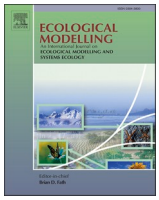
Ecological Modelling 478 (2023) 110303

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



An evaluation of multistate occupancy models for estimating relative abundance and population trends

Valerie A. Steen [a,*], Adam Duarte [b], James T. Peterson [c]

[a] Oregon Cooperative Fish and Wildlife Research Unit, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR, USA
[b] U.S.D.A. Forest Service, Pacific Northwest Research Station, Olympia, WA, USA
[c] U.S. Geological Survey, Oregon Cooperative Fish and Wildlife Research Unit, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, OR, USA

ARTICLE INFO

Keywords:
Abundance estimation
Count data
Detection probability
Multistate occupancy
Population trend
Power analysis

ABSTRACT

Detecting spatiotemporal changes in the abundances of organisms is key to effectively conserving species. While indices of abundance have long been used, there has been a shift toward model-based estimators that account for the detection process. Popular approaches including traditional occupancy models and N-mixture models entail tradeoffs. The traditional occupancy approach requires the researcher coarsen the characterization of abundance to the probability that a site is occupied or unoccupied. Conversely, N-mixture models make use of variation in counts, but perform poorly when individuals have low detectability or move into or out of sites between visits. Multistate occupancy models that differentiate relatively abundant from non-abundant states have the potential to fill this gap but have been underexplored. We conducted a simulation study to test whether multistate occupancy models could capture spatial abundance patterns and detect population declines in the face of low individual detection probability ($p \leq 0.3$) and unmodeled heterogeneity (e.g., that arising from individual movement). We considered 10,773 scenarios to examine the effects of differing amounts of heterogeneity as well as alternative study designs, population parameters, and modeling choices. We tracked bias in the proportion of sites estimated to be in the abundant state for single-season models, and power to detect a declining trend across multiple years. We also evaluated data diagnostic metrics to provide guidance to users. Multistate occupancy models were able to differentiate sites with higher abundances from sites with lower abundances when there were at least medium levels of spatial heterogeneity in true abundances. If different sites were randomly selected each year, power to detect even large population declines (65%) was poor (power < 0.8). However, if the same sites were surveyed each year, and a dynamic multistate occupancy was used, multistate occupancy models could detect (power ≥ 0.8) relatively small declines (5-40%) in 20% of scenarios, and frequently detect large declines of 45-60% (mean power $= 0.92$). Conservation decisions rely on detecting change reliably, rarely needing absolute abundance information. Multistate occupancy models can improve our ability to detect changing abundance while accommodating low individual detection probability and heterogeneity in count monitoring data.

1. Introduction

Estimating abundances of animals through space and time is the goal of many monitoring programs. Armed with this information, conservation practitioners can understand population trends and prioritize locations for protection and restoration. While raw count data are often used as indices of abundance, variation in such estimates can be confounded with patterns in the detection process and yield misleading results (Bayley and Peterson, 2001; Williams et al., 2002; Royle et al., 2007; Kéry et al., 2009). Furthermore, given that conservation of species often has diverse stakeholders involved, the tenability of reported trends or spatial patterns is often as important as the trends themselves (Duarte and Peterson, 2021). Therefore, analytical methods that explicitly address the detection process are ever more widely used (Mackenzie et al., 2017; Kéry and Royle, 2020).

Numerous analytical approaches have been developed to explicitly model detectability of individuals within a population and obtain direct estimates of abundance. Among the oldest of these, mark-recapture methods use unique identifiers to estimate abundance based on the number of marked individuals relative to unmarked individuals (Otis

* Corresponding author.
E-mail address: valerie.steen@gmail.com (V.A. Steen).

https://doi.org/10.1016/j.ecolmodel.2023.110303
Received 28 November 2022; Received in revised form 10 January 2023; Accepted 24 January 2023
Available online 7 February 2023

et al., 1978; Pollock et al., 1990; Nichols, 1992; Williams et al., 2002; Amstrup et al., 2005; Caci et al., 2013; Alonso et al., 2015). Aside from the few animals with unique and visible markings, this requires the intensive work of capturing and marking animals. Developed more recently, N-mixture models estimate individual detection probabilities and abundance from repeat visits to sites using counts of unmarked individuals and thus do not require capturing animals (Royle, 2004). Given this convenience, these models have exploded in popularity. However, N-mixture models have recently been shown to produce biased abundance estimates in the face of low individual detection probability ($p \leq 0.3$) and unmodeled heterogeneity in the count data (Barker et al., 2018; Duarte et al., 2018; Knape et al., 2018; Link et al., 2018), rendering them a poor choice for many study systems.

Another related approach to explicitly modeling detectability is to use occupancy models where counts of animals are reclassed as detected (1) or not detected (0) to estimate occupied and unoccupied site-level status (MacKenzie et al., 2002; Tyre et al., 2003). In occupancy models, detection probability refers to the detectability of the species – rather than the individuals – within a sample unit. Whereas unmodeled heterogeneity in count or capture-based models arises from a lack of site closure to, for example, individual movement, in occupancy models this lack of individual-level closure is less problematic unless it produces a change in occupancy state. In the common case where counts number higher than one, occupancy modeling entails an information reduction. However, this reduction can produce gains in the precision and accuracy of model estimates due to the more lenient population closure assumption (Mackenzie, 2006). Then, variation in abundance might be inferred through variation in occupancy or detectability estimates (Royle and Nichols, 2003); monitored as change in occupied area over time (e.g. Briscoe et al., 2021); or, inferred by taking advantage of the presumed occupancy-abundance relationship in ecological systems (Gaston et al., 2000). However, simulation-based studies have found: changes in abundance need to be large (Ellis et al., 2014) or species detectability needs to be high (Fuller et al., 2016) before occupancy models can detect a change; severe positive bias in abundance estimates can occur when model assumptions are violated (Stauffer et al., 2021); and the occupancy-abundance relationship does not always hold (Ellis et al., 2014; Stauffer et al., 2021). Thus, occupancy models may be biased or have low sensitivity to detect population change in many common circumstances.

