Single-visit dynamic occupancy models: an approach to account for imperfect detection with Atlas data

Michelle A. Peach* iD**, Jonathan B. Cohen and Jacqueline L. Frair**

Department of Environmental and Forest Biology, State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY, USA

Summary

1. Atlas data provide biodiversity information at a relatively fine spatial grain over a broad spatial extent and, increasingly, at multiple points in time, which make them invaluable for understanding processes that affect species distributions over time. The effect of survey effort on species detection has long been appreciated and Atlases typically include survey standards and records of effort, but challenges remain in analysing Atlas data that have not been collected using a repeated sampling protocol designed to correct for imperfect detection.

2. We developed a single-visit dynamic occupancy model to quantify the effects of climatic and land-use drivers on local species extinction and colonization while accounting for imperfect detection using repeat Atlas data. We evaluated model stability using data simulated under alternative scenarios and, ultimately, applied the model to empirical data for Canada warbler *Cardellina canadensis*, a wide-spread species exhibiting a long-term population decline.

3. At sample sizes that are realistic for many Atlases (n = 1000–10 000 independent survey blocks), our models produced unbiased estimates of detection, occupancy, colonization and extinction parameters. Slope estimates for explanatory covariates were somewhat less stable than overall occupancy, colonization and extinction rates, with covariate effects being sensitive to the total number of, and relationships among, explanatory variables.

4. In comparison to other analyses of Canada warbler distributions that indicated minor changes over time, our approach identified a widespread decline in occupancy probability across New York, consistent with the broader population trend, particularly in the areas where it was initially more likely to occur.

5. *Synthesis and applications.* A single-visit dynamic occupancy model is a novel method for analysing common, ecologically valuable datasets, such as Atlases, that lack repeated sampling necessary to correct for imperfect detection using alternative multi-season occupancy modelling approaches. As a result, using this method can improve understanding of species distributions and factors that shape them over time, thereby providing more accurate information to guide conservation and management.

Key-words: Bayesian hierarchical models, biodiversity, breeding bird atlas, *Cardellina canadensis*, colonization, dynamic occupancy modelling, extinction, simulation, species distributions, survey effort

Introduction

To understand drivers of change in biodiversity over time, comparable data are required on species' occurrence and environmental conditions. Given the multi-scale nature of species–habitat selection and broad species ranges, such data must be collected at a relatively fine-grain over broad spatial extents. Moreover, because the goal is to make inferences about changes in species distribution over time, both sets of data ideally should reference multiple points in time rather than different locations at a single point in time (i.e. a space-for-time substitution). Although landscape-level environmental data have been available and periodically updated for decades (e.g. National Land Cover Dataset 1992 cited in Vogelmann *et al*. 2001), comparable and repeated surveys of species' occurrences are more recent.

*Correspondence author. E-mail: mapeachl@syr.edu

As a result, researchers have only recently begun exploring how to best utilize large-scale, repeat surveys to evaluate spatio-temporal drivers of biodiversity change.

Coordinated citizen science efforts that incorporate long-term monitoring of permanent survey locations, such as Atlases and breeding bird surveys, provide substantial biodiversity data (Donald & Fuller 1998; Tulloch *et al.* 2013). Atlases record species' occurrences from surveys that are conducted in relatively small sampling blocks (on the order of 25 km^2) that often cover entire states or countries. First established in the 1950s, Atlas data are currently available for avian, mammalian, reptilian, amphibian and fish species world-wide. Breeding bird atlas data in particular are available in over 50 countries on 6 continents (Gibbons *et al.* 2007; Dunn & Weston 2008). Data from single Atlases have been used to analyse patterns of species richness and beta diversity (Desrochers, Kerr & Currie 2011; Fitterer *et al.* 2012), describe species–habitat relationships (Glennon & Porter 2005; Ferenc *et al.* 2014) and predict species' sensitivity to climate change (Virkkala *et al.* 2008; DesGranges & Morneau 2010). Although still rare, repeated Atlases that provide data for the same locations at multiple points in time are in progress or completed in at least nine US states (U.S. Geological Survey 2014). Repeated Atlases enable dynamic analyses of how species' ranges have shifted over time in response to changing habitat amount and fragmentation (Zuckerberg & Porter 2010; van der Hoek, Renfrew & Manne 2013), climate (Thomas & Lennon 1999; Zuckerberg, Woods & Porter 2009; Virkkala *et al.* 2014) and interacting biotic and abiotic factors (Melles *et al.* 2011; Bradshaw *et al.* 2014). Although repeated Atlases enable broader exploration of changing species' distributions, imperfect species detection during the surveys poses a non-trivial challenge to accurate inference.

The sampling design leading to many Atlas-based products pre-dates modern occupancy methods that formally estimate detection probability as part of the survey process (Mackenzie *et al.* 2006), instead adopting survey standards, such as minimum number of hours spent surveying a block, to favour higher species detection probabilities. Failure to account for imperfect detection, even when detection probability is high, consistently underestimates the probability of site occupancy and yields biased inference about drivers of population change (Gu & Swihart 2004). Estimates tend towards zero when detection is not accounted for, but the direction and magnitude of bias depends on the relationship between detection probability and occupancy (Kéry, Gardner & Monnerat 2010). In an occupancy framework, quantifying detection probability typically relies on repeated sampling over a limited time at a given location. In a novel dynamic (or multi-season) modelling effort using Atlas data, Sadoti *et al.* (2013) used spatial rather than temporal replicates by aggregating two to four survey blocks into 'sites' enabling an estimate of detection probability for Canada Warbler *Cardellina canadensis* (CAWA hereafter) at the site level and occupancy probability at the block level. A critical assumption of this approach, as with all occupancy models, is that replicate samples (here adjacent BBA blocks) have the same true state of occupancy (i.e. population closure). With spatial replication, differences between blocks in sampling and environmental characteristics can lead to inaccurate estimates of detection probability (Kéry & Royle 2015). Moreover, the spatial-replicate approach reduces the effective sample size and consequent statistical power, which will be particularly problematic for rare species that are already data-limited (Sadoti *et al.* 2013).

A promising alternative when repeat visits are lacking involves environmental covariates that are strong predictors of detectability and occupancy to help statistically separate these two processes in an occupancy modelling framework (Lele, Moreno & Bayne 2012). Formulated as a single-visit static (or single-season) occupancy model, the approach requires one unique continuous variable informing occupancy probability and another for detection probability. Other variables (continuous or categorical) can be included and shared between detection and occupancy. The Lele, Moreno & Bayne (2012) model improved analyses of Atlas-style data by removing a source of bias (imperfect detection) without the attendant assumptions and reductions in sample size associated with spatial replicates. Importantly, Atlas data typically include a record of effort, or the amount of time spent surveying, which provides the unique, continuous variable needed to inform detection probability. Effort has regularly been included as the best available surrogate for detection probability in a non-occupancy modelling framework (e.g. Gaston *et al.* 2008), although always as a linear predictor. The probability of detection might be expected to increase linearly with low levels of effort, but eventually approaches unity where additional effort has negligible impact. As a result, including effort as a power term would more accurately describe the expected relationship with detection probability.

