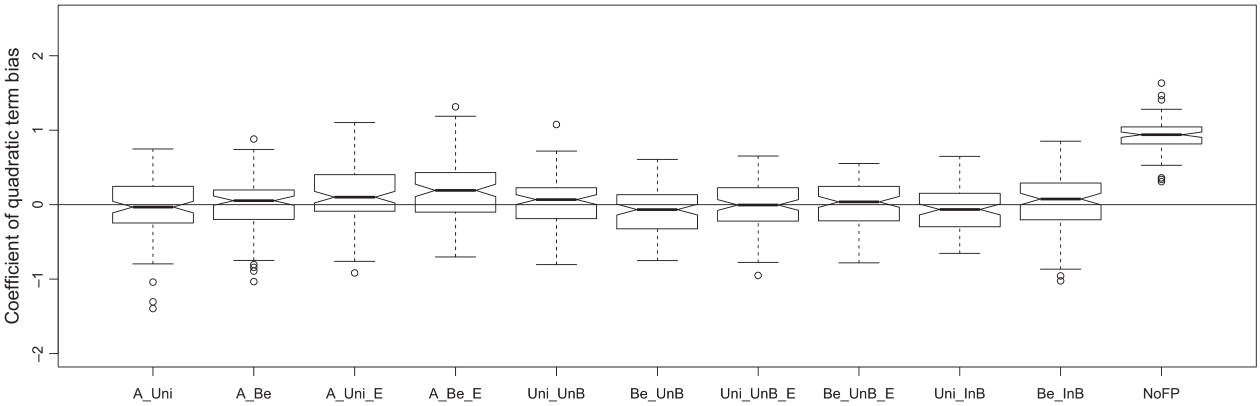

Fig. 7. Bias of the coefficient for the quadratic term in the scenario where a strong quadratic function related occupancy probabilities to a covariate. Bias in the coefficient is presented from all converged model runs. Box plot details can be found in the legend for Fig. 2.

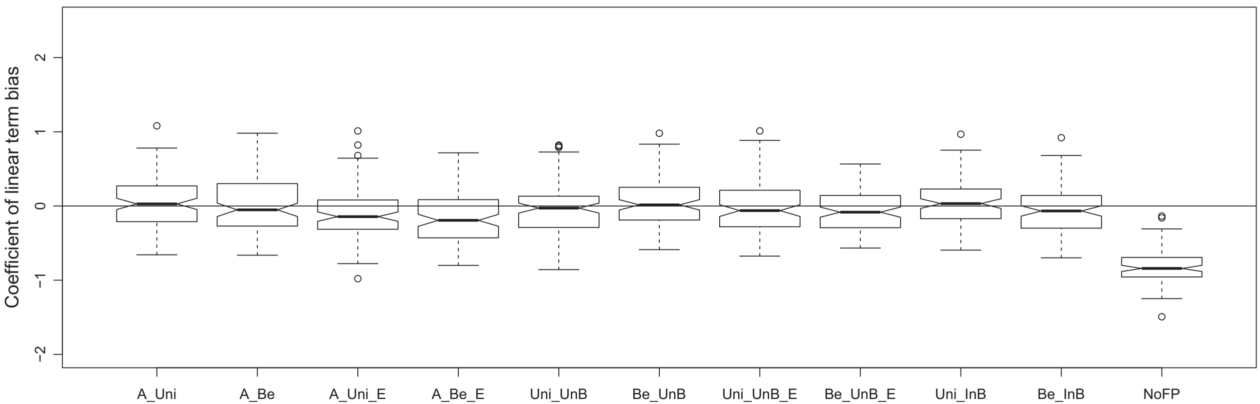

Fig. 8. Bias of the coefficient for the linear term in the scenario where a strong quadratic function related occupancy probabilities to a covariate. Bias in the coefficient is presented from all converged model runs. Box plot details can be found in the legend for Fig. 2.