Multistate occupancy models estimate the detectability and status of three or more occupancy states simultaneously (MacKenzie et al., 2009). By adding one or more additional states, they increase the sensitivity of the basic occupancy model to detect important differences in occupied sites. They have commonly been used to differentiate breeding categories of occupied sites (e.g., Nichols et al., 2007; Peterson and Shea, 2015; McGrady et al., 2017; Duarte et al., 2020; Clawson et al., 2022) and to identify disease status of individuals or populations (e.g., Elmore et al., 2014; Mosher et al., 2018; Rodrigues et al., 2020; Reddell et al., 2021). When the abundance across sites is of interest, relative abundance categories can be included, such as relatively abundant and not abundant. This can increase the ability of occupancy models to detect changing abundance. While the utility of this approach has been shown in a variety of systems (Royle and Link, 2005; Martin et al., 2010; Jensen and Vokoun, 2013; Peterson and Barajas, 2018; Whitlock et al., 2020) - it has not been widely applied. This is likely because multistate occupancy models have not been systematically evaluated for their ability to estimate spatiotemporal variation in abundance. Furthermore, in study systems where no a priori definition of abundance categories exist, guidance is needed on how to optimally define the number of animals that would constitute abundant versus not abundant states.

We used simulations to evaluate the ability of multistate occupancy models to estimate variation in abundance among sites (spatial variation) and changes in abundance when populations are declining across years or primary sampling occasions (temporal variation). We restricted simulation scenarios to situations where the count-based N-mixture model performs poorly (i.e., low detection and unmodeled heterogeneity of individuals). Specifically, we explored (1) how to determine the threshold count value that separates abundant and non-abundant states; (2) bias in estimates of the proportion of sites (spatial variation) in the abundant state given varying study designs elements (number of sites and visits); and (3) the power to detect a declining trend (temporal variation) in abundance given two common survey designs: a random-site design with different sites surveyed from year to year, and a fixed-site design using dynamic occupancy models with the same sites surveyed from year to year.

2. Materials and methods

2.1. Simulations

2.1.1. Overview

We generated observed animal count data for each visit at each site and then classed the observations into one of three states: not-detected; detected in the non-abundant state; and detected in the abundant state. To do this, we simulated true latent (i.e., unobserved) site occupancy; for occupied sites we simulated true latent abundances. We then generated the number of individuals observed (counts) from the latent abundances assuming a low individual detection probability. To generate sample data under different scenarios, we varied study design and population parameters by what we consider to be low, medium, or high levels for many studies and systems that do not meet the criteria needed to reliably fit N-mixture models (for distributions and parameter levels used see Table 1). Each unique combination of mean true abundance μ, number of sites i, and number of visits j, along with specified levels of among-site (spatial) and within-site heterogeneity (described below) in true abundance defined the basic study design and population parameter components of a scenario. For population trend simulations, we also evaluated random-site and fixed-site designs, alternative levels of population decrease, the number of years sampled, and whether the 'abundant' state was determined based on counts from just the first year or all years (described below). Collectively, this resulted in the evaluation of 10,773 different scenarios.

2.1.2. Single season; heterogeneity exploration

Spatial. - We explored the effects of spatial heterogeneity in true abundance on bias in abundant state probabilities and the ability to detect a trend (Fig. 1.2a). We evaluated four levels of spatial heterogeneity including a very high level because we wanted to verify our expectation that increasing levels of spatial heterogeneity would benefit multistate abundant occupancy models; as count distributions become longer right-tailed, a larger spread of values presumably facilitates differentiating abundance states (Table 1, Fig. 1.2). Our evaluation contained 108 scenarios across three levels of abundance, visits, and sites, and the four levels of spatial heterogeneity.

Within-site. - We also explored the effects of within-site heterogeneity in true abundance on bias in abundant state probabilities and the power to detect trends (Fig. 1.2b). These scenarios represent conditions where individuals enter/leave sample units between replicate surveys. In addition to including scenarios with no within-site heterogeneity, we explored four additional levels of heterogeneity (Table 1). Within-site heterogeneity was simulated for visits $j+1$ where true abundances were drawn from a gamma distribution based on the true abundance from the first visit (j) and a level of variation (Table 1). This evaluation included 405 scenarios across the three highest levels of spatial heterogeneity; three levels of abundance, visits and sites; and the five levels of within site heterogeneity.

2.1.3. Trends

We evaluated the power to detect changes in occupancy states across primary sampling occasions using two common survey designs: random-site and fixed-site. For the former, sites were randomly selected during

Table 1
Parameters, associated method used to simulate sample data in multistate occupancy evaluations, levels, and justification for choice of levels.