Herein, our goal was to extend the Lele, Moreno & Bayne (2012) approach to a multi-season, or dynamic, model for use with repeated Atlas-style data. Our specific objectives were to: (i) evaluate the ability of a dynamic occupancy model to produce unbiased estimates of site occupancy, extinction and colonization, and their drivers, under sample sizes typical for Atlas data, (ii) improve how effort informs the detection process by incorporating a power function, and (iii) compare occupancy, extinction and colonization estimates for CAWA in New York State using our dynamic occupancy model vs. the spatial-replicate approach applied by Sadoti *et al.* (2013). Ultimately, we provide guidance on applying our single-visit dynamic occupancy model to data for other systems to enable accurate estimation of population change and the drivers of change when replicate samples are lacking.

Materials and methods

OCCUPANCY MODEL

Expanding the single-visit static model (Lele, Moreno & Bayne 2012) to a dynamic occupancy approach still requires detection to be modelled with at least one unique continuous covariate. Occupancy, colonization and extinction models can be identical, but each needs to include a continuous covariate not in the detection model. Following the Bayesian dynamic occupancy approach of Royle & Kéry (2007), we modelled occupancy probability in time one using the equation

$$\text{logit}(\psi_{1,j}) = \beta_{0o} + \beta_{1o}x_{1,j} + \ldots + \beta_{Uo}x_{U,j} \qquad \text{eqn 1}$$

where $\psi_{1,j}$ is the probability of occupancy at site j in time one as a function of U covariates. We modelled the probability of colonization and extinction using the equations

$$\text{logit}(\gamma_j) = \beta_{0c} + \beta_{1c}x_{1,j} + \ldots + \beta_{Uc}x_{U,j} \qquad \text{eqn 2}$$

$$\text{logit}(\varepsilon_j) = \beta_{0e} + \beta_{1e}x_{1,j} + \ldots + \beta_{Ue}x_{U,j} \qquad \text{eqn 3}$$

where γ_j and ε_j are probabilities of colonization and extinction, respectively, each of which is a function of U covariates. Detection is usually modelled as a continuous function of effort (e.g. Sadoti *et al.* 2013), but such an approach does not account for the nonlinear relationship between time spent in an area and the probability of detecting a species. Thus, we took a novel approach incorporating effort as a power term using the equation

$$p_{*i,j} = 1 - (1 - \text{logit}^{-1}(y))^{E_{i,j}} \qquad \text{eqn 4}$$

where $p_{*i,j}$ is the probability of detection at site j in time i and $\text{logit}^{-1}(y)$ describes the detection rate for one unit of effort. Here, y could be modelled as intercept-only or using linear covariates. Equation (4) is similar to the model describing the probability of detecting a species at least once during repeat surveys (Mackenzie & Royle 2005), but units of effort ($E_{i,j}$) replace the number of surveys (K).

SIMULATION

Using eqns (1–4) we simulated one dataset for each of 69 scenarios to explore the effect of sample size, detection probability, occupancy probability and covariate relationships on parameter estimation in a single-visit dynamic occupancy framework. In each scenario we evaluated the estimates of population parameters (occupancy, colonization, extinction and detection) as well as estimates of regression coefficients for predictor variables.

Our simplest scenarios included a single covariate in each of the occupancy, colonization, extinction and detection models. The occupancy covariate (eqn 1) was simulated from

$$x_{1,j} \sim \text{Uniform(min,max)} \qquad \text{eqn 5}$$

We assumed colonization and extinction models (eqns 2 and 3) were identical and a function of the same covariate as occupancy, with new values to reflect random fluctuations over time. The colonization and extinction covariate was simulated from

$$x_{2,j} \sim \text{Normal}(x_{1,j}, \sigma) \qquad \text{eqn 6}$$

In our simplest scenarios, we assumed y in eqn (4) was a constant (i.e. an intercept-only linear model). The number of hours spent surveying, $E_{i,j}$, was simulated at time i in site j as

$$E_{i,j} \sim \text{TruncatedNormal}(\mu, \sigma, a, b) \qquad \text{eqn 7}$$

where μ represents the mean value, σ is the standard deviation, and a and b set upper and lower truncation values. For simplicity, we assumed that y, and therefore the probability of detection, remained constant over both time periods in all scenarios.

We used eqns (1–7) as the basis for our simulations (simulation values in Table S1, Supporting Information). We generated more complex occupancy scenarios by incorporating additional covariates into eqn (1) and replacing β_{0o} with a random effect $\alpha_{j(i)}$ modelled as

$$\alpha_j \sim \text{Normal}(\mu, \sigma) \qquad \text{eqn 8}$$

We assumed sites were evenly distributed among $j = 1 \ldots J$ groups. We generated 10 groups for each scenario except the smallest sample sizes ($n = 1000$), which had five groups. In scenarios where we added covariates to $\psi_{1,j}$ (eqn 1), the same covariates were added to γ_j and ε_j (eqns 2 and 3) and assumed to remain constant over time or to deviate based on a normal distribution (eqn 6). We created more complex detection probability scenarios by modelling y in eqn (4) as a linear function of one or two variables that were either unique or shared with the occupancy, colonization and extinction models.

We generated sample sizes roughly equivalent to those of an Atlas in a small state (1000), a large state (5000) and multiple states (10 000). To vary occupancy probability, we altered the slope of a linear covariate to establish low ($\psi = 0.11$ on average), medium ($\psi = 0.40$), high ($\psi = 0.71$) and very high ($\psi = 0.90$) initial occupancy states. In order to more systematically explore the limitations of our approach, we also simulated data with fixed occupancy probabilities ranging from 0.1 to 0.9 using our simplest models and largest sample size. We varied detection probabilities by changing the intercept or one of the slopes in y to represent species with low ($p_* = 0.01$ on average), intermediate ($p_* = 0.05$) and high ($p_* = 0.30$) detection probabilities per unit effort (hours in the case of atlas data).

