Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 900–909 doi: 10.1111/2041-210X.12542

Uncertainty in biological monitoring: a framework for data collection and analysis to account for multiple sources of sampling bias

Viviana Ruiz-Gutierrez[1]*, Mevin B. Hooten[1,2,3] and Evan H. Campbell Grant[4]

[1]*Department of Fish, Wildlife, and Conservation Biology, 109 Wagar Building, Colorado State University, Fort Collins, CO 80523, USA;* [2]*Colorado Cooperative Fish and Wildlife Research Unit, 201 Wagar Building, U.S. Geological Survey, Fort Collins, CO 80523, USA;* [3]*Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA; and* [4]*Patuxent Wildlife Research Center, S.O. Conte Anadromous Fish Laboratory, One Migratory Way, U.S. Geological Survey, Turners Falls, MA 01376, USA*

Summary

1. Biological monitoring programmes are increasingly relying upon large volumes of citizen-science data to improve the scope and spatial coverage of information, challenging the scientific community to develop design and model-based approaches to improve inference.

2. Recent statistical models in ecology have been developed to accommodate false-negative errors, although current work points to false-positive errors as equally important sources of bias. This is of particular concern for the success of any monitoring programme given that rates as small as 3% could lead to the overestimation of the occurrence of rare events by as much as 50%, and even small false-positive rates can severely bias estimates of occurrence dynamics.

3. We present an integrated, computationally efficient Bayesian hierarchical model to correct for false-positive and false-negative errors in detection/non-detection data. Our model combines independent, auxiliary data sources with field observations to improve the estimation of false-positive rates, when a subset of field observations cannot be validated *a posteriori* or assumed as perfect. We evaluated the performance of the model across a range of occurrence rates, false-positive and false-negative errors, and quantity of auxiliary data.

4. The model performed well under all simulated scenarios, and we were able to identify critical auxiliary data characteristics which resulted in improved inference. We applied our false-positive model to a large-scale, citizen-science monitoring programme for anurans in the north-eastern United States, using auxiliary data from an experiment designed to estimate false-positive error rates. Not correcting for false-positive rates resulted in biased estimates of occupancy in 4 of the 10 anuran species we analysed, leading to an overestimation of the average number of occupied survey routes by as much as 70%.

5. The framework we present for data collection and analysis is able to efficiently provide reliable inference for occurrence patterns using data from a citizen-science monitoring programme. However, our approach is applicable to data generated by any type of research and monitoring programme, independent of skill level or scale, when effort is placed on obtaining auxiliary information on false-positive rates.

Key-words: bias, detection, false-negative, false-positive, monitoring, occupancy, probit link

Introduction

Biological monitoring programmes are fundamental to much of our understanding about the natural world and play an integral role in biodiversity conservation by providing valuable knowledge about the state of ecosystems and populations (Jones 2011). To increase the spatial coverage of information, many monitoring programmes are solely based on, or have incorporated, data collected by the general public (i.e. citizen scientists) (Dickinson *et al.* 2012). For example, the North American Breeding Bird Survey (BBS), an avian monitoring programme made up of 2800 volunteer survey routes, generates the largest source of information used to determine the status and trends of bird populations for Canada and the United States (Sauer, Fallon & Johnson 2003; Sauer & Link 2011). These and other citizen-science programmes have improved our understanding of the presence, spread and response to management of invasive species (Hooten *et al.* 2007; Gallo & Waitt 2011) and emerging diseases (Hosseini, Dhondt & Dobson 2004).

Global trends of ecosystem degradation and loss of biodiversity (Hooper *et al.* 2012), coupled with uncertainty related to changing environmental conditions (Bellard *et al.* 2012) and

*Correspondence author. E-mail: vr45@cornell.edu

a need to make observations over large spatial spatial extents with limited resources, are challenging the scientific community to develop efficient ways of using large volumes of citizen-science data to address these and other emerging issues. As with any scientific study, accuracy may be improved by addressing aspects of sampling design (e.g. selection of study sites, training observers) and those which are model based (e.g. accounting for heterogeneity in detection). Much effort has been placed on design-based suggestions for improving biological monitoring programmes in general (Nichols & Williams 2006), which includes a wealth of sampling design and data analysis techniques to account for key sources of heterogeneity in our observations (Williams, Nichols & Conroy 2002; Mackenzie & Royle 2005; Mackenzie *et al*. 2006). However, even the best study design may require model-based approaches to improve accuracy, as not all heterogeneity can be accommodated through improved sampling designs (Pacifici *et al*. 2015). Sources of heterogeneity are common in all data based on the observation of organisms, even when collected by professional scientists or well-trained technicians (McClintock *et al*. 2010b). However, the skill level of citizen-science volunteers can be more variable and harder to control, increasing the likelihood of errors such as false-positive detections (Farmer, Leonard & Horn 2012; Hochachka *et al*. 2012). As volunteer-based programmes increase in popularity, the development of model-based approaches to improve accuracy of large volumes of citizen-science data is key to increase the level of trust, scope and applicability of the resulting inference (Dickinson *et al*. 2012; Tulloch *et al*. 2013).

The value of any type of monitoring data depends on how these data can be used to infer true occurrence or absence of an event of interest (Farmer, Leonard & Horn 2012). Therefore, model-based approaches aimed at improving accuracy of monitoring data should focus on ways to account for known sources of bias within this context, such as the probability of missing an event that occurred, and the probability of reporting an event that did not occur. We illustrate the need to correct for these defined sources of error using Bayes' theorem, which is a natural way to characterize the probability structure of the occurrence or absence of an event (Efron 2013; Hobbs & Hooten 2015). For example, if interest is in estimating the probability of an event (e.g. the presence of a disease), when interpreting results from testing a sample for a specific disease, the probability that the sampled individual has the disease, given that the individual tested positive for the disease, can be described as

$$p(A|B) = \frac{p(B|A)p(A)}{p(B|A)p(A) + p(B|A^c)p(A^c)} \quad \text{eqn 1}$$

where $p(B|A)$ is the probability of a positive detection of the disease (B) given that the individual has the disease (A), $p(A)$ is the prevalence of the disease in the sampled population, $p(B|A^c)$ is the probability of testing positive for the disease when the individual does not have the disease, and $p(A^c)$ is the probability of being healthy (Fig. 1).