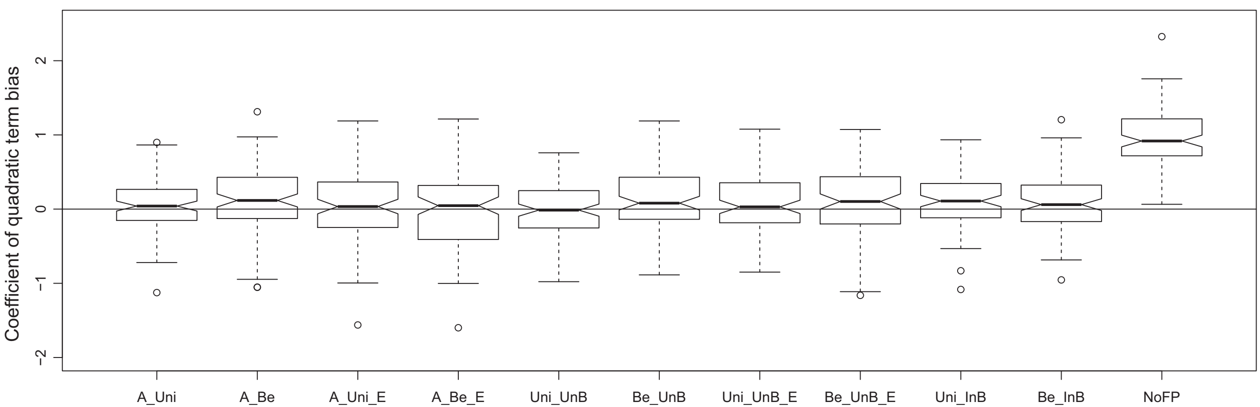

Fig. 9. Bias of the coefficient for the quadratic term in the scenario where a weak quadratic function related occupancy probabilities to a covariate. Bias in the coefficient is presented from all converged model runs. Box plot details can be found in the legend for Fig. 2.

cating whether there was a confirmed observation for each survey at each site was input into the model. In the model, a confirmed observation has probability z_i*b (0 if the site is unoccupied and b if the site is occupied), so the confirmed observation matrix contributes to inference about b and z_i. Given a confirmed observation matrix, if b is overestimated, z_i may be inferred to be 0 too often, which could contribute to underestimating occupancy probabilities. An informative prior for b or an informative prior for P_{10}, which affects inference about occupancy probabilities, tended to improve inference about b, P_{10} and ψ_i. Posterior occupancy probabilities from the no false positive model were biased high

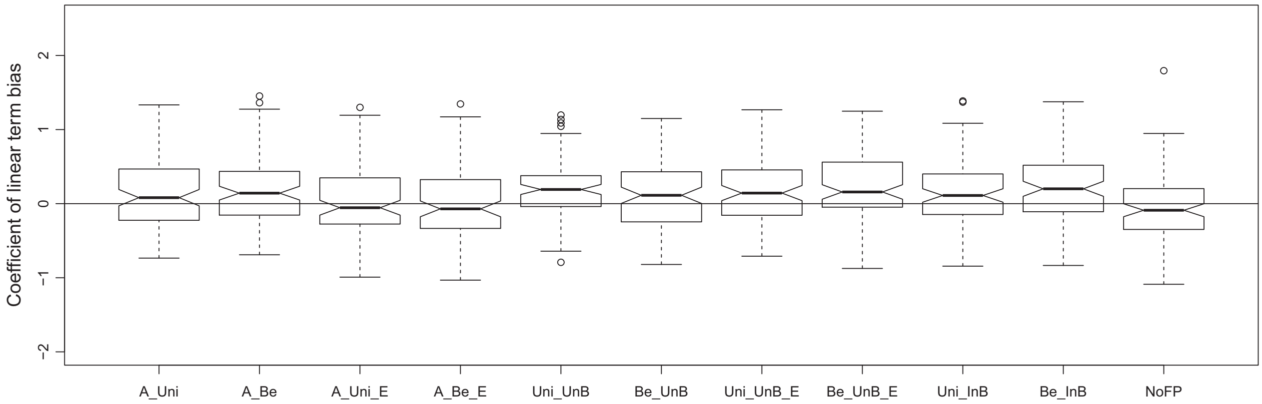

Fig. 10. Bias of the coefficient for the linear term in the scenario where a weak quadratic function related occupancy probabilities to a covariate. Bias in the coefficient is presented from all converged model runs. Box plot details can be found in the legend for Fig. 2.