We modelled latent occupancy (i.e. true occupancy) using the equations

$$Z_{i,j} \sim \text{dbern}(\psi_{i,j}) \qquad \text{eqn 9}$$

$$\psi_{1,j} = 1/(1 + \exp(\text{logit}^{-1}(\psi_{1,j}))) \qquad \text{eqn 10}$$

$$\psi_{2,j} = Z_{1,j} \times (1 - 1/(1 + \exp(\text{logit}^{-1}(\varepsilon_j)))) \\ + (1 - z_{1,j}) \times 1/(1 + \exp(\text{logit}^{-1}(\gamma_j))) \qquad \text{eqn 11}$$

where $Z_{i,j}$ represents occupancy in time i at site j (Mackenzie *et al.* 2006). We then used latent occupancy estimates to derive occupancy probability (mean of $Z_{i,j}$) and colonization and extinction probabilities (differences between $Z_{1,j}$ and $Z_{2,j}$).

All simulations were completed using ʀ 3.2.0 and Bayesian model fitting was performed using the ʀ package rjags which calls the JAGS software (Plummer 2002, 2015; R Core Team 2015), using uninformative priors, three chains, 50 000

iterations and a burn-in of 500 iterations (see ʀ code in Appendix S1).

APPLICATION TO NEW YORK BREEDING BIRD ATLAS DATA

We reanalysed BBA CAWA data used by Sadoti *et al.* (2013) with input from the authors and data from the New York State Department of Environmental Conservation (2007a, b). CAWA is considered a High Priority Species of Greatest Conservation Need in NY (New York State Department of Environmental Conservation 2015) and has experienced an overall decline in the US over the past 40 years (Sauer *et al.* 2011). It is associated with a variety of forested habitats and was observed in 32% of blocks in the first BBA.

For the BBA, New York was divided into 5332 25-km^2 blocks, each of which was surveyed by volunteers during the periods 1980–1985 and 2000–2005 (Andrle & Carroll 1988; McGowan & Corwin 2008). Sadoti *et al.* (2013) aggregated blocks into sites containing two to four contiguous blocks, with each block serving as a spatial replicate for site-level detection. In order to reduce violations to the population closure assumption, they excluded all blocks with <50% forest cover to ensure adequate habitat availability and <90% of species detected in a single sampling year for each BBA period. They also excluded blocks lacking effort data in both BBA periods but retained blocks with effort for a single period. Finally, they excluded all sites containing <2 blocks, leaving a total of 1458 blocks and 504 sites.

We replicated the Sadoti *et al.* (2013) covariates and block selection to the extent possible (G. Sadoti provided a list of blocks and covariate values to ensure accurate replication). While they could retain blocks for analysis in a single BBA period even if they were excluded from the other period, our approach required all blocks be included in both periods. When effort was unavailable for a single BBA period, we substituted the average effort for that period and coded the observed occupancy as NA. This allowed us to retain the block for information on detection probability for the BBA period during which effort data were available without including occupancy information when they were not. We again estimated latent occupancy to derive population-level estimates of occupancy, colonization and extinction (eqns 9–11).

In order to make direct comparisons with the spatial-replicate analyses we aggregated blocks to the site level and considered a site to be occupied if any of the blocks within it was predicted to be occupied ($Z_{i,j} = 1$). We estimated block-level occupancy probabilities based on block-level covariates and estimated model coefficients derived from the single-visit dynamic occupancy model. We did the same for site-level occupancy probabilities using coefficients estimated from Sadoti *et al.*'s (2013) model. Block- and site-level occupancy probabilities in the second BBA were calculated using the equation

$$\psi_{1,j} \times (1 - \varepsilon_j) + (1 - \psi_{1,j}) \times \gamma_j$$

We conducted a second analysis of the BBA CAWA data with an expanded sample size because we did not need to exclude as many blocks to meet our model assumptions. For the expanded analysis, we focused on block-level results and only excluded blocks with no observation data in either BBA period ($n = 5007$).

When analysing both BBA data subsets, we estimated occupancy in the first BBA ($\psi_{1,j}$), as well as colonization (γ_j) and extinction (ε_j) in the second BBA at the block level using the equations

$$\text{logit}(\psi_{1,j}) = \beta_{0o} + \beta_{1o} \times \text{ELEV}_j + \beta_{2o} \times \text{ELEV2}_j \\ + \beta_{3o} \times \text{FOREST}_j + \beta_{4o} \times \text{ACOV}_{1,j}$$

$$\text{logit}(\gamma_j) = \beta_{0c}$$

$$\text{logit}(\varepsilon_j) = \beta_{0e} + \beta_{1e} \times \text{EDGE j} + \beta_{2e} \times \text{NEIGHBORS}_j$$

where the variables included were identified in the most competitive model of Sadoti *et al.* (2013) and β_i are the regression coefficients for the variables indicated (See Appendix S3 for a description of covariates).

We modified Sadoti *et al.*'s (2013) detection equation to reflect our approach to account for the nonlinear relationship between effort in block *j* during BBA survey *i* and detection using eqn (6) where

$$\text{logit}(y) = \beta_{0d} + \beta_{1d} \times \text{FOREST}_i + \beta_{2d} \times \text{DETECT80}_{i,j} \\ + \beta_{3d} \times \text{NEIGH3}_{i,j} + \beta_{4d} \times \text{Y1980}_j + \beta_{5d} \times \text{Y1981}_j \\ + \beta_{6d} \times \text{Y1982}_j + \beta_{7d} \times \text{Y1983}_j + \beta_{8d} \times \text{Y1985}_j + \beta_{9d} \\ \times \text{Y2000}_j + \beta_{10d} \times \text{Y2001}_j + \beta_{11d} \times \text{Y2002}_j + \beta_{12d} \\ \times \text{Y2003}_j + \beta_{13d} \times \text{Y2004}_j + \beta_{14d} \times \text{Y2005}_j$$

and $E_{i,j}$ is the effort in block *j* during BBA survey *i* in hours.

Ultimately, we had to remove NEIGH3, a spatial autocovariate, and Y1980 through Y2005, which were dummy variables indicating the primary year (Y) of sampling using 1984 as a reference, due to a lack of convergence in estimates of coefficients for those and other variables when they were included in the model. We expect these modifications to have a relatively minor effect on predictions. Five of 11-year parameters included 0 in the 95% confidence interval. The remaining year parameters, primarily from the second BBA period, were negative, which implies that later years contributed less to detection probability than the 1984 baseline. Given that average effort increased in the second BBA, we suspect the negative year effects in the second BBA period were an artefact of confounding true site-level occupancy probability declines with false reductions in block-level detection probability. NEIGH3 was strongly correlated with other spatial covariates remaining in the model.

All analyses were completed using ʀ 3.2.0 (R Core Team 2015). Spatial covariates were calculated using the packages maptools, rgdal and spdep (Bivand, Hauke & Kossowski 2013; Bivand, Keitt & Rowlingson 2015; Bivand & Lewin-Koh 2015). We used JAGS and the R package rjags to conduct Bayesian analyses (Plummer 2002, 2015). As with the simulated data, we used uninformative priors.