Parameters	Method	Levels	Justification
Occupancy, ψ^1	Bernoulli distribution	0.5	During exploratory analyses, we found that the occupancy rate ψ^1 had no effect on estimated bias or power to detect trends in abundance, thus we used a single fixed level.
Mean abundance, μ	NA	3, 6, and 12	Low, medium, and high levels of μ.
Coefficient of variation, CV	NA	————	————
Among-site (spatial)		0.2, 0.6, 1.2, and 1.8	Low, medium, high, and very high levels (spatial CV; Gibbs et al., 1998)
Within-site		0.0, 0.3, 0.6, 0.9, and 1.2	and none, low, medium, high, and very high levels (within-site CV).
Standard deviation, sd	μ*CV	NA	NA
Gamma distribution parameters	$\alpha = (\mu/sd)^2$, $\beta = (sd^2)/\mu$	NA	Gamma distribution allowed us to specify the level of heterogeneity.
Number of sites sampled, *i*	NA	25, 50, and 100	Low, medium, and high levels of *i*.
Number of visits per site, *j*	NA	2, 3, and 6	Low, medium, and high levels of *j*.
True conditional abundance, N_i	Gamma distribution rounded to nearest integer, (α, β)	NA	NA
Individual detection probability, p_i	Uniform distribution, (min., max.)	Min. = 0.1; max. = 0.3	Realistic range for many animals (e.g., Hammond and Anthony, 2006; Duarte et al., 2011; Price and Peterson, 2010; Couturier et al., 2013; Howell et al., 2016; Lituma et al., 2017), yet where N-mixture models produce biased estimates (Duarte et al., 2018).
Observed count, y_{ij}	Binomial distribution, (N_i, p_i)	NA	NA

each simulated year, whereas the latter included the random selection of sites during the first year and the same sites on subsequent years. For both designs, we set annual decreases in abundance of 5% and 10% for two- and ten-year simulations, although given the stochastic nature of the simulations, actual population declines varied (Fig. 1.3). We also simulated declines between two primary periods of 5%, 10%, 20%, and 40% as commonly in natural resource studies, surveys conducted at two points in time (whether separated by e.g., 5, 10, or 20 years or more) are all that is available. To simulate declines in annual abundance for the random-site design, we imposed declines on the overall mean true abundance from year to year. We then randomly sampled from the yearly overall mean to obtain values for each new randomly selected site. For the fixed-site design, we randomly generated true abundances for the first year at each site and imposed a proportional change in abundance for each subsequent year for each specific site. Because these decreases would result in fractions of animals (e.g., a 5% decline for an initial abundance of 10 is 9.5 animals), we randomly generated the true abundance in subsequent years using a normal distribution with a mean equal to the abundance at a site after decrease (e.g., 9.5 above example) and a CV of 0.1 rounded to the nearest whole number. This resulted in spatial variation in abundance that equaled that specified. We simulated 9,720 scenarios given varying spatial and within-site heterogeneity, mean true abundance, visits, sites, years, population decline, sample design, and whether we defined abundant based on the first year or all years' data.

2.2. Defining the abundant state

2.2.1. True states

We used quantiles of the true abundances where N>0 to identify simulated sites that were in the abundant state and defined the true abundant state as abundance greater than the 60th quantile (Fig. 1.1). For simulations with within-site heterogeneity in true abundance, we used the maximum true abundances across visits (e.g., five is the maximum count for a site with simulated counts for three visits of zero, one, and five) to calculate the threshold abundance value and classify sites into their true states.

2.2.2. Observed states

Multistate occupancy models assume that lower level states (occupied, not abundant) are not misclassified as upper-level states (detected in the abundant state). Therefore, we explored various thresholds to separate simulated count data into abundant and non-abundant states that minimized bias in the estimated proportion of abundant sites (Fig. 1.1). We evaluated several candidate thresholds using quantiles of the maximum non-zero counts at each site across multiple visits and tested thresholds of 0.70, 0.75, 0.8, 0.85, and 0.9. We defined abundant counts as those values greater than *or* equal to the threshold (note, this was different than how the threshold was applied in defining true states). We evaluated bias for 540 scenarios representing all combinations of abundance, spatial heterogeneity, sites, and visits simulated for 200 iterations (Table 1). We summarized bias for each scenario based on the mean values across the 200 replicates for each scenario. We found that a threshold quantile of 0.8 had the lowest median bias (See the Supporting Information). Given that the sensitivity of bias to the optimal threshold quantile between scenarios was low, we use this threshold for all subsequent steps in this paper (Fig. S1 in the Supporting Information). For users wanting to optimize their threshold for a given dataset, see *Data Diagnostics* sections in the Supporting Information.

2.3. Multistate occupancy model fitting

We fit single-season multistate occupancy models with program MARK (White and Burnham, 1999) in R (R Core Team, 2022) using RMark (version 2.2.7; Laake, 2013) to estimate the probability of sites being in the abundant state, and the power to detect a trend with the random-sites design (below). We used the conditional binomial formulation, which estimates parameters: ψ^1 the probability of occupancy regardless of abundance state, ψ^2 the probability of abundant given that the site is occupied, p^1 the probability of detecting occupancy given it is not in the abundant state, p^2 the probability of detecting occupancy given it is in the abundant state, and δ the probability of detecting the abundant state given that the species was detected and the true state was abundant. We used intercept-only models for all parameters and allowed p^1 and p^2 to be estimated separately. We calculated the probability of a site being in the abundant state as $\psi^1 \times \psi^2$, and the probability of being unoccupied as $1 - \psi^1$.