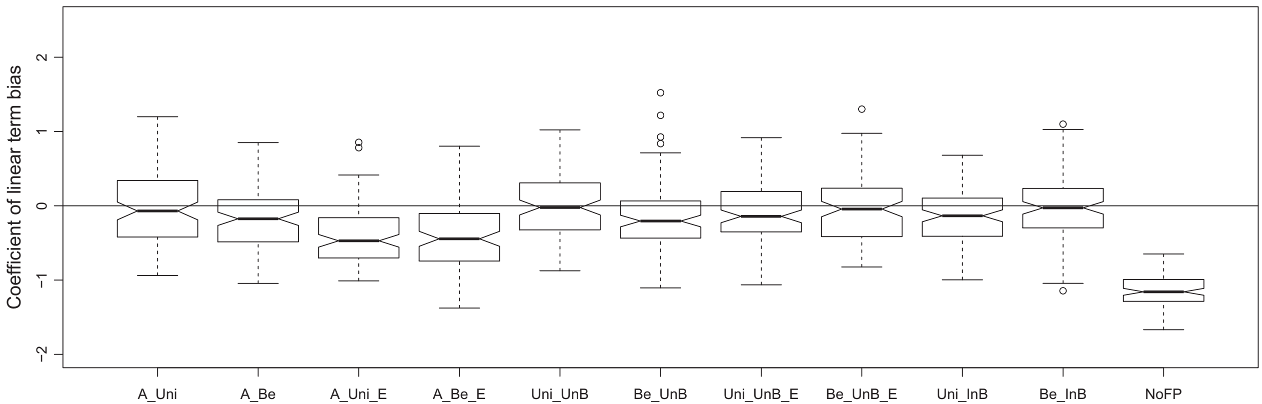

Fig. 11. Bias of the coefficient in the scenario where a linear function related occupancy probabilities to a covariate. Bias in the coefficient is presented from all converged model runs. Box plot details can be found in the legend for Fig. 2.

because false positive detections were treated as true positives. Similarly, when data had observation confirmation errors, posterior occupancy probabilities from the CP model were biased high because some false positive detections were treated as confirmed true positives.

In the scenarios with heterogeneity, all CACP and CP models generated unbiased posterior occupancy probabilities, but some posterior covariate coefficients seemed biased based on boxplots of the parameter's absolute error. However, there was not strong evidence, based on the notches around boxplot medians, that the bias was different among the CACP and CP models when there were no observation confirmation errors. Even for models where the notches nearly did not overlap, the CP model with vague priors and the CP model with an informative prior for P_{10} and vague prior for b in the linear scenario, the 95% BCIs for the covariate coefficients from these two models overlapped in all 100 simulations. Additionally, Figs S2–S7 do not indicate severely biased inference about the relationship between occupancy probabilities and covariates. The CACP and CP models also were similarly accurate for all other parameters. So we conclude that these models performed adequately.

In summary, biases in results from the CP model can be explained, and efforts can be made to reduce the potential for these biases by developing informative priors, collecting confirmed absence data or eliminating observation confirmation errors. The CACP and CP models are useful additions to the available suite of occupancy models because they can generate accurate inference when data have complexities that traditional models cannot accommodate but are anticipated to occur in ecological field data (false positive and false negative errors, heterogeneity in occupancy and detection probabilities, and detections of phantom species).

Application

EXAMPLE

Based on the simulations presented here, we selected the CP model with vague priors to model the relationship between passerines and residential development. Urban development and exurban development are considered principal causes of world-wide habitat loss (Brown *et al.* 2005; Hansen *et al.* 2005). We focused on development in the southern Appalachian region, specifically Macon County, North Carolina. Macon County has a relatively high per cent forest cover, but amenity-driven residential development has contributed to forest fragmentation, especially at higher elevations on previously undeveloped slopes (Gragson & Bolstad 2006). For this example, we will consider one model describing the relationship between black-and-white warbler (*Mniotilta varia*; BAWW) occupancy and forest cover. The BAWW is a forest-dwelling, insectivorous, Neotropical migrant.

We conducted point counts and collected forest cover data at 272 sites representing a range of elevations and land-use/land-cover classes around Macon County (doi: 10.5061/dryad.t68v8). Each site was surveyed three times during the breeding season in either 2010 or 2011. Two independent observers conducted eight minute point counts between twilight and 10:00 am. Each observer recorded the species they heard or saw and indicated whether the species was within 25 m and/or between 25 and 100 m. We considered detections to be confirmed when, during a point count, both independent observers detected a species by sight and/or sound within 25 m. We determined the per cent forest cover within 12·5 ha (circle with 200 m radius) of point count sites by digitizing aerial photographs of Macon County in ArcGIS using National Land Cover Dataset 2001 classes. We standardized the per cent forest cover data to have a mean of zero and variance of one. We modelled a constant observation confirmation probability and year-specific true-positive and false positive detection probabilities because observers differed between the 2 years. We ran models in OPENBUGS version 3.2.2 (Lunn *et al.* 2009) using the R2OPENBUGS package and R version 2.15.3 (Sturtz, Ligges & Gelman 2005, R Development Core Team 2013). We ran three MCMC chains with at least 100 000 iterations, a burn-in of at least 50 000 and thinning of 5.