Results

SIMULATION

In general, results for the most complex scenarios followed the same patterns as those for the simplest scenarios. For ease of interpretation, we present the results for the simplest scenarios (where occupancy, colonization and extinction are predicted by a single linear covariate and

detection has a constant hourly rate with effort included as a power term) except where more complex scenarios produced different outcomes.

Our simulations revealed that extremes in initial occupancy probability challenge inference regarding colonization and extinction probabilities. Colonization probability estimates exhibited a slight negative bias when initial occupancy probability was very high and true detection probability low (Figs 1 and S2), whereas extinction probability estimates exhibited a slight positive bias when initial occupancy probability was low regardless of detection probability (Fig. 1, Table S2). Aside from these exceptions, and consistent with expectations, estimates of population parameters (i.e. occupancy, colonization and
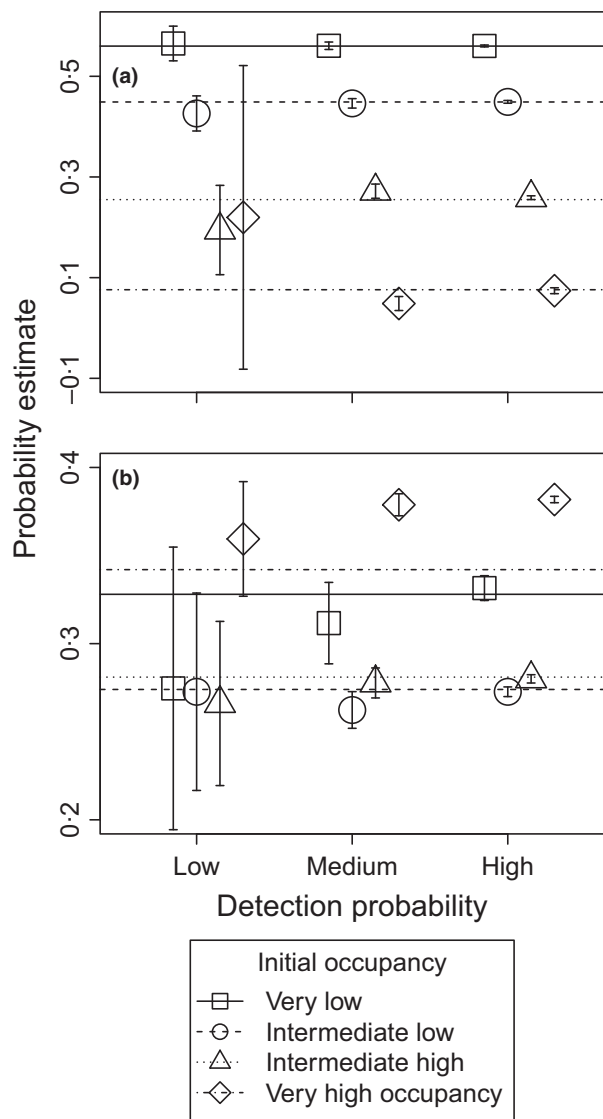
Fig. 1. Estimates of (a) colonization and (b) extinction probability across three levels of detection probability and four levels of initial occupancy probability (n = 1000). Estimates are based on a single linear covariate predicting colonization and extinction and a constant describing detection. Actual values are indicated by horizontal dashed lines. Standard deviation bars are shown but may be small than symbols.

extinction probability) generally became more precise given increasing sample size, higher initial occupancy probability and higher detection probability (Fig. S1). Similar patterns were observed for estimates of covariate effects on occupancy, colonization and extinction probabilities. Parameter estimates for the occupancy covariate when associated with high occupancy and low detection probability exhibited a slight negative bias (Fig. 2a). The colonization covariate was biased at very high occupancy levels and became increasingly imprecise as detection probability decreased (Fig. 2b). In contrast, the covariate associated with extinction probability showed a negative bias given extremely low occupancy combined with low detection probability (Fig. 2c). Population parameters were considered more robust than covariate coefficient estimates because the latter indicated bias at less extreme initial occupancy levels.

True detection probability exhibited a strong influence on precision of population parameters (Fig. 1) and covariate effects (Fig. 2). The only exception was for estimates of detection parameters themselves, which, consistent with expectations as a proportion increases from near 0 to 0·5, became less precise as detection probability increased (Figs 2d and S4). Moreover, estimates of regression coefficient effects on detection were more robust to variation in sample size, occupancy level and detection probability than other covariates except at the highest detection levels (Fig. S4).

Estimates of population parameters remained unbiased under more complex scenarios in which we added an additional covariate to occupancy, colonization and extinction models as well as a unique linear covariate to detection (Fig. 3a), a linear covariate to detection that was also shared with occupancy, colonization and extinction models (Fig. 3b), and two additional linear covariates – one unique to detection and the other shared with occupancy, colonization and extinction (Fig. 3c). Estimates of covariate effects were less robust, showing a slight and inconsistent bias across multiple levels of true occupancy probability given a detection model based on a single unique linear covariate (Figs 4a and S5), a linear covariate shared with occupancy, colonization and extinction models (Figs 4b and S5), or one linear covariate unique to detection as well as one shared with the occupancy, colonization and extinction models (Figs 4c and S5).