While accounting for both false-positive [$p(B|A^c)$] and false-negative [$p(B|A)$] errors is common in epidemiology, most work on the application of Bayes' theorem in ecology has focused on the detection of an event given that has occurred, $p(B|A)$ (MacKenzie *et al*. 2002; MacKenzie 2006), and there is an extensive body of work that examines how ignoring false-negative errors influence our interpretation of key processes in population and community dynamics (Risk, de Valpine & Beissinger 2011; Ruiz-Gutierrez & Zipkin 2011). As a result, data collection and analysis techniques have been developed that allow for the direct estimation of $p(B|A)$, including other sources of heterogeneity using occupancy models (Mackenzie *et al*. 2006). The main modification is the addition of repeated observations within sampling periods, during which a site is assumed to be closed to changes in the true occurrence of an event. This repeated sampling design is suggested as the best approach for monitoring the state of ecosystems and biodiversity (Elphick 2008; Jones 2011), and despite their increasing popularity in applied ecology, few citizen-science monitoring programmes result in data appropriate for analysis using an occupancy framework (e.g. Kery *et al*. 2010; van Strien *et al*. 2010).

Less attention has been focused on ways to account for the probability of mistakenly detecting an event that has not occurred, $p(B|A^c)$, or false-positive probability, aside from a few notable exceptions (e.g. Royle & Link 2006; McClintock *et al*. 2010a; Miller *et al*. 2012). So far, this work indicates that false-positive errors are also important sources of bias, where rates as low as $p(B|A^c) = 0{\cdot}01$ can cause the overestimation of the probability of occurrence of an event by as much as 30% (Royle & Link 2006). This is of considerable concern given the mounting evidence for the presence of false-positive errors across diverse systems, where rates as high as $p(B|A^c) = 0{\cdot}11$ have been estimated in avian (Simons *et al*. 2007) and anuran studies (McClintock *et al*. 2010a). False-positive errors are likely to occur in situations where it is difficult to distinguish between individual observed events, such as species with similar calls (e.g. grey treefrogs in the Eastern US), or morphology (e.g. *Contpus* sp. flycatchers). Given the increasing trend to monitor biological diversity with the help of citizen scientists, the ability to reduce the occurrence of, and correct for, both false-positive and false-negative errors are critical for future large-scale monitoring programmes, in particular when the objective is to detect the occurrence of uncommon or isolated events (e.g. introduced species on the invasion front) or to monitor changes in a system. False-positive errors have also been shown to bias our understanding of fundamental ecological processes such as species turnover, colonization and extinction (McClintock *et al*. 2010a,b; Miller *et al*. 2013). The main challenge with correcting for false-positive errors is that, unlike the false-negative probability, design-based modifications alone do not provide enough information to estimate $p(B|A^c)$ without the need for additional field sampling information (McClintock *et al*. 2010b; Farmer, Leonard & Horn 2012).

Examples of approaches to deal with false-positive errors include limiting the inclusion of a subset of unreliable data (e.g. Molinari-Jobin *et al*. 2012) or eliminating them altogether (e.g. Hochachka *et al*. 2012) from analyses. More formal approaches range from categorizing observations based on

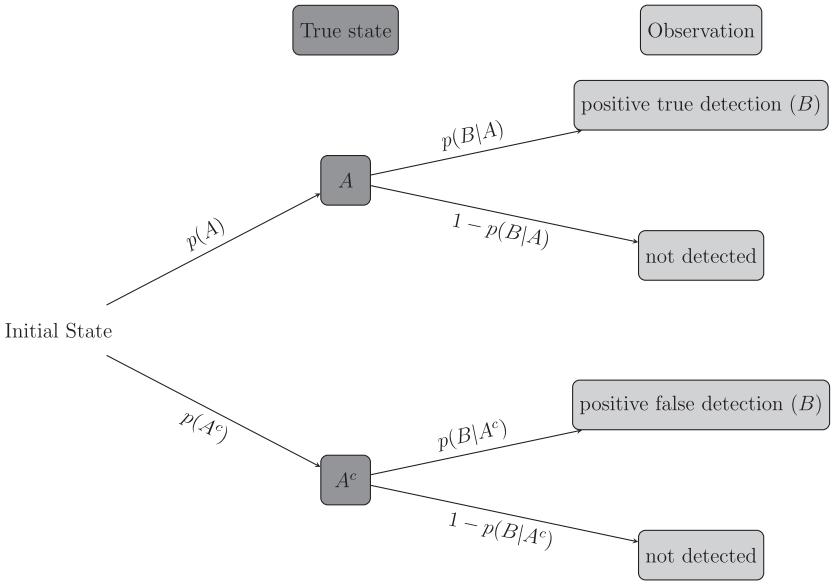


Fig. 1. The probability structure of Bayes' theorem for starting out at an initial state and moving into state A with probability $p(A)$. The true presence of an event (e.g. true state: A) can be positively observed (B) with probability $p(B|A)$ or not observed with probability $1 - p(B|A)$. From the initial state, the probability of not being present, or being in state A^c is $p(A^c)$, and the probability of a false-positive observation (B) of state A is $p(B|A^c)$, and a negative observation can occur with the probability $1 - p(B|A^c)$.

their degree of reliability, to the development of statistical models that formally deal with uncertainty in observations (see detailed summaries in Miller *et al*. 2015; Chambert, Miller & Nichols 2015). Chambert, Miller & Nichols (2015) categorized current false-positive statistical models into three categories, largely based on how different types of information are modelled to inform false-positive rates. The *Site Confirmation Model* relies on being able to assume, or having additional direct information on, the true state of occurrence of a site for a subset of observations (e.g. Clement *et al*. 2014). Several examples exist of this popular approach, which has been found to perform well for cases when the probability of occurrence is high (Miller *et al*. 2011, 2013). The *Calibration Model* is described as incorporating independent sources of data to inform both false-positive and false-negative probabilities, in addition to field observation data. Lastly, the *Observation Model Confirmation* expands on previous efforts to correct inference *a posteriori* (e.g. Gardiner *et al*. 2012) and uses different types of additional information collected during the sampling event to categorize field observations as belonging to one of four categories: true absences, true presences only, false positives only and a combination of false absences and presences (Chambert, Miller & Nichols 2015).

The *Site Confirmation Model* shares similarities with the *Observation Model Confirmation* approach in that a subset of field observations must be classified as true or false detections, either by knowing or assuming the true state in the former, or by confirming the true state based on field observations *a posteriori* in the latter. The utility of these approaches to biological monitoring programmes is constrained by the feasibility of obtaining field data that can be assumed or classified as perfect (i.e. free from error). Therefore, the *Calibration Model* is the most promising approach for long-term or large-scale monitoring programmes, for it relaxes the need for additional field-based information, and could potentially make use of a wide range of auxiliary sources of information on false-positive errors (Chambert, Miller & Nichols 2015).