The BAWW was detected 157 times, and 23 of those detections were confirmed presences. In the 2 years surveyed, the true-positive detection probabilities were 0·28 (95% BCI: 0·18–0·42) and 0·40 (0·32–0·49), and the false positive detection probabilities were 0·01 (0·00–0·03) and 0·08 (0·04–0·13). The observation confirmation probability was 0·07 (0·04–0·11). Occupancy probabilities increased with increasing per cent forest and ranged 0·01–0·90 (Fig. 12).

Our results indicate that data included false positive detections. This is not unexpected as the BAWW song can resemble that of the American redstart. When we accounted for false positive errors, we estimated that having more than a 0·5 probability of BAWW occupancy was associated with more than 70% forest cover within 12·5 ha (Fig. 12). These findings provide insights into BAWW distribution, the effects of development and conservation strategies. Understanding the relationship between BAWW occupancy and forest cover can help inform decision-making about land use and development. If we had assumed there were no false positive detections, biased results would have suggested that occupancy was higher at low per cent forest cover than was true (Figs 12 and S4). This would downplay the importance of forest cover for the BAWW and could lead to inappropriate management decisions.

SUGGESTIONS

Regardless of the specific application, it is prudent to include a false positive detection probability in the model. Evidence suggests that ecological data typically are not free from false positive errors (Miller *et al.* 2012), and the bias resulting from ignoring even a small number of false positive errors is likely to be larger than the bias from modelling false positive detections when there are none (Fig. S8). If data do not have false positive errors, but the model includes a false positive detection probability, the model is expected to generate a posterior false positive probability with a BCI of *c.* 0–0·1. Posterior occupancy probabilities may be slightly biased low, while posterior true-positive detection probabilities may be slightly biased high (Fig. S8).

To apply our models (Appendix S2), a researcher needs a detection/non-detection matrix (indexed by the number of sites and the number of surveys per site), a confirmed/unconfirmed matrix, any relevant covariate data and prior distributions for unknown values in the model. When selecting sites to sample, researchers can consider that with a smaller sample size, posterior distributions can be expected to be more biased, but the patterns of relative model performance presented here should not change (Fig. S9). A researcher can collect detection histories as in a traditional occupancy study, but at least some data with confirmed presences are required to distinguish false positive and false negative detection processes. Some examples of data collection procedures were presented in the 'Data' section of the study. Any method for making confirmed observations is suitable if the probability of having an error is small enough that model results would be robust to errors. Researchers may have information about how reliable an identification method is in their study system, or they could perform an identification experiment as in

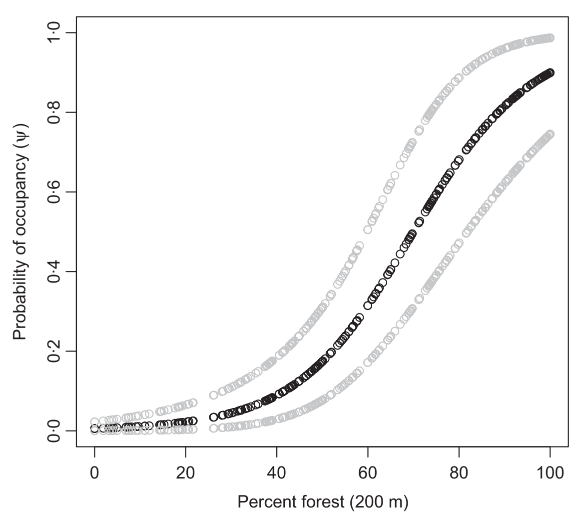

Fig. 12. Posterior occupancy probabilities for the Black-and-white Warbler ordered by the per cent forest within 200 m of point count sites. For each point count site, the mean of the posterior distribution is presented in black and the 95% Bayesian credible interval is in grey.