APPLICATION TO NEW YORK BREEDING BIRD ATLAS DATA

Our estimates of occupancy probability for CAWA (ψ_1 = 0·73 and ψ_2 = 0·66) were intermediate between those assuming perfect detection (ψ_1 = 0·71 and ψ_2 = 0·6) and those based on a space-for-time substitution to correct for imperfect detection (ψ_1 = 0·81 and ψ_2 = 0·71). When mapping occupancy probability for CAWA across the state, our model produced more heterogeneous predictions (compare the a panels to the b panels in Fig. 5).
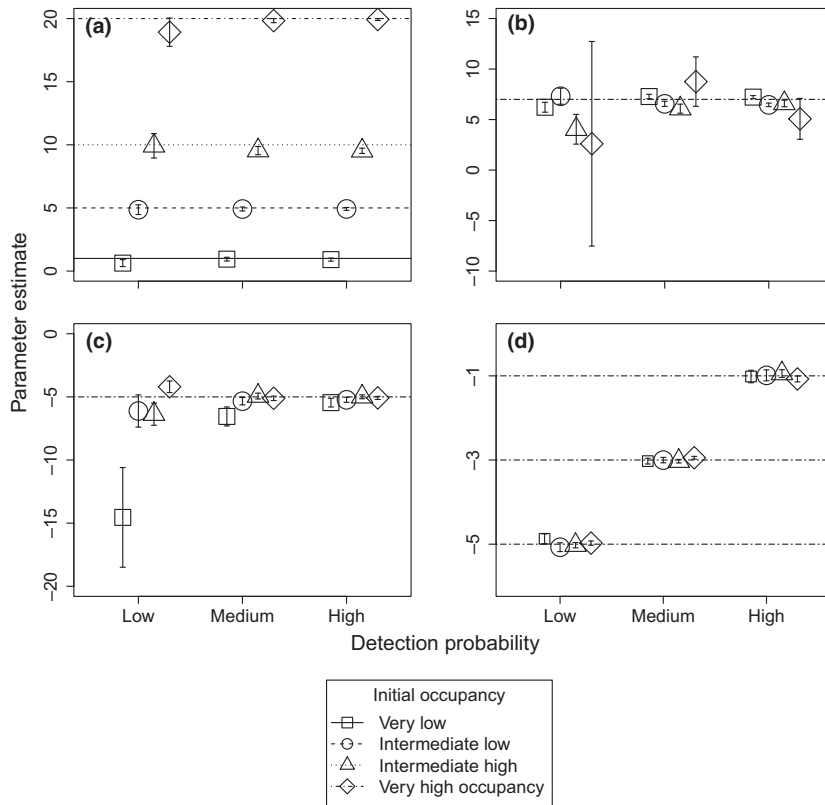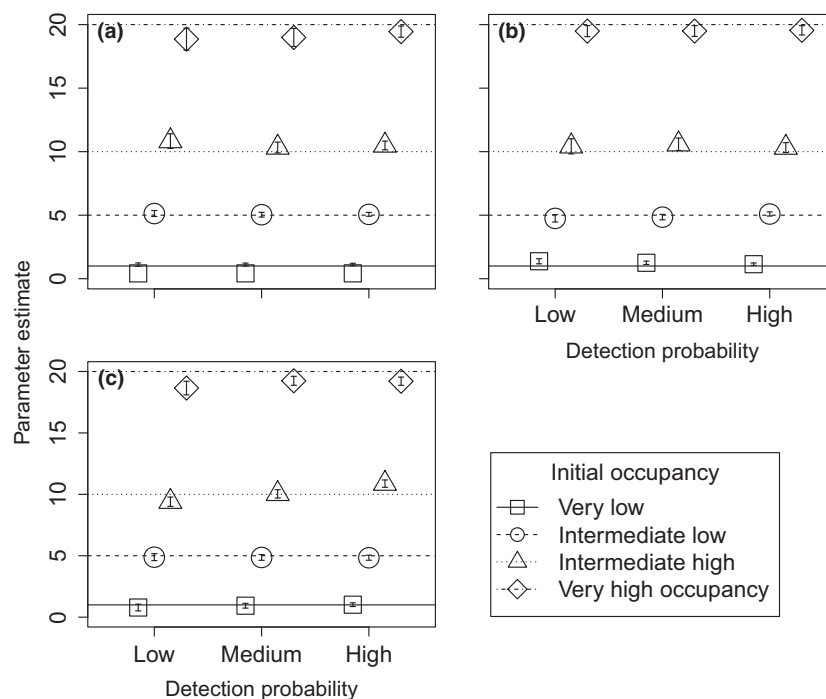
Fig. 2. Estimates of a regression coefficient for the effect of a random uniform covariate on (a) occupancy, (b) colonization and (c) extinction probabilities and (d) a constant that when inverse-logit transformed represents the hourly detection across three levels of detection probability and four levels of occupancy probability (n = 10 000). Actual values are indicated by horizontal dashed lines. Standard deviation bars are shown but may be small than symbols.
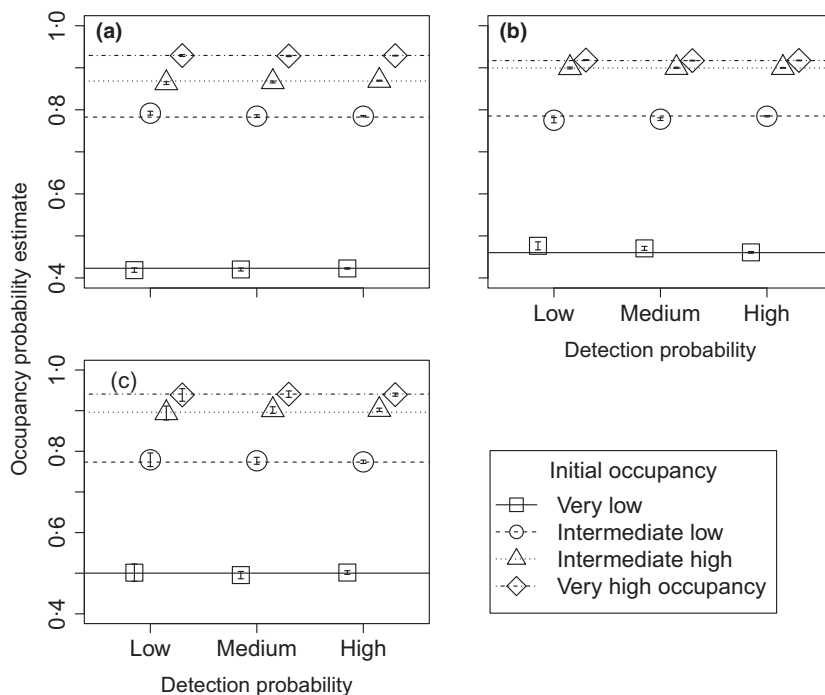
Fig. 3. Estimates of the occupancy probability in time one when the linear part of detection model contained (a) a single, unique covariate, (b) a single covariate that was also shared with the occupancy model and (c) one unique covariate and one covariate that was shared with the occupancy model. Actual values are indicated by horizontal dashed lines. Wtandard deviation bars are shown but may be small than symbols.