Although data to inform false-positive rates may not be reliably collected by observers conducting standard surveys, they may be informed by independent data collected outside of the normal survey design. Such independent sources of information on false-positive errors can be collected using experimental trials to estimate the probability of detecting an event that has not occurred [i.e. $p(B|A^c)$]. For example, one can play a number of bird calls under field or laboratory conditions for a set of observers, and record how often a species was listed as present when their call was not actually played. Such information can be collected at small-scales, such as testing the ability of a few field technicians working on a specific research project. For larger monitoring programmes, an online platform could be used to test thousands of volunteer citizen-science participants using both visual and auditory test trials (e.g. Frog Quiz: https://www.pwrc.usgs.gov/frogquiz/). Each trial is assumed to be independent of other trials, between and among individuals, and information about each participant (e.g. skill level, years of experience) could be further associated as covariates of false-positive probabilities.

Here, we expand upon the proposed framework for the *Calibration Model* with the objective of improving accuracy of biological monitoring data, including that of citizen-science-based programmes. We developed a Bayesian hierarchical model, which incorporates an independent, auxiliary source of information about false-positive errors to improve inferences on occurrence patterns. Our model differs from the one proposed by Chambert, Miller & Nichols (2015) in the fact that we include auxiliary information only for false-positive errors, which are more difficult to estimate from within-sample data, and use information collected during repeat sampling events to inform false-negative rates. In addition, we used a simulation study to explore the performance of our model across a range of volumes of independent auxiliary information on false-positive rates under different scenarios of occupancy, false-negative and false-positive errors. We further provide the first application of a *Calibration Model* using data from a large-scale, citi-

zen-science monitoring programme and contrasted these inferences with a standard occupancy model to demonstrate the improvement in inference (i.e. reduced bias) by incorporating within- and out-of-sample data to jointly inform false detection rates. Lastly, we discuss potential extensions to the model, such as including covariates on effort and observer skill level on false-positive probabilities.

All full-conditional distributions in our model can be derived in closed form, allowing for the use of a computationally efficient Gibbs sampling algorithm (Hooten, Larsen & Wikle 2003; Dorazio & Rodriguez 2012) to sample all model parameters, reducing the tuning needed to reach model convergence. This is especially important to address increasing needs of large-scale biological monitoring programmes, including those based on citizen scientists. The false-positive model we present provides a flexible framework to correct the presence–absence data for observation errors in the absence of additional information collected during sampling events and may be applied to both professional- and citizen-scientist-based monitoring programmes.

Methods

FALSE-POSITIVE MODEL

Existing Bayesian hierarchical occupancy models provide the basic framework to model the probability of occurrence of events (McClintock *et al.* 2010c). We extend this class of models by incorporating auxiliary sources of information to allow for inference on occupancy probability [$p(A|B)$], while accounting for known sources of bias (Fig. 1). In our model, we assume that observation y_{ij} is a binomial random variable that indicates if a species was detected ($y_{ij} = 1$) or not detected ($y_{ij} = 0$) at sampling location i, during the jth repeated sampling event within a season, for $i = 1, \ldots, n$ and $j = 1, \ldots, J_i$. Similar to standard occupancy models, we use a mixture model specification to describe the data generating mechanism in terms of probability distributions as

$$y_{ij} \sim \begin{cases} \text{Bern}(p_0) & \text{if } z_i = 0 \\ 0 & \text{if } u_{ij} \leq 0, z_i = 1 \\ 1 & \text{if } u_{ij} > 0, z_i = 1 \end{cases} \quad \text{eqn 2}$$

where z_i is an indicator of true occurrence of a species at sampling location i and u_{ij} is a latent normal random variable. We assume that $u_{ij} \leq 0$ when $y_{ij} = 0$ and $z_i = 1$, and $u_{ij} > 0$ when $y_{ij} = 1$ and $z_i = 1$. We note that this model specification differs from those of more typical occupancy models (e.g. Royle & Dorazio 2006) in that it uses a latent auxiliary variable (i.e. u_{ij}) to account for the false-positive or true detection process, following previous developments described by Dorazio & Rodriguez (2012) and Johnson *et al.* (2013). This modification to the more conventional model structure maintains the original concept of the detection process, but it can be computationally advantageous to implement. This becomes increasingly important as the amount of data increases – as is the case with large-scale citizen-science monitoring programmes.

Our model further differs in that when an event has not occurred (i.e. $z_i = 0$), we assume that the probability of observing the event (i.e. $y_{ij} = 1|z_i = 0$) is p_0, or the false-positive probability. The current challenge is that information needed to estimate p_0 cannot be derived from the observational data (i.e. y_{ij}) without additional assumptions (Royle & Link 2006). We address this issue by incorporating an auxiliary data source to improve estimation of p_0, therefore allowing us to differentiate between the true occurrence of an event ($z_i = 1$), and a false-positive detection of the same ($y_{ij} = 1|z_i = 0$). The auxiliary data are derived from a series of tests or experiments where we know that an event has not occurred, but it is reported as detected or determined as having occurred during the experiment. For example, suppose that n_c is the total number of audio trials where observers are asked to identify calls of individual species from a predefined candidate set of species. Under this scenario, c_{01} is the total number of times the call of a specific species was indicated as detected during an audio trial, when the call for that species was not played. We model this auxiliary source of information as

$$c_{01} \sim \text{Binom}(n_c, p_0) \quad \text{eqn 3}$$

where c_{01} is the total count of positive detections of an event out of n_c independent test trials where the event is known to not have occurred. We assume that c_{01} arises from a binomial distribution with n_c independent trials and probability p_0. This is a similar approach to defining a strong prior on p_0, but it differs in that an increase in the amount of independent trials (i.e. n_c) will directly improve the accuracy of p_0; thus, sample effort can be investigated to optimize the power of a monitoring programme to provide unbiased estimates of occupancy rates.