We restricted our trends evaluations to simulations that resulted in a true decline in simulated abundances. To test for a true decline, we used a generalized linear model with a Poisson distribution and modeled the maximum abundance (across visits) for the occupied sites (first year only) as a function of year. A true decline was any iteration that found a statistically significant decline in simulated true abundances ($\alpha = 0.05$, 0.1).

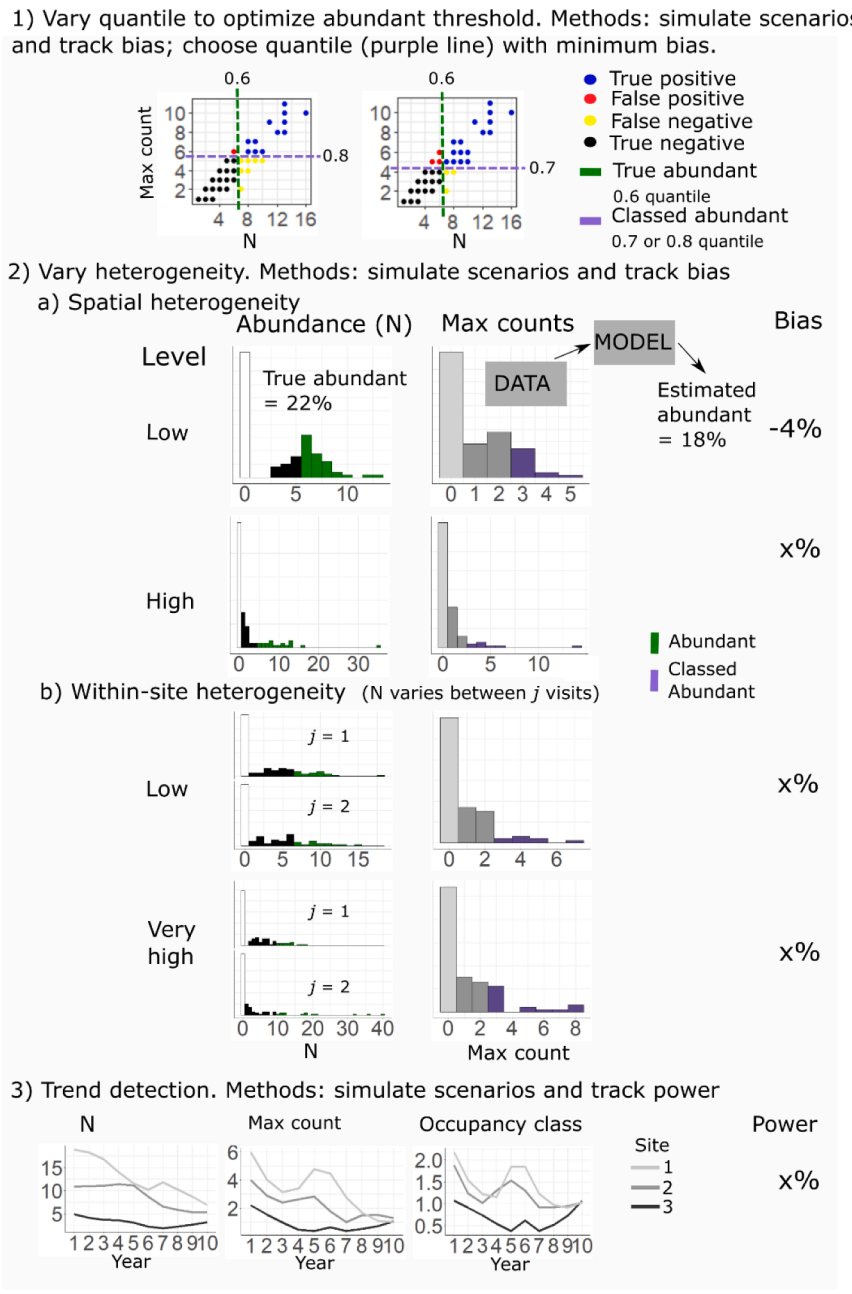


Fig. 1. Example simulated data for each of the three major components of the paper which are shown in the order they were conducted. We (1) optimized the threshold to class sites as in the relatively abundant state based on count data. The example shows non-zero latent (true, unobserved) site-level abundances (N) versus the maximum count across visits. Thresholding the latent abundance data by a 0.6 quantile defines sites in the true abundant state (right of line) and true non-abundant state (left of line). The two graphs show the tradeoff between different threshold quantiles creating *fewer* false positives but *higher* false negatives as the threshold is raised. These data were simulated using a mean latent abundance μ of 6, latent occupancy ψ^1 of 0.5, site-level detection probability p_i between 0.1 and 0.3, a medium amount of spatial heterogeneity, no within-site heterogeneity, 200 sites, and 3 visits. Bias is calculated as the estimated proportion of sites in the abundant state minus the true proportion of sites in the abundant state. We (2) explored the effects of varying spatial and within-site heterogeneity with single-season simulations. Based on the true and optimized classed abundant thresholds from (1), true and classed abundant sites are identified. Within-site heterogeneity causes true abundance to vary between visits. 'Estimated abundant' is estimated in a multistate occupancy model utilizing the multiple visit data. We (3) simulated multi-year data and looked at the power to detect a declining trend using multistate abundant occupancy models. Graphs show 3 sites from an example simulated with an initial (year 1) μ of 6, p_i 0.1 - 0.3, high spatial heterogeneity, low within-site heterogeneity, 10 years of data, and 10% annual declines. Given that true abundance varied between visits due to the within-site heterogeneity, $N = N_{i\text{-}max}$. Lines are plotted using a smoothing function to clarify the patterns. Power is calculated as the proportion of simulation replicates, for a given scenario, where a significant declining trend is detected. If this example had 25, 50, or 100 sites (only three are shown), it would represent one replicate for that scenario.