This pattern was driven by the larger grain used to determine detection in the spatial-replicate approach (100 km² sites vs. 25 km² blocks) and the tendency of spatial replicates to underestimate detection and, by extension, over-estimate occupancy. Differences between the two approaches were most pronounced during the second BBA period (a2 vs. b2 in Fig. 5), where the spatial-replicate approach indicated the majority of landscape to have an occupancy probability >0·8 for CAWA, whereas our single-visit dynamic occupancy model estimated probability of occupancy to be no higher than 0·8 anywhere in the state, consistent with the long-term decline observed for

Fig. 4. Estimates of an occupancy regression coefficient for the effects of one of two random uniform covariates when the linear part of detection mode (eqn 6) contained (a) a single, unique covariate, (b) a single covariate that was also shared with the occupancy model and (c) one unique covariate and one covariate that was shared with the occupancy model. Actual values are indicated by horizontal dashed lines. Standard deviation bars are shown but may be small than symbols.

this species. Although overall patterns in block-level occupancy probabilities were similar for our model whether based on the full dataset or the restricted subset applicable to the spatial-replicate approach (c1 vs. b1 in Fig. 5), being able to fit models using the full set of survey blocks is a clear advantage of our modelling approach. Importantly, to make a fair comparison of these two approaches, occupancy estimates for the second BBA period using our model were restricted to an intercept-only colonization model, which may have been reasonable for the Sadoti *et al.* (2013) model that focused only on areas having a threshold level of suitable habitat, but would be inappropriate when including blocks encompassing a greater range of suitable habitat. Moreover, our estimates of regression parameters were not directly comparable to Sadoti *et al.*'s (2013) because we included effort as a power term in the detection model and had to remove other variables due to convergence issues.

Discussion

We extended the single-visit, single-season occupancy model of Lele, Moreno & Bayne (2012) to the multi-season case, providing a substantively improved set of analytical tools for ecologists working with repeated instances of single-visit monitoring data to understand the drivers of change in species distributions over time. Several advantages were apparent when comparing our single-visit dynamic occupancy model to a spatial-replicate dynamic occupancy model (sensu Sadoti *et al.* 2013) in the case of a declining songbird (CAWA) in New York State. Foremost, our approach maintained a larger sample size. Rather than aggregating blocks to effectively yield spatial replicates at a larger site-level, our approach retained the

original data resolution enabling finer scale inference, greater coverage of potential environmental or land-use drivers affecting the distributional shifts of species, and greater statistical power. Importantly, our single-visit dynamic occupancy model was proven to provide unbiased and precise estimates of population parameters and covariate effect sizes given the sample size, detection probability and occupancy levels observed for CAWA in NY. In light of this, and given the greater number of assumptions and data manipulations required by the spatial-replicate approach, we believe our approach provides a more robust picture of CAWA status in NY. Given a more explicit treatment of the detection process, our model estimated a higher probability of local CAWA extinction and lower probability of persistence between 1980 and 2005 than the spatial-replicate approach. We documented that declines in occupancy probability were particularly high in areas where CAWA were initially more likely to occur which, in combination with the lower probability of colonization estimated by our model, points to species decline rather than distributional shift. Such a case warrants more attention to local habitat characteristics in population strongholds to better understand local drivers of decline, for example, changes in understorey habitat quality. In contrast, maps from the spatial-replicate approach indicated relatively small changes in CAWA patterns of occupancy over time, providing little justification or guidance for conservation efforts.

Importantly, using spatial replicates confounds species detection with site occupancy processes. Under the spatial-replicate design, when one block in a site was occupied, all others were assumed to be occupied as well, which tended to underestimate detection probability because in reality some blocks could be unoccupied (Kéry

First BBA period   Second BBA period
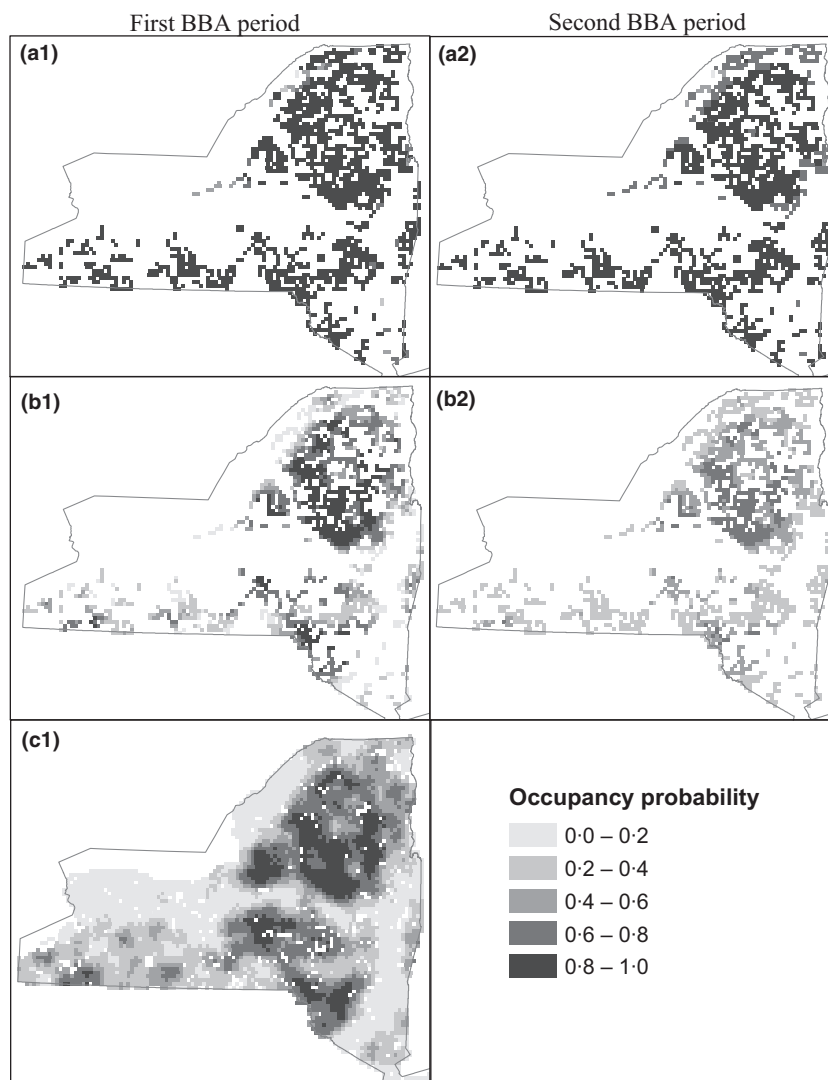
Fig. 5. Estimated occupancy probabilities for *Cardellina canadensis* in New York State BBA blocks based on (a) spatial-replicate model, (b) a single-visit dynamic occupancy model using an identical subset of BBA blocks as the spatial-replicate model (Sadoti *et al*. 2013) and (c) a single-visit dynamic occupancy model using a larger subset of BBA blocks for the (1) first and (2) second Atlas periods. A single-visit dynamic occupancy model using a larger subset of BBA blocks was not possible for the second Atlas period given the Sadoti *et al*. (2013) variables and models. Occupancy probabilities in the first Atlas period were estimated based on the occupancy equations in each model.