To model false-negative probabilities, we allow detection to be a function of covariates measured during the J_i sampling events, at all n sampling locations, by modelling u_{ij} as

$$u_{ij} \sim \text{Normal}(\mathbf{w}'_{\mathbf{ij}}\boldsymbol{\alpha}, \mathbf{1}) \quad \text{eqn 4}$$

where $\mathbf{w}_{ij}$ is a vector of q covariates for the ith sampling location and the jth sampling visit, and $\boldsymbol{\alpha}$ are coefficients associated with the covariates for detection probability [$p(B|A)$]. Our specification implicitly assumes a probit link $\Phi^{-1}(p_{ij}) = \mathbf{w}'_{ij}\boldsymbol{\alpha}$, where p_{ij} is the probability of detecting an event ($y_{ij} = 1$) given that it has occurred ($z_i = 1$), and Φ indicates the standard normal cumulative distribution function. The true occurrence of an event at sampling location i, denoted z_i, was defined as a latent variable (i.e. the true state is imperfectly observed) and we model z_i as

$$\begin{aligned} z_i &\sim \text{Bern}(\psi) \\ \psi &\sim \text{Beta}(\alpha_\psi, \beta_\psi) \end{aligned} \quad \text{eqn 5}$$

where ψ is the probability of a specific event occurring at i, after accounting for both false-positive (i.e. $p(B|A^c)$ or p_0) and false-negative (i.e. $p(B|A)$ or p_{ij}) probabilities (Table 1).

EXAMPLE 1: SIMULATION STUDY

To assess the performance of the false-positive model, we simulated the effect of different values of p_0 under varying scenarios of occurrence, detection, and quantity of auxiliary information used to estimate p_0. Our values of p_0 were obtained from the literature, where they have been found to range from $p_0 = 0{\cdot}01$ to $p_0 = 0{\cdot}11$ for avian (Simons

Table 1. A list of each probability defined in Bayes theorem, along with their individual definitions, equivalent parameters in the false-positive model, and interpretation

Probability	Definition	Parameter	Interpretation
$p(A)$	Prevalence	ψ	Probability of occurrence of an event
$1 - p(B\|A^c)$	Specificity	p_0	False-positive probability
$p(B\|A)$	Sensitivity	p	False-negative probability

et al. 2007) and anuran species surveys (McClintock *et al.* 2010b). For simulations, we chose four values of p_0: 0·01, 0·05, 0·10 and 0·20. For each of these values, we explored four scenarios for the true probability of occurrence of an event (very low = 0·2, low = 0·4, medium = 0·6, high = 0·8). We modelled initial detection probability (p_{ij}) under two scenarios (low ~0·4, high ~0·6) and included linear and squared effects of sampling day to reflect changes in p_{ij} throughout the season. We also predicted that precision and bias would both improve with an increase in the total number of independent trials (n_c) used to estimate p_0 and that scenarios with low probabilities of occurrence and detection, and high probabilities of false positives, would require the most amount of effort for the same gain in precision and reduction in bias. To examine the effect of increasing the size of an auxiliary data set, we simulated the following numbers of total independent trials n_c: 500, 1000, 1500, 2000, 3000, 5000, 10 000, and 15 000. All simulations and analyses were carried out using the R Statistical Computing Environment (R Core Team, 2013). We carried out 100 simulations for each case scenario and generated 15 000 MCMC samples using a burn-in period of 5000 iterations for each simulation. A more detailed description of the model implementation can be found in Appendix S1 (Supporting information).

Simulation results

The false-positive model performed without bias under simulated scenarios of probabilities of occurrence, false-positive and false-negative errors and effort in auxiliary data. Bias in posterior mean estimates never exceeded more than 2% across all values of p_0. As predicted, an increase in effort, or total number of independent trials (n_c), resulted in an increase in precision of posterior means of probability of occurrence at around 5000 trials (Fig. 2). When false-positive probabilities are around $p_0 = 0{\cdot}1$, the degree of precision does not improve after 5000 trials and in fact decreases with increasing probability of occurrence (Fig. 2). Under conditions with high false-positive probabilities ($p_0 = 0{\cdot}2$), we observed little improvement in precision with increasing effort, independent of probabilities of occurrence (ψ), although credible intervals were smaller at lower probabilities of detection ($p = 0{\cdot}04$) (Fig. 2).

EXAMPLE 2: ANURAN MONITORING PROGRAMME

To examine occurrence patterns of anurans in the north-eastern United States, we used North American Amphibian Monitoring Program (NAAMP) records for 152 surveyed routes, for the following 10 species: American toad (*Anaxyrus americanus*), Fowler's toad (*Anaxyrus fowleri*), northern cricket frog (*Acris crepitans*), grey treefrog complex (*Hyla versicolor*/*chrysocelis*), green treefrog (*Hyla cinerea*), spring peeper (*Pseudacris crucifer*), American bullfrog (*Lithobates catesbeianus*), green frog (*Lithobates clamitans*), pickerel frog (*Lithobates palustris*), and wood frog (*Lithobates sylvaticus*). We used independent information on false-positive rates from replicated field experiments using a remote broadcasting system to simulate simple anuran call surveys for details see McClintock *et al.* 2010a). To capture the variability in detection for the duration of the survey period across all surveyed states (February–July), we modelled detection probabilities for each species as $\Phi^{-1}(p_{ij}) = \mathbf{w}_{ij}'\boldsymbol{\alpha}$, where $\mathbf{w}_{ij}$ is a vector of the linear and quadratic terms representing sampling day (e.g. February 24 was day 1) for route i during the jth visit.

To evaluate the effects of false-positive errors on estimates of anuran occurrence rates, we contrasted results obtained using our false-positive model with those obtained using a standard model that only corrects for false-negative errors (Fig. 3). For the standard occupancy model, we modified the mixture specification of the false-positive model to follow the commonly used notation for occupancy models (Dorazio & Rodriguez 2012), such that

$$y_{ij} = \begin{cases} 0 & \text{if } z_i = 0 \\ 0 & \text{if } u_{ij} \leq 0, z_i = 1 \\ 1 & \text{if } u_{ij} > 0, z_i = 1 \end{cases} \quad \text{eqn 6}$$

where $y_{ij} = 0$ when $z_i = 0$. We fit both types of models to each species' data independently using the R statistical computing environment (R Core Team, 2013). For each model fit, we generated 30 000 MCMC samples and used a burn-in period of 10 000 iterations.