McClintock *et al*. (2010a) to estimate the identification method's probability of error. Then, they could use the code provided in Appendix S2 to verify that accurate inference can be made with this probability of error. We also recommend the use of simulations to explore how informative priors may affect posterior distributions given the expected sample size and range of detection and occupancy probabilities.

Conclusion

Our results confirm the importance of accounting for false positive detections in occupancy models and illustrate the ability of the CACP and CP models to do so with complex ecological data (heterogeneity in occupancy and detection probabilities, detections of 'phantom species', false positive detection probabilities greater than true-positive detection probabilities). False positive detections occur in field data, and if they are ignored, inference about species occupancy and how it relates to environmental factors will be biased. The CACP and CP models are robust and flexible occupancy models, as evidence by the range of biologically realistic scenarios we simulated and applications we discussed, that overcome limitations in existing occupancy models.

Acknowledgements

We are grateful to Drs. J.F. Chamblee, R.J. Cooper and J.C. Maerz for feedback on this project and Dr. D.A.W. Miller for suggestions about the manuscript. Funding was provided by the Coweeta Long Term Ecological Research project (NSF grant DEB-0823293), USDA CSREES McIntire-Stennis Project (GEOZ-0159-MS), Warnell School of Forestry and Natural Resources at the University of Georgia, University of Georgia Graduate School, Georgia Ornithological Society and Georgia Museum of Natural History.

Data accessibility

The point count data are archived in Dryad http://dx.doi.org/10.5061/dryad.t68v8. The R code for the occupancy model is archived in Appendix S2 of the manuscript's supporting information.

References

Alldredge, M.W., Pacifici, K., Simons, T.R. & Pollock, K.H. (2008) A novel field evaluation of the effectiveness of distance and independent observer sampling to estimate aural avian detection probabilities. *Journal of Applied Ecology*, **45**, 1349–1356.

Brooks, S.P. & Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.

Brown, D.G., Johnson, K.M., Loveland, T.R. & Theobald, D.M. (2005) Rural land-use trends in the conterminous United States, 1950-2000. *Ecological Applications*, **15**, 1851–1863.

Falster, D.S., Murray, B.R. & Lepschi, B.J. (2001) Linking abundance, occupancy and spatial structure: an empirical test of a neutral model in an open-forest woody plant community in eastern Australia. *Journal of Biogeography*, **28**, 317–323.

Farmer, R.G., Leonard, M.L. & Horn, A.G. (2012) Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. *The Auk*, **129**, 76–86.

Feigelson, H.S., Criqui, M.H., Fronek, A., Langer, R.D. & Molgaard, C.A. (1994) Screening for peripheral arterial-disease – the sensitivity, specificity, and predictive value of noninvasive tests in a defined population. *American Journal of Epidemiology*, **140**, 526–534.

Ferguson, P., Conroy, M. & Hepinstall-Cymerman, J. (2015) Data from: Occupancy models for data with false positive and false negative errors and heterogeneity across sites and surveys. *Methods in Ecology and Evolution*, http://dx.doi.org/10.5061/dryad.t68v8.

Fitzpatrick, M.C., Preisser, E.L., Ellison, A.M. & Elkinton, J.S. (2009) Observer bias and the detection of low-density populations. *Ecological Applications*, **19**, 1673–1679.

Gragson, T.L. & Bolstad, P.V. (2006) Land use legacies and the future of Southern Appalachia. *Society and Natural Resources*, **19**, 175–190.

Hansen, A.J., Knight, R.L., Marzluff, J., Powell, S., Brown, K., Gude, P.H. & Jones, K. (2005) Effects of exurban development on biodiversity: patterns, mechanisms, and research needs. *Ecological Applications*, **15**, 1893–1905.

Link, W.A. & Barker, R.J. (2010) *Bayesian Inference: With Ecological Applications*. Academic Press, Burlington.

Link, W.A. & Eaton, M.J. (2011) On thinning of chains in MCMC. *Methods in Ecology and Evolution*, **3**, 112–115.

Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. (2009) The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, **28**, 3049–3067.

Lunn, D., Jackson, C., Best, N., Thomas, A. & Spiegelhalter, D. (2012) *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Chapman & Hall/CRC, Boca Raton.

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

McClintock, B.T., Bailey, L.L., Pollock, K.H. & Simons, T.R. (2010a) Experimental investigation of observation error in Anuran call surveys. *Journal of Wildlife Management*, **74**, 1882–1893.