& Royle 2015). This resulted in an inflated estimate of occupancy probability as we observed in our application for CAWA – a problem circumvented when applying our single-visit dynamic occupancy model. Although we limited our consideration of covariates to those identified by Sadoti *et al*. (2013), our approach enables greater resolution of potential forest cover effects (no need for *a priori* exclusion of blocks of low habitat amount), as well as additional covariates given the increased statistical power associated with retaining the original sample size of survey blocks.

In addition to Lele, Moreno & Bayne's (2012) requirement that detection and occupancy models each have a unique, continuous covariate, our simulations indicated that three other guidelines apply to single-visit occupancy models whether static or dynamic in order to achieve unbiased and precise estimates: (i) ensure adequate numbers of occurrence, colonization and extinction events, (ii) include a continuous predictor of detection probability as a power term, and (iii) consider the value of including covariates shared by the detection and occupancy models. We address each of these in the paragraphs to follow.

First, our results highlighted the importance of considering not only sample size but also the number of occurrence, colonization and extinction events when determining whether a single-visit approach will provide accurate parameter estimates. In particular, estimates of occupancy probability remained unbiased across our scenarios, whereas colonization and extinction estimates became biased as occupancy probability approached extremes (i.e. 0·1 or 0·9). Although we did not explicitly vary colonization and extinction rates or set initial occupancy probabilities <0·1 or >0·9, it is logical to conclude that this bias applies generally to species with low numbers of absolute colonization, extinction or occupancy events. Moreover, with ubiquitous species, that is, when baseline occupancy probability is high, parameter estimates are consistently biased regardless of sample size (Field, Tyre & Passingham 2005; Mackenzie & Royle 2005). Our estimates for CAWA remain robust despite these restrictions given its intermediate level of occupancy, colonization and extinction, However, ubiquitous species such as red-winged blackbird (*Agelaius phoeniceus*), rare species such as spruce grouse (*Falcipennis canadensis*) and

species with little change in distribution in NY might generate inaccurate estimates using our approach.

Second, a nonlinear detection model provides accurate parameter estimates and assumes a more realistic relationship between detection and effort. A nonlinear relationship also eliminates the parameter non-identifiability problem described by Knape & Korner-Nievergelt (2015) for single-visit occupancy models. When detection is fitted as a linear process, more than one set of parameter values could produce the same predictions for certain link functions. In our model, detection and occupancy regression parameters are uniquely identifiable regardless of the link functions chosen because our method models detection similarly to a repeat-visit framework. Each unit of time is treated as a repeat visit with the assumption that there is a detection rate associated with a base unit of effort (i.e. 1 h). Given adequate variability in effort, only a single combination of parameter values could produce the given set of observations. Our assumption that hours can be considered discrete units of effort is testable in the field.

Third, shared covariates can result in biased estimates for regression coefficients, although estimates of occupancy, colonization and extinction probabilities remain robust. We suspect that the relationship between covariates within and between occupancy, colonization, extinction and detection models will influence the accuracy of parameter estimates, particularly the magnitude of unique covariates relative to shared ones, and future studies could improve understanding of these relationships. Our CAWA analyses are unaffected by this caveat because the occupancy, colonization and extinction probabilities we examined are robust to these challenges. Interpretation of regression coefficients would have been subject to potential bias because per cent forest cover was used to model both occupancy and detection, but, based on our simulation results, that bias would have been slight.

One non-trivial consideration in the use of our approach is the time required because working with large sample sizes in a Bayesian framework can be processing intensive and thus time-limiting, particularly as the number of covariates or sample size increases. However, it is possible to apply a single-visit dynamic occupancy approach in a maximum likelihood framework for which we provide sample code in Appendix S2. We preferred the Bayesian framework for this application due to the ease of incorporating a random effect and estimating latent occupancy, which makes it easy to derive population level occupancy, colonization and extinction probabilities. With different analysis questions or larger datasets, those benefits may not be worth the trade-offs in processing time.

Our modelling approach addresses a gap between the requirements of other multi-season occupancy models (i.e. repeated sampling) and many existing datasets. Although we applied it specifically to Atlas data, this method may be relevant to other types of data that lack repeated sampling. For example, Hostetler & Chandler (2015) compared a variety of state-space models describing

abundance that relied on covariates to predict detection. While they concluded that direct information about detection probability is preferable to using covariates, such information is not always available. In these cases, the single-visit dynamic approach is a useful method for more accurately describing population parameters and environmental relationships than ignoring the effects of imperfect detection. An accurate understanding of species distribution and the factors that shape them at meaningful spatial scales are essential for effective conservation and management. Our CAWA results highlight the necessity of properly accounting for imperfect detection to identify species in decline and guide conservation efforts.

Authors' contributions

Acknowledgements

Data accessibility

ʀ scripts to simulate and analyse simulated data: uploaded as online supporting information. NY BBA dataset as used by Sadoti *et al.* (2013); available from New York State Department of Environmental Conservation (2007a, b) was used in this manuscript. To request data contact: Breeding Bird Atlas Project, NYS Department of Environmental Conservation, 625 Broadway, Albany, NY 12233-4754 http://www.dec.ny.gov/cfmx/extapps/bba/DataUsePolicy.cfm.

References

Andrle, R.F. & Carroll, J.R. (1988) *The Atlas of Breeding Birds in New York State.* Cornell University Press, Ithaca, NY, USA.

Bivand, R.S., Hauke, J. & Kossowski, T. (2013) Computing the Jacobian in Gaussian spatial autoregressive models: an illustrated comparison of available methods. *Geographical Analysis*, **2**, 150–179.

Bivand, R., Keitt, T. & Rowlinson, B. (2015) rgdal: Bindings for the geospatial data abstraction library. R package version 1.0-4. Available at: https://CRAN.R-project.org/package=rgdal (accessed 27 January 2010).

Bivand, R.S. & Lewin-Koh, N. (2015) maptools: tools for reading and handling spatial objects. R package version 0.8036. Available at: https://CRAN.R-project.org/package=maptools (accessed 27 January 2010).

Bradshaw, C.J.A., Brook, B.W., Delean, S., Fordham, D.A., Herrando-Pérez, S., Cassey, P., Early, R., Sekercioglu, C.H. & Araújo, M.B. (2014) Predictors of contraction and expansion of area of occupancy for British birds. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140744.

DesGranges, J.-L. & Morneau, F. (2010) Potential sensitivity of Quebec's breeding birds to climate change. *Avian Conservation and Ecology*, **5**, 5.

Desrochers, R.E., Kerr, J.T. & Currie, D.J. (2011) How, and how much, natural cover loss increases species richness. *Global Ecology & Biogeography*, **20**, 857–867.

Donald, P.F. & Fuller, R.J. (1998) Ornithological atlas data: a review of uses and limitations. *Bird Study*, **45**, 129–145.