Anuran case study results

We obtained inference for occupancy probabilities for 10 anuran species in the north-eastern United States using both the false-positive and standard occupancy models. Two very commonly detected species, spring peeper and green frog, had posterior means for $\psi \approx 1{\cdot}0$; therefore, false-positive rates did not bias occupancy estimates for these species. For the remaining eight species, we found varying degrees of influence of the auxiliary data on estimates of occurrence. We did not find a pronounced effect of false-positive errors on occupancy for American toad, grey treefrog complex and American bullfrog. The former two species had posterior means for p_0 that were approximately 0·2, and all three species had high values for posterior means of $\psi > 0{\cdot}68$ (Table 2). Effects of false-positive errors were found in the four species with posterior means with $\psi < 0{\cdot}3$, with the strongest effect in pickerel frog (Table 2). A relatively high posterior mean for p_0 equal to 0·19, combined with low posterior means for occupancy and detection ($\psi = 0{\cdot}26$ and $p = 0{\cdot}1$), resulted in a difference of 0·1 in posterior means for occupancy between the false-positive model and the standard model (Table 2). To gain a better understanding of the bias associated with inference on the number of occupied routes, we calculated the number of occupied routes, or (N), as a derived quantity ($N = \sum_{i=1}^{n} z_i$). Results were more pronounced in the four species with posterior means of $\psi < 0{\cdot}3$, where the standard model overestimated the number of occupied routes by 20–70%. The species with the largest discrepancy was northern cricket frog, which has a posterior mean of $N = 15$ occupied routes with the false-positive model, in contrast to $N = 25$ with the standard model (Table 2).

Discussion

To understand the potential impacts of habitat and environmental changes, we need precise and unbiased estimates of population trends at large spatial and temporal scales. Citizen-science monitoring programmes have the potential to fulfil this role at scales not achievable by most conventional research, but correcting for known sources of bias inherent to all types of observational data is critical for reliable inference. The model we present was successful at correcting for both false-positive and false-negative errors by coupling auxiliary information of the frequencies of false-positive observations with detection-non-detection monitoring data to estimate occurrence patterns. This approach eliminates the need to assume that false-positive errors are non-existent for certain types of observations, invest effort in verifying observations or define arbitrary criteria to determine what constitutes a reliable obser-

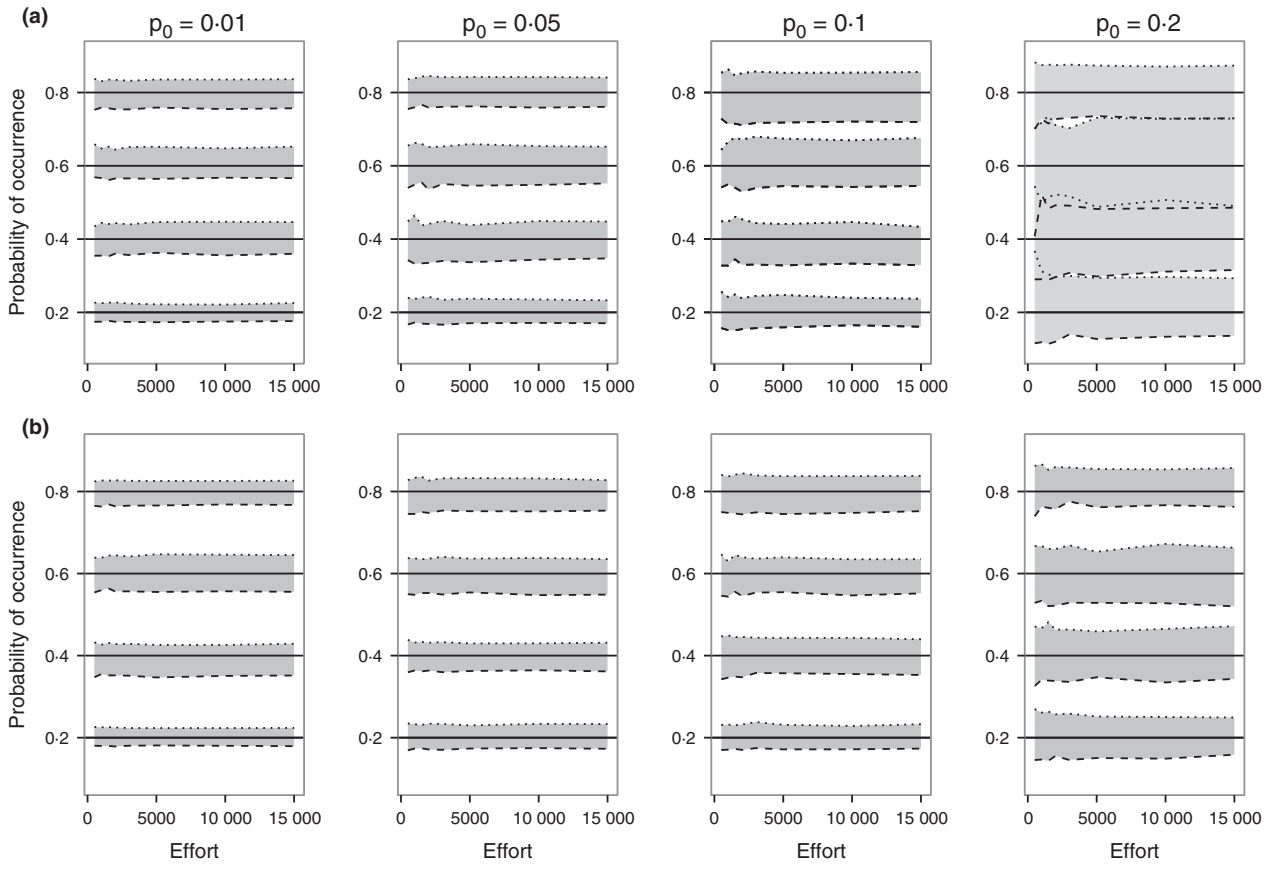
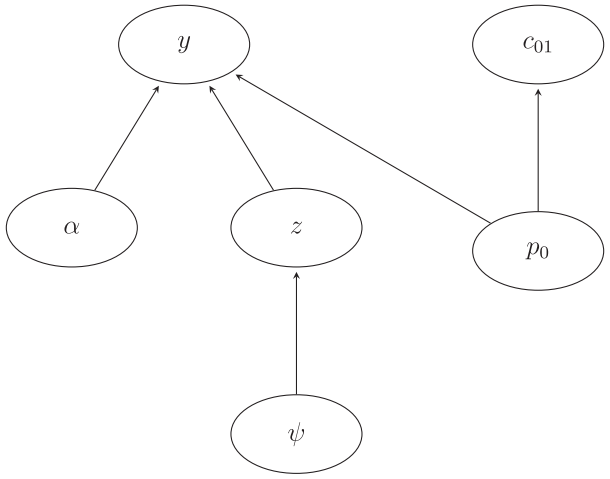


Fig. 2. Simulated probabilities of occurrence and credible intervals for posterior means using the false-positive model. Data were simulated under low (a: p = 0·4) and high (b: p = 0·6) detection probabilities, across a range of values for probability of occurrence (solid black line), false-negative probability (p_0) and total number of trials (Effort).