To evaluate our ability to estimate trends in abundance across primary occasions using a random-site design with single-season multistate occupancy models, we included a year covariate on ψ^2.

To evaluate our ability to estimate trends in abundance across primary occasions with a fixed-site design, we used dynamic multistate occupancy models in RMark to model transition probabilities from state 2 to state 1 (ψ^{BA}) and from state 1 to state 0 (ψ^{A0}). We fixed all other transition probabilities to zero, because we did not expect, for example, transitions to occur from absent to present given our simulations were focused on declining trends. For fixed-site and random-site simulations, we tracked power, the proportion of simulation replicates where the parameter estimate was negative (i.e., declining trend) and statistically significant ($\alpha = 0.05, 0.1$).

2.3.1. Convergence

For all iterations, we excluded results if parameter estimates did not converge. We ascribed this if parameter standard errors were zero or larger than 10 on the logit scale. Because p^2 is often large (near or at 1 on the real scale) due to the strong relationship among N, p, and the number of individuals in the count data (Bayley and Peterson, 2001; Royle and Nichols, 2003), we did not apply this rule to the standard error for p^2; standard errors are inestimable by their nature if the parameter is at a boundary. We tracked the proportion of simulations that converged for each scenario. We ran each scenario until we had 500 replicates that converged or for a minimum of 100 replicates where the time exceeded 30 minutes. We report scenarios where fewer than 100 replicates converged.

2.4. Inference on important relationships

For each set of simulations (defining the abundant state; heterogeneity; trend with random-site design and fixed-site design), we assessed which simulation parameters most affected bias or power (also, convergence; see the Supporting Information for details). We evaluated

the relative importance of parameters using ensemble regression trees with the *randomForest* package in R using the default settings (Liaw and Wiener, 2002). We tracked permutation importance – a metric that describes the relative explanatory ability of variables – to estimate variable importance. We report the average relativized importance across ten random forest models (Cutler et al., 2012). We discuss 'important' variables as those with a minimum relative importance of 0.10.

2.5. Data diagnostics

See the Supporting Information.

3. Results

3.1. Single-season; heterogeneity

Low levels of spatial heterogeneity resulted in unacceptable levels of bias, but otherwise the estimator performed well (Fig. 2A). Medium levels resulted in a small positive bias on average, while high or very high levels resulted in a small negative bias. This mix of positive and negative bias was due to defining a single threshold cutoff optimized across our varying levels of heterogeneity, whereas as a threshold could be optimized for specific levels of heterogeneity (see DATA DIAGNOSTICS in the Supporting Information). Increasing the number of visits or sites did not lower bias in a consistent fashion (Fig. 2B).

Within-site heterogeneity had little impact on the ability of multi-state occupancy models to estimate the proportion of sites in the abundant state. Mean bias with no within-site heterogeneity was -0.007 and with very high levels was 0.001. Across levels of spatial heterogeneity, increasing within-site heterogeneity from none to very high levels shifted bias to be more positive (Fig. 2C). Increasing the number of sites or visits did not lower bias in a consistent fashion (Fig. 2D).

3.2. Power to detect a trend

3.2.1. Random-site design

Power to detect a trend for the years comparison with the random-site design was low on average ($\bar{x} = 0.33$ at $\alpha = 0.10$; Fig. 3A). To detect a declining trend, large population decreases were needed; mean power across 2-year scenarios (with 5% and 10% decreases) was 0.10, while for 10-year simulations mean power was 0.41 for 5% annual decreases (40% total decline from year 1 to year 10) and 0.69 for 10% annual decreases (65% total decline). Besides the large impact of overall