Dunn, A. & Weston, M. (2008) A review of terrestrail bird atlases of the world and their application. *Emu*, **108**, 42–67.

Ferenc, M., Sedláček, O., Fuchs, R., Dinetti, M., Fraissinet, M. & Storch, D. (2014) Are cities different? Patterns of species richness and beta diversity of urban bird communities and regional species assemblages in Europe. *Global Ecology and Biogeography*, **23**, 479–489.

Field, S.A., Tyre, A.J. & Passingham, H.P. (2005) Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management*, **69**, 473–482.

Fitterer, J.L., Nelson, T.A., Coops, N.C., Wulder, M.A. & Mahony, N.A. (2012) Exploring the ecological processes driving geographical patterns of breeding bird richness in British Columbia, Canada. *Ecological Applications*, **23**, 888–903.

Gaston, K.J., Jackson, S.F., Cantú-Salazar, L. & Cruz-Piñón, G. (2008) The ecological performance of protected areas. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 93–113.

Gibbons, D.W., Donald, P.F., Bauer, H.-G., Fornasari, L. & Dawson, I.K. (2007) Mapping avian distributions: the evolution of bird atlases: capsule an increasing proportion of atlases now map patterns of abundance but they are still a minority even though they require no more input of time or fieldworkers. *Bird Study*, **54**, 324–334.

Glennon, M.J. & Porter, W.F. (2005) Effects of land use management on biotic integrity: an investigation of bird communities. *Biological Conservation*, **126**, 499–511.

Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, **116**, 195–203.

van der Hoek, Y., Renfrew, R. & Manne, L.L. (2013) Assessing regional and interspecific variation in threshold responses of forest breeding birds through broad scale analyses. *PLoS ONE*, **8**, 1–12.

Hostetler, J.A. & Chandler, R.B. (2015) Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology*, **96**, 1713–1723.

Kéry, M., Gardner, B. & Monnerat, C. (2010) Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, **37**, 1851–1862.

Kéry, M. & Royle, J.A. (2015) *Applied hierarchical modelling in ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS*, Volume 1: Prelude and Static Models. Elsevier, Amsterdam, The Netherlands.

Knape, J. & Korner-Nievergelt, F. (2015) Estimates from non-replicated population surveys rely on critical assumptions. *Methods in Ecology and Evolution*, **6**, 298–306.

Lele, S.R., Moreno, M. & Bayne, E. (2012) Dealing with detection error in site occupancy surveys: what can we do with a single survey? *Journal of Plant Ecology*, **5**, 22–31.

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

Mackenzie, D.I., Royle, J.A., Pollock, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modelling: Inferring Patterns and Dynamics of Species Occurrence*. Elsevier, Amsterdam, The Netherlands.

McGowan, K.J. & Corwin, K. (2008) *The Second Atlas of Breeding Birds in New York State*. Cornell University Press, Ithaca, NY, USA.

Melles, S.J., Fortin, M.-J., Lindsay, K. & Badzinski, D. (2011) Expanding northward: influence of climate change, forest connectivity, and population processes on a threatened species' range shift. *Global Change Biology*, **17**, 17–31.

New York State Department of Environmental Conservation (2007a) New York State Breeding Bird Atlas. 1980-1985. Release 1.0, Albany, NY. Available at: http://www.dec.ny.gov/animals/7312.html (accessed 28 April 2016).

New York State Department of Environmental Conservation (2007b) New York State Breeding Bird Atlas 2000. 2000-2005. Release 1.0, Albany, NY [online]. Available at: http://www.dec.ny.gov/animals/7312.html (accessed 28 April 2016).

New York State Department of Environmental Conservation (2015) New York state species of greatest conservation need [online]. Available at: http://www.dec.ny.gov/docs/wildlife_pdf/sgnc2015list.pdf (accessed 16 September 2016).

Plummer, M. (2002) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria.

Plummer, M. (2015) Rjags: Bayesian graphical models using MCMC. R prackage version 4-6. Available at: https://CRAN.R-project.org/package=rjags (accessed 5 November 2015).

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Royle, J.A. & Kéry, M. (2007) A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 1813–1823.

Sadoti, G., Zuckerberg, B., Jarzyna, M.A. & Porter, W.F. (2013) Applying occupancy estimation and modelling to the analysis of atlas data. *Diversity and Distributions*, **19**, 804–814.

Sauer, J.R., Hines, J.E., Fallon, J.E., Pardieck, K.L., Ziolkowski Jr, D.J.. & Link, W.A. (2011) The North American breeding bird survey, results and analysis 1966-2009. Version 3.23.2001. USGU Patuxent Wildlife Research Center, Laurel, MD, USA.

Thomas, C.D. & Lennon, J.J. (1999) Birds extend their ranges northwards. *Nature*, **399**, 213.

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.

U.S. Geological Survey (2014) Status of North American atlases. Available at: http://www.pwrc.usgs.gov/bba/ [online] (accessed 16 September 2016).

Virkkala, R., Heikkinen, R.K., Lehikoinen, A. & Valkama, J. (2014) Matching trends between recent distributional changes of northern-boreal birds and species-climate model predictions. *Biological Conservation*, **172**, 124–127.

Virkkala, R., Heikkinen, R.K., Leikola, N. & Luoto, M. (2008) Projected large-scale range reductions of northern-boreal land bird species due to climate change. *Biological Conservation*, **141**, 1343–1353.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.L., Wylie, B.K. & van Driel, J.N. (2001) Completion of the 1990's National Land Cover Data Set for the coterminous United States. *Photogrammetric Engineering and Remote Sensing*, **67**, 650–662.

Zuckerberg, B. & Porter, W.F. (2010) Thresholds in the long-term responses of breeding birds to forest cover and fragmentation. *Biological Conservation*, **143**, 952–962.

Zuckerberg, B., Woods, A.M. & Porter, W.F. (2009) Poleward shifts in breeding bird distributions in New York State. *Global Change Biology*, **15**, 1866–1883.

Supporting Information

Details of electronic Supporting Information are provided below.

Appendix S1. R code used to simulate and analyse data in a Bayesian framework.

Appendix S2. R code to simulate and analyse data in a maximum likelihood framework.

Appendix S3. Sample R code to analyse data in a likelihood framework.

Fig. S1. Estimates of occupancy probabilities at varying sample sizes.

Fig. S2. Estimates of colonization and extinction probabilities at varying sample sizes.

Fig. S3. Estimates of the standard deviation of the random intercept in the occupancy model.

Fig. S4. Estimates of detection parameters for intercept-only models.

Fig. S5. Estimates of regression coefficients with varying models.

Table S1. Values used to simulate datasets.

Table S2. Accuracy of population parameter estimates.