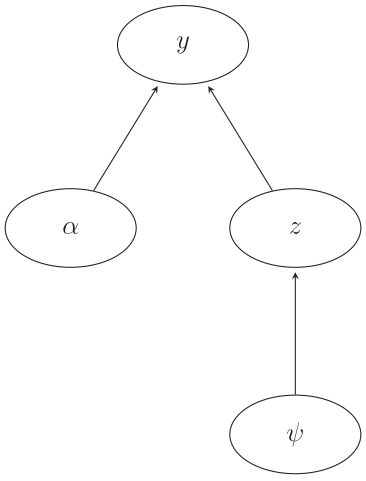
vation *a posteriori*, all which are not likely to completely eliminate false-positive errors (Miller *et al.* 2015). In addition, our model performed well across a wide range of occupancy values (ψ = 0·2–0·8) and false-positive rates (ψ = 0·2–0·8), although careful consideration of survey design should be given to species with high false-positive rates, given the low levels of precision that we obtained even with high volumes of auxiliary information.

The application of our false-positive model requires auxiliary sources of data, and attention should be paid by biological monitoring programmes on how to best collect this information. The number of independent trials needed to increase accuracy by correcting false-positive errors can be determined prior to initiating a survey programme based on the desired precision and best available estimates of probability of occurrence, false-negative rates, and false-positive rates. The range of false-positive rates estimated for most frog and bird species (p_0 = 0·01–0·05) suggest that 5000 independent trials should be sufficient to improve precision. When false-positive rates are relatively high (p_0 = 0·1–0·2), resources should be invested into increasing the number of trials beyond 15 000, independent of probabilities of detection and occurrence. In addition, consideration should be given to the test trials themselves, and effort should be made to best represent conditions experienced by the observers. For example, provide both audio and visual trials for bird monitoring projects, or provide realistic combinations of multiple calls of species commonly found together, whenever possible.

The cost associated with high numbers of trials can be reduced by sharing information across multiple monitoring programmes, or using online platforms or mobile phone technology to collect information, instead of the field experiments as were used here. For example, the North American BBS has an online quiz (http://www.mbr-pwrc.usgs.gov/bbs/trend/birdquiz.html) using both visual and audio test trials. Such platforms can potentially generate large volumes of trials to inform false-positive rates. These independent tests can be carried out after field data collection. For example, camera trap photographs presented to observers to test their ability to identify individual animals have been used to estimate false-positive error rates in data collected from opportunistic observations by community-based monitoring programmes (Meek *et al.* 2014). Specimens housed at museums, and other collections such as sound and video archives, could also be used to conduct in-person independent test trials. Ultimately, reducing uncertainty in citizen-science data demands a shift in our current data collection paradigm to define, identify and prioritize the collection of external sources of information on false-positive rates as part of an overall sampling programme. This shift will likely increase the return on investment of



(a) false positive occupancy model



(b) Occupancy model

Fig. 3. Comparison of the false-positive occupancy model (a) and the standard occupancy model (b). The figure illustrates the relationship between the data model (y), the covariates on detection probability (α), the true state of occurrence of an event (z) and the probability of occurrence of an event (ψ). The uniqueness of the false-positive model is an additional data source (c_{01}) to inform the false-positive probability (p_0).

any biological monitoring programme, along with well-defined objectives and protocols (Tulloch *et al.* 2013).

In 2008 and 2010, the Amphibian Research and Monitoring Initiative of the U.S. Geological Survey (ARMI) invested in a series of replicated experiments to estimate and explore the consequences of false-positive errors on anuran calling surveys. This effort has provided much of the knowledge on the effects of false-positive errors on occupancy and related dynamics (McClintock *et al.* 2010a; Miller *et al.* 2015) and provided a framework to examine these effects using calling anuran surveys, although we note that the general guidelines and model we present are applicable to other surveys where auxiliary information can be collected (e.g. data collected by Meek *et al.* 2014). Our analyses of these data indicate that six of the 10 species appear to be widespread in the routes surveyed by NAAMP in 2004, with posterior means for occupancy probabilities ranging from 0·68 to 0·99. Other species with low posterior means for occupancy probabilities such as the pickerel frog ($\psi = 0{\cdot}26$) are not of conservation concern, but the bias in the expected number of occupied routes does have some important implications for the estimated range of this species. For example, the core of the distribution of *H. cinerea* is the south-eastern United States, and it is not a surprise that this common species was estimated to be present at only 9 of the 152 surveyed routes in the north-eastern states. However, the standard occupancy model overestimated the number of occupied routes by 50%, and these errors could potentially mask range expansions into northern states driven by changes in climatic conditions, for example. Similarly, the range for *A. crepitans* is restricted to the south-east of the Appalachian Mountains, and the posterior mean for occupancy was similarly estimated to be low ($\psi = 0{\cdot}11$) in our study. The species is estimated to have suffered extreme range contractions, and there is interest in identifying features of habitats that allow successful persistence of this species (Beasley *et al.* 2005) and could be used to develop conservation and management strategies. Therefore, the use of our false-positive model is

Table 2. Table of model results. For each species, scientific and, common names and are provided, as well as posterior means for false-positive probability (p_0), detection probability (p_{fp}), occupancy (ψ_{fp}) and mean number of occupied sites (N_{fp}) using the false-positive model, and estimates of detection probability (p_s), occupancy (ψ_s) and number of occupied sites (N_s) using the standard occupancy model, including the percentage of bias in the estimated number of sites between the standard and the false-positive models (% ΔN)

		False-positive model				Standard model			
Scientific name	Common name	p_0	p_{fp}	ψ_{fp}	N_{fp}	p_s	ψ_s	N_s	Bias % ΔN
Anaxyrus americanus	American toad	0·025	0·2	0·68	104	0·2	0·7	108	4
Anaxyrus fowleri	Fowler's toad	0·01	0·09	0·29	44	0·09	0·35	53	20
Acris crepitans	Northern cricket frog	0·02	0·05	0·105	15	0·05	0·17	25	67
Hyla versicolor	Grey treefrog	0·018	0·25	0·72	109	0·25	0·74	112	3
Pseudocaris crucifer	Spring peeper	0·023	0·6	0·99	152	0·6	0·99	152	0
Hyla cinerea	Green treefrog	0·11	0·04	0·06	9	0·04	0·1	14	56
Lithobates catesbeianus	American bullfrog	0·01	0·43	0·76	116	0·43	0·77	118	2
Lithobates clamitans	Green frog	0·01	0·54	0·98	150	0·54	0·98	150	0
Lithobates palustris	Pickerel frog	0·19	0·1	0·26	39	0·1	0·36	55	41
Lithobates sylvaticus	Wood frog	0·023	0·23	0·73	112	0·23	0·79	120	7

advised, considering that the standard model overestimated the number of occupied routes by 67%. The distribution for *A. fowleri* is not as restricted as the previous two species, but it has a scattered distribution pattern and a low posterior mean for occupancy rates ($\psi = 0{\cdot}29$), likely driven by specific habitat requirements preferred for breeding (Vogel & Pechmann 2010). The standard model overestimated occupancy by 20%, and this bias could potentially mislead conservation efforts aimed at protecting critical breeding areas.