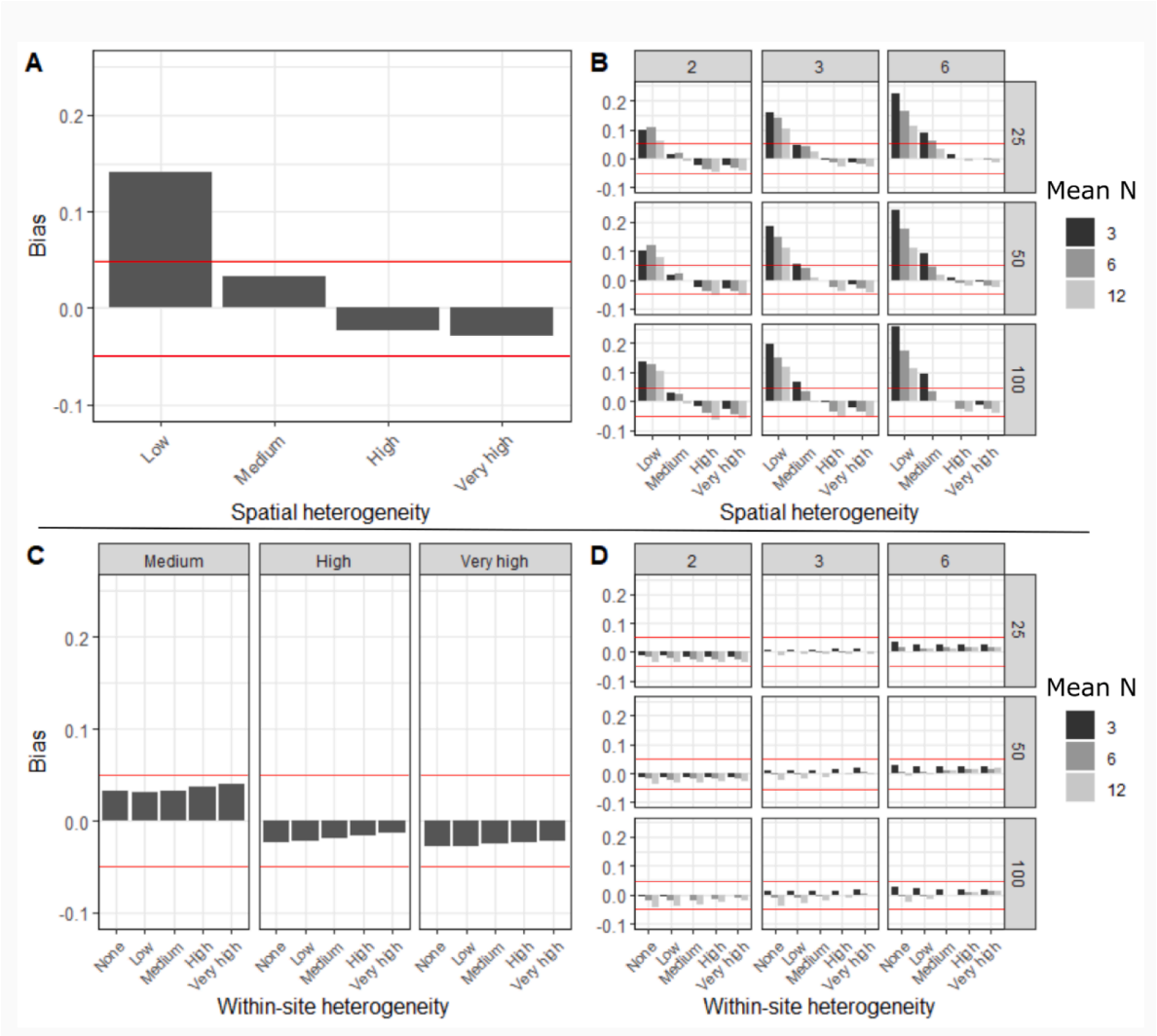


Fig. 2. Bias in estimates of the proportion of sites in the abundant state given variation in spatial heterogeneity (A); given visits (columns) and sites (rows, B); and within-site (between visit) heterogeneity at medium, high, and very high levels of spatial heterogeneity (C) in true abundance; and, given visits and sites (D). Red lines show limits of acceptable bias (+/- 0.05).