McClintock, B.T., Bailey, L.L., Pollock, K.H. & Simons, T.R. (2010b) Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, **91**, 2446–2454.

Miller, D.A., Nichols, J.D., McClintock, B.T., Campbell Grant, E.H., Bailey, L.L. & Weir, L.A. (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.

Miller, D.A.W., Weir, L.A., McClintock, B.T., Campbell Grant, E.H., Bailey, L.L. & Simons, T.R. (2012) Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecological Applications*, **22**, 1665–1674.

Miller, D.A.W., Nichols, J.D., Gude, J.A., Rich, L.N., Podruzny, K.M., Hines, J.E. & Mitchell, M.S. (2013) Determining occurrence dynamics when false positives occur: estimating the range dynamics of wolves from public survey data. *PLoS One*, **8**, e65808.

Peck, S.L. (2004) Simulation as experiment: a philosophical reassessment for biological modeling. *Trends in Ecology and Evolution*, **19**, 530–534.

R Development Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org (accessed 26 May 2013).

Royle, J.A. & Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.

Shea, C.P., Peterson, J.T., Wisniewski, J.M. & Johnson, N.A. (2011) Misidentification of freshwater mussel species (Bivalvia: Unionidae): contributing factors, management implications, and potential solutions. *Journal of the North American Benthological Society*, **30**, 446–458.

Simons, T.R., Alldredge, M.W., Pollock, K.H. & Wettroth, J.M. (2007) Experimental analysis of the auditory detection process on avian point counts. *The Auk*, **124**, 986–999.

Sturtz, S., Ligges, U. & Gelman, A. (2005) R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, **12**, 1–16.

Received 17 April 2015; accepted 2 July 2015
Handling Editor: Nigel Yoccoz

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. details about methods.

Table S1. Performance of the confirmed absences and confirmed presences (CACP) model in scenarios with no heterogeneity.

Table S2. Performance of the confirmed presences (CP) model in scenarios with no heterogeneity, where priors for the false positive probability (P_{10}) were vague U = U(0, 0·5) or informative I = Beta(1, 9)) and for the observation confirmation probability (b) were vague UU = U(0, 1), UB = Beta(0·5, 0·5), or informative IU = U(0·01, 0·05), IB = Beta(10, 300).

Table S3. Performance of the confirmed absences and confirmed presences (CACP) model when data had observation confirmation errors and no heterogeneity.

Table S4. Performance of the confirmed presences (CP) model when data had observation confirmation errors and no heterogeneity.

Table S5. Performance of models when there was a 'phantom' species and no heterogeneity.

Table S6. Performance of the model that assumes false positive errors do not occur when applied to simulated data that contained false positive errors.

Table S7. Performance of models in scenarios with heterogeneity: a strong quadratic function related occupancy probabilities to a covariate.

Table S8. Performance of models in scenarios with heterogeneity: a weak quadratic function related occupancy probabilities to a covariate.

Table S9. Performance of models in scenarios with heterogeneity: a linear function related occupancy probabilities to a covariate.

Fig. S1. Simulated effect of percent forest on occupancy probabilities and effect of temperature on true positive detection probabilities under three scenarios: (a) strong quadratic, (b) weak quadratic, and (c) linear.

Fig. S2. Simulated (black) and estimated (grey) effects of percent forest on occupancy probabilities from the strong quadratic scenario.

Fig. S3. Simulated (black) and estimated (grey) effects of percent forest on occupancy probabilities from the weak quadratic scenario.

Fig. S4. Simulated (black) and estimated (grey) effects of percent forest on occupancy probabilities from the linear scenario.

Fig. S5. Bias of posterior occupancy probabilities as a function of covariate values from the strong quadratic scenario.

Fig. S6. Bias of posterior occupancy probabilities as a function of covariate values from the weak quadratic scenario.

Fig. S7. Bias of posterior occupancy probabilities as a function of covariate values from the linear scenario.

Fig. S8. Bias in posterior occupancy (psi), true positive detection (P_{11}), false positive detection (P_{10}), and observation confirmation probabilities (b) when the presence of false positives in data and in a model did not match.

Fig. S9. Bias in posterior occupancy (psi), true positive detection (P_{11}), false positive detection (P_{10}), and observation confirmation probabilities (b) with different sample sizes: three surveys at either 250 sites or 100 sites.

Appendix S2. Occupancy model code.