Our inferences on occurrence of anurans in the north-eastern United States could have been influenced by the relatively low sample size of external information on false-positive rates (n_c = 1960–2822 trials for all 10 species). We used information collected during experimental field tests of volunteers who participate in anuran survey programmes. Therefore, the total number of trials was limited, but the trials were very realistic in terms of condition experienced during sampling surveys. In addition, posterior means for $\psi < 0{\cdot}3$ and overall low posterior means for detection rates were found in four species. Because low occupancy and detection rates are more susceptible to the effects of false positives in the data, these species will require a greater amount of auxiliary data in future analyses. Field trails such as the ones we employed can be labour intensive and costly; therefore, we suggest that information on false-positive rates from the online NAAMP Frog Quiz also be applied in the future.

FUTURE POTENTIAL EXTENSIONS

Our false-positive model can be adapted to incorporate factors that have been found to influence p_0. Covariates such as the number of years a participant has collected data for a specific survey or age are currently used to correct for potential observer errors in large data sets (e.g. Hochachka *et al.* 2012). In our example, a categorical variable of calling intensity of vocalizations recorded in the NAAMP surveys could potentially be included as a covariate in future analyses (Miller-Rushing, Primack & Bonney 2012), where higher intensity of calling would be expected to reduce false-positive rates. To accomplish this, we would denote y_{ij} as

$$y_{ij} = \begin{cases} \text{Bern}(p_{0ij}) & \text{if } z_i = 0 \\ 0 & \text{if } u_{ij} \leq 0, z_i = 1, \\ 1 & \text{if } u_{ij} > 0, z_i = 1 \end{cases} \quad \text{eqn 7}$$

where $\Phi^{-1}(p_{0ij}) = \mathbf{w}'_{0ij}\boldsymbol{\alpha}_0$, and $\mathbf{w}_{0ij}$ is a vector of calling intensity scores for the ith sampling location and the jth sampling visit, and $\boldsymbol{\alpha}_0$ are coefficients associated with the covariates for false-positive probability. We would model this additional information as $c_l \sim \text{Bern}(p_{0l})$.

The present model can also be applied to formally examine disease prevalence rates for cases when information exists on false-positive and/or negative rates. Often, *post hoc* approaches are applied, such as defining thresholds on the number of positive detections to confirm the presence of a pathogen when false-positive rates are considered to be low (Cheng *et al.* 2011). Our flexible approach can directly incorporate raw data from clinical trials [e.g. number of positive tests (c_{01}) and number of control trials (n_c) known to be negative] to inform false-positive probabilities. Published values from diagnostic tests can also be incorporated, such as using estimates of sensitivity as an informative prior for p_{ij} (Table 1).

Other potential extensions include the addition of site-specific covariates for ψ using the same approach we presented to include covariates on detection. Incorporation of covariates may additionally improve bias in the occupancy estimator (Miller *et al.* 2015). Our approach for incorporating auxiliary information can also be applied in models where the interest is correcting for species misidentification errors. Conn *et al.* (2013) contrasted the use of information from multiple observers with defining strong priors to inform the mixture probability of correctly assigning an observation to one of two potential states (e.g. species 1 vs. species 2), to correct for species misidentification errors (see also Hanks, Hooten & Baker 2011). Under this approach, auxiliary information on the probability of correctly assigning a species to an observation could be incorporated directly, in the same fashion we have presented here.

Conclusions

The estimation approach we present is likely to increase the overall accuracy of inferences based on observations collected by biological monitoring programmes, specifically those which rely heavily on citizen-science data. The ability of the latter to aid in the monitoring of the state of a system is key for efficient development and application of conservation and management actions, especially when other sources of irreducible uncertainty are present, such as factors related to climate change (Nichols *et al.* 2011; Williams, Eaton & Breininger 2011). The potential extensions of this model also provide a flexible framework to characterize simple binary event scenarios, like the one we presented, or provide estimates of the occurrence of a series of events and their related dynamics. To take advantage of the full spectrum of potential applications, biological monitoring programmes must include both, the collection of information useful for estimating false-positive error rates, and a system for the estimation of false-negative errors. The improvements in accuracy via our modelling approach should provide an opportunity to expand the scope, scale and coverage of information on plant and animal populations which can be informed by citizen scientists.

Acknowledgements

We would like to thank all North American Amphibian Monitoring Program (NAAMP) volunteers for the collection of the data. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This is contribution 532 of the Amphibian Research and Monitoring Initiative (ARMI) of the US Geological Survey.

Data accessibility

All data used in this manuscript can be found in the North American Amphibian Monitoring Program data download website: https://www.pwrc.usgs.gov/naamp/index.cfm?fuseaction=app.dataDownload.

References

Beasley, V., Faeh, S., Wikoff, B., Staehle, C., Eisold, J., Nichols, D. *et al.* (2005) Risk factors and declines in northern cricket frogs (*Acris crepitans*). *Amphibian Declines* (ed. M. Lannoo), pp. 75–86. University of California Press, Berkeley, CA, USA.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**, 365–377.

Chambert, T., Miller, D.A.W. & Nichols, J.D. (2015) Modeling false positive detections in species occurrence data under different study designs. *Ecology*, **96**, 332–339.

Cheng, T.L., Rovito, S.M., Wake, D.B. & Vredenburg, V.T. (2011) Coincident mass extirpation of neotropical amphibians with the emergence of the infectious fungal pathogen *Batrachochytrium dendrobatidis*. *Proceedings of the National Academy of Science of the United States of America*, **108**, 9502–9507.

Clement, M.J., Rodhouse, T.J., Ormsbee, P.C., Szewczak, J.M. & Nichols, J.D. (2014) Accounting for false-positive acoustic detections of bats using occupancy models. *Journal of Applied Ecology*, **51**, 1460–1467.

Conn, P.B., McClintock, B.T., Cameron, M.F., Johnson, D.S., Moreland, E.E. & Boveng, P.L. (2013) Accommodating species identification errors in transect surveys. *Ecology*, **94**, 2607–2618.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T. & Purcell, K. (2012) The current state of citizen Science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, **10**, 291–297.