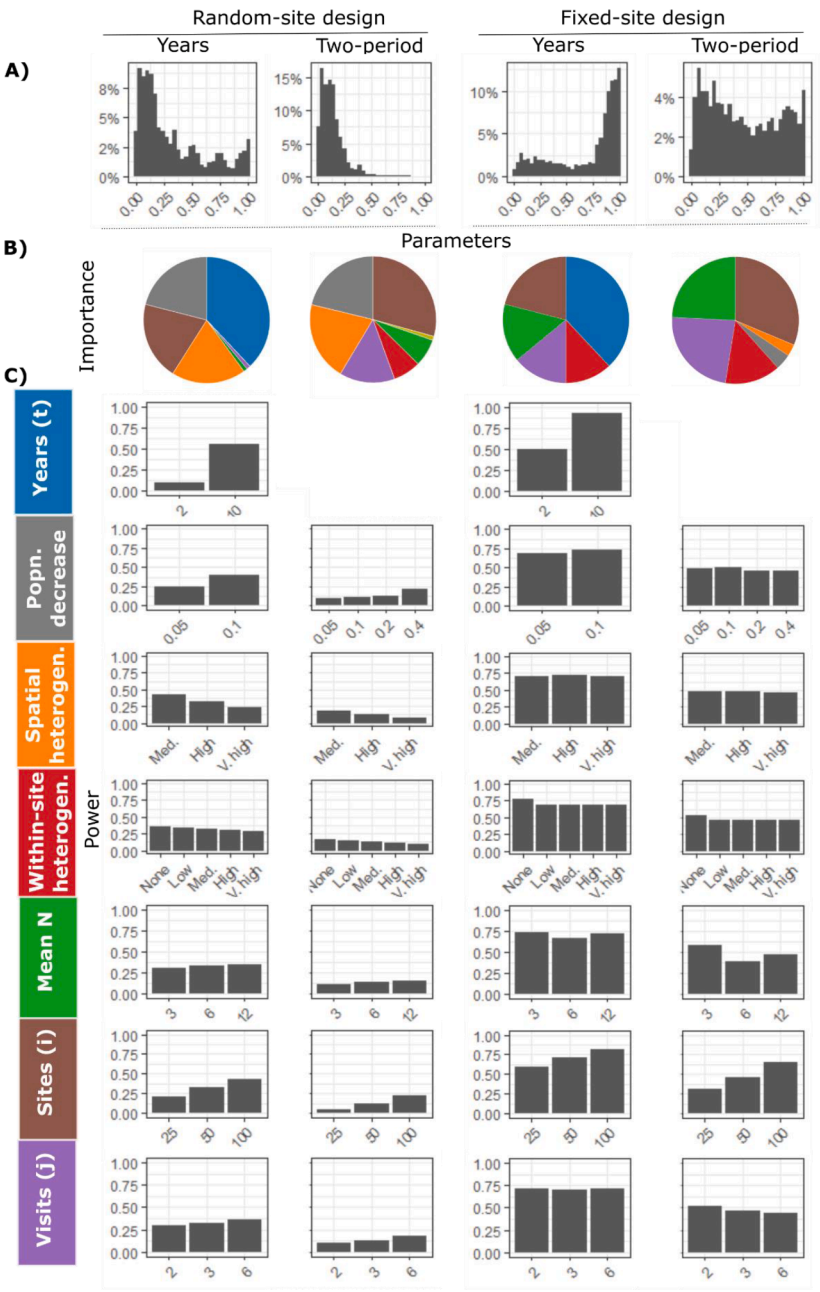


Fig. 3. Simulation results given random-site and fixed-site designs for scenarios that varied by the number of 'years' (two and ten) and by the population decrease (5, 10, 20, and 40%) between 'two-periods'. A) Distribution across scenarios in average power to detect a trend. B) Relative importance of model parameters in explaining the power to detect a trend (see Table S1 in the Supporting Information for the values used to create these charts). C) Mean power given variation in simulation parameters.

declines (number of years and population decrease), increasing the number of sites improved power the most followed by lower levels of spatial heterogeneity (Fig. 3B, 3C; Table S1 and Fig. S4 in the Supporting Information).

Power was low using the random-site design to detect a population decline as much as 40% between two time periods. Mean power for a decrease of 5% was 0.10, for 10% it was 0.10, for 20% it was 0.13, and for 40% it was 0.22 (at $\alpha = 0.10$; Fig. 3A,C). The number of sites influenced power the most, followed by the population decline, spatial heterogeneity, and the number of visits (Fig. 3B; Table S1). All effects were small and positive except spatial heterogeneity which had a small negative effect (Fig. 3C).

3.2.2. *Fixed-site design*

Across scenarios for 2- and 10-year periods with 5% and 10% declines when using a fixed-site design, mean power to detect a declining trend was on average higher for the non-abundant to unoccupied state transition (0.71 at $\alpha = 0.10$; 0.65 at $\alpha = 0.05$) than for the abundant to non-abundant state transition (0.64 at $\alpha = 0.10$; 0.62 at $\alpha = 0.05$). Thus, we focus on power to detect a trend in the non-abundant to unoccupied state transition (at $\alpha = 0.10$). Mean power was high for 10-year (0.92) and moderate for 2-year simulations (0.50; Fig. 3A). Besides the number of years, varying sites, mean true abundance, visits, and within-site heterogeneity impacted power the most (Fig. 3B; Table S1 and Fig. S5 in the Supporting Information). Power was highest when within-site heterogeneity was zero (Fig. 3C).

For two periods, again mean power was greater for the non-abundant to unoccupied state transition (0.48 at $\alpha = 0.10$; 0.39 at $\alpha = 0.05$) than the abundant to non-abundant state transition (0.29 at $\alpha = 0.10$; 0.26 at $\alpha = 0.05$). Focusing on the non-abundant to unoccupied transition, the amount of decline for a two-year period had little impact on power (Fig. 3B). The number of sites, followed by varying mean true abundance, visits, and within-site heterogeneity impacted power the most (Fig. 3B; Table S1 in the Supporting Information). Increasing sites

increased power while increasing visits lowered it (Fig. 3C). The minimum of 100 converged replicates was not achieved for 130 2-year scenarios (see Supporting Information DataS1 for a list of which scenarios did not converge).

3.3. Defining the abundant state

See the Supporting Information.

3.4. Convergence

See the Supporting Information.

3.5. Data diagnostics

See the Supporting Information.

4. Discussion

Conservation decisions are often based on differentiating areas of higher animal populations from lower, and with detecting trends; they rarely require information on total abundance (Nichols and Williams, 2006). Therefore, we evaluated the ability of multistate occupancy models that differentiate relatively low and high abundance states to detect spatial variation in abundance or a declining trend in the face of low individual detection probability and unexplained heterogeneity in true abundance between visits. We found, not surprisingly, that this approach would not be appropriate if sites differ little in their true abundance levels. We otherwise found that this approach could be used to differentiate sites with relatively high abundance from sites with relatively low abundance even in the face of high levels of individual movement, deaths, or recruitment among visits. Multistate occupancy models also had a good ability to detect a declining trend if the same sites were surveyed each year (i.e., the fixed site design), especially when more sites were surveyed.

The good performance of the estimator for differentiating relative abundance patterns among sites under high within-site heterogeneity is likely at least partly due to our definition of true abundance. We defined true abundance based on the maximum abundance across visits, describing the so-called 'super-population' (Schwarz and Arnason, 1996; Williams et al., 2011) which included the surrounding area from which animals moved into and out of the site. When we lack closure during abundance estimation, we are in fact always estimating super population abundance (Kendall, 1999). Classing a site in a relatively abundant (or non-abundant) state based on the maximum count, therefore, did not generate any additional bias between the true superpopulation abundant state and the estimate. However, if we had based true abundance on that from the first visit instead of the maximum across visits, we would have likely found that bias increased with increasing within-site heterogeneity.

While we tried to mitigate false positive classifications of the abundant state by choosing a reasonable optimal threshold for classing counts as abundant, this did not eliminate this set of misclassed observations (Fig. 1.1). This violates a key assumption of standard occupancy models. The effect of these false positives is to produce positively biased estimates; sampling more simply worsens this problem ("the more you sample, the worse you do", Kéry and Royle, 2020). This was apparent as we increased the number of visits to sites and – although the pattern is less clear – may also be why increasing the number of sites did not decrease bias in single season models (Fig. 2B–D). This problem was less or not at all apparent in our evaluation of the power to detect a trend where increasing sites positively impacted power, and increasing visits often did (Fig. 3).

When sites are randomly selected each year, spatial heterogeneity is not seen by a statistical model as explicitly spatial given variation is likely occurring within and among sites. Thus, any temporal signal needs to overcome the reduction in precision caused by the spatial heterogeneity to register as statistically significant (Larsen et al., 2001). In other words, a random-site study design can only detect declines that are large. We found this to be the case in our multistate occupancy model simulations where we only had good power to detect a decline in our 10-year simulations which was improved by lower spatial heterogeneity and higher annual population declines. In additional simulations, we found that a very large number of sites (N = 300) allowed declines of 40% to be detected (power = 0.8) across two periods with a medium level of spatial heterogeneity; if spatial heterogeneity was very high, many more sites were needed (N = 1,000; Fig. S7 in the Supporting Information).

Many studies report that large declines are needed in order to detect a statistically significant trend. For example, in one study, a 50% decline in occupancy could be detected for only 22% of a large, taxonomically diverse suite of species (136 species of birds, reptiles, and mammals; Southwell et al., 2019). Studies using density estimates or abundance indices reported needing large declines: 64% for tropical bats (Meyer et al., 2010; note, power cutoff was 0.9); greater than 40% for the Sonoran Desert Tortoise (Zylstra et al., 2010; power was 0.43 to detect a 40% decline); and greater than 50% for 70% of 127 marine mammal species (Taylor et al., 2007). With the relatively precise abundance estimates from mark-recapture or spatial mark-recapture methods (but, see Green et al., 2020), the minimum decrease detectable was 45% for a population of dolphins (Cantor et al., 2012) and ranged from 20-80% for Asian bears (Morin et al., 2022). Using a fixed-site design analyzed with dynamic multistate occupancy models, we could detect declines of 5 - 40% between two periods in 20% of the scenarios we looked at (power > 0.8); with 10 years of 5-10% annual declines we had relatively high power (≥ 0.75) across all scenarios (Supporting Information Fig.S5A). Power appeared unaffected by the actual amount of the population decline using a fixed-site design (Fig. 3), indicating that the increase in power from two to ten years was due to an increased ability to estimate transition probabilities with the repeated samples (i.e., a greater sample size), and not because declines themselves were large (i.e., 10% annual declines produce 65% declines after 10 years).

When animals are solitary and hard to find, occupancy-based trends analysis can be appropriate; our results indicate that when counts of animals vary between sites, multistate occupancy models can be sensitive to relatively small declines in abundance. While we only looked at two relative abundance states, more relative abundance states might increase the sensitivity of this model. We used a three-state multistate occupancy model as this parameterization allowed us to account for potential misclassifications in the observation data via the δ parameter (Nichols et al., 2007). Although more complicated parameterizations can be considered, more data would be needed to achieve model convergence (Fig. S2, S3, and *Results* in the Supporting Information). This is especially true when trying to estimate trends in abundance across years. In our trend simulations, convergence rates ranged from low to high (Fig. S3 in the Supporting Information) decreasing with increasing model complexity (lower for the more complex dynamic multistate occupancy model) and increasing with number of years and sites. This supports the notion that even in the simplest of cases a multistate dynamic occupancy model requires substantial data to be fitted. While we opted to fit our model using a frequentist statistical approach, better convergence could possibly be achieved using a Bayesian approach where information can be borrowed across parameters (Kéry and Schaub, 2011).

5. Conclusion

Based on our simulations, multistate occupancy models are able to model varying spatial abundance patterns and detect declining trends where N-mixture models fail. In particular, dynamic multistate occupancy models utilizing surveys conducted at the same sites from year to year (fixed-site design) were effective at detecting declining trends,

especially with more years and sites (Fig. S7 in the Supporting Information). By utilizing data diagnostic summaries (ICC, QCV-spatial, MMC0) along with study design elements (i.e., number of visits, sites, and years) users can evaluate, for a given dataset, whether bias in the estimated proportion of abundant sites will be low, and/or the likelihood they will be able to detect a declining trend should one exist (Fig. S6 in the Supporting Information).

CRediT authorship contribution statement

Valerie A. Steen: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. **Adam Duarte:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **James T. Peterson:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Valerie Steen reports financial support was provided by Metropolitan Water District of Southern California.

Data availability

All data used in this article were simulated. Simulation code is provided in the supplementary materials.

Acknowledgments

This project was funded by the Metropolitan Water District of Southern California, Grant No 187386. A. Duarte would also like to acknowledge support from the U.S.D.A. Forest Service, Pacific Northwest Research Station. We thank Corey Philips, Shawn Acuña, and Steve Whitlock for comments that improved this manuscript. The Oregon Cooperative Fish and Wildlife Research Unit is jointly sponsored by the U.S. Geological Survey, the