Dorazio, R.M. & Rodriguez, D.T. (2012) A Gibbs sampler for Bayesian analysis of site-occupancy data. *Methods in Ecology and Evolution*, **3**, 1093–1098.

Efron, B. (2013) Bayes' theorem in the 21st century. *Science*, **340**, 1177–1178.

Elphick, C.S. (2008) How you count counts: the importance of methods research in applied ecology. *Journal of Applied Ecology*, **45**, 1313–1320.

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

Gallo, T. & Waitt, D. (2011) Creating a successful citizen Science model to detect and report invasive species. *BioScience*, **61**, 459–465.

Gardiner, M.M., Allee, L.L., Brown, P.M.J., Losey, J.E., Roy, H.E. & Smyth, R.R. (2012) Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-Science programs. *Frontiers in Ecology and the Environment*, **10**, 471–476.

Hanks, E.M., Hooten, M.B. & Baker, F.A. (2011) Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological Applications*, **21**, 1173–1188.

Hobbs, N.T. & Hooten, M.B. (2015) *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press, Princeton, NJ.

Hochachka, W.M., Fink, D., Hutchinson, R.A., Sheldon, D., Wong, W.K. & Kelling, S. (2012) Data-intensive Science applied to broad-scale citizen Science. *Trends in Ecology & Evolution*, **27**, 130–137.

Hooper, D.U., Adair, E.C., Cardinale, B.J., Byrnes, J.E.K., Hungate, B.A., Matulich, K.L. *et al.* (2012) A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, **486**, 105–108.

Hooten, M., Larsen, D. & Wikle, C. (2003) Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, **18**, 487–502.

Hooten, M.B., Wikle, C.K., Dorazio, R.M. & Royle, J.A. (2007) Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*, **63**, 558–567.

Hosseini, P., Dhondt, A. & Dobson, A. (2004) Seasonality and wildlife disease: how seasonal birth, aggregation and variation in immunity affect the dynamics of *Mycoplasma gallisepticum* in house finches. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, **271**, 2569–2577.

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

Jones, J.P.G. (2011) Monitoring species abundance and distribution at the landscape scale. *Journal of Applied Ecology*, **48**, 9–13.

Kery, M., Royle, J.A., Schmid, H., Schaub, M., Volet, B., Haefliger, G. & Zbinden, N. (2010) Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, **24**, 1388–1397.

MacKenzie, D. (2006) Modeling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly. *Journal of Wildlife Management*, **70**, 367–374.

Mackenzie, D. & Royle, J. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

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

Mackenzie, D.I., Nichols, J.D., Royle, J., Pollock, K.H., Hines, J.E. & Bailey, L.L. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, San Diego, CA, USA.

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.

McClintock, B.T., Nichols, J.D., Bailey, L.L., MacKenzie, D.I., Kendall, W.L. & Franklin, A.B. (2010c) Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters*, **13**, 659–674.

Meek, P.D., Ballard, G., Claridge, A., Kays, R., Moseby, K., O'Brien, T. *et al.* (2014) Recommended guiding principles for reporting on camera trapping research. *Biodiversity and Conservation*, **23**, 2321–2343.

Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., 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., Grant, E.H.C., 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. doi: 10.1371/journal.pone.0065808

Miller, D.A.W., Bailey, L.L., Grant, E.H.C., McClintock, B.T., Weir, L.A. & Simons, T.R. (2015) Performance of species occurrence estimators when basic assumptions are not met: a test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, **6**, 557–565.

Miller-Rushing, A., Primack, R. & Bonney, R. (2012) The history of public participation in ecological research. *Frontiers in Ecology and the Environment*, **10**, 285–290.

Molinari-Jobin, A., Kery, M., Marboutin, E., Molinari, P., Koren, I., Fuxjaeger, C. *et al.* (2012) Monitoring in the presence of species misidentification: the case of the Eurasian lynx in the Alps. *Animal Conservation*, **15**, 266–273.

Nichols, J.D. & Williams, B.K. (2006) Monitoring for conservation. *Trends in Ecology & Evolution*, **21**, 668–673.

Nichols, J.D., Koneff, M.D., Heglund, P.J., Knutson, M.G., Seamans, M.E., Lyons, J.E. *et al.* (2011) Climate change, uncertainty, and natural resource management. *Journal of Wildlife Management*, **75**, 6–18.

Pacifici, K., Reich, B.J., Dorazio, R.M. & Conroy, M.J. (2015) Occupancy estimation for rare species using a spatially-adaptive sampling design. *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.12499

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Risk, B.B., de Valpine, P. & Beissinger, S.R. (2011) A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. *Ecology*, **92**, 462–474.

Royle, J.A. & Dorazio, R.M. (2006) Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 249–263.

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

Ruiz-Gutierrez, V. & Zipkin, E.F. (2011) Detection biases yield misleading patterns of species persistence and colonization in fragmented landscapes. *Ecosphere*, **2**, 1–14. doi: 10.1890/ES10-00207.1

Sauer, J., Fallon, J. & Johnson, R. (2003) Use of North American Breeding Bird Survey data to estimate population change for bird conservation regions. *Journal of Wildlife Management*, **67**, 372–389.

Sauer, J.R. & Link, W.A. (2011) Analysis of the North American Breeding Bird Survey using hierarchical models. *The Auk*, **128**, 87–98.

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.

van Strien, A.J., Termaat, T., Groenendijk, D., Mensing, V. & Kery, M. (2010) Site-occupancy models may offer new opportunities for dragonfly monitoring based on daily species lists. *Basic and Applied Ecology*, **11**, 495–503.

Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J. & Martin, T.G. (2013) Realising the full potential of citizen Science monitoring programs. *Biological Conservation*, **165**, 128–138.

Vogel, L.S. & Pechmann, J.H.K. (2010) Response of fowler's toad (*Anaxyrus fowleri*) to competition and hydroperiod in the presence of the invasive coastal plain toad (*Incilius nebulifer*). *Journal of Herpetology*, **44**, 382–389.
Williams, B.K., Eaton, M.J. & Breininger, D.R. (2011) Adaptive resource management and the value of information. *Ecological Modelling*, **222**, 3429–3436.
Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations: Modeling, Estimation, and Decision Making*. Elsevier, San Diego, CA, USA.

Received 31 October 2015; accepted 15 January 2016
Handling Editor: Nigel Yoccoz

Supporting Information

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

Appendix S1. Single-season, single-species, false-positive occupancy model, with covariates on detection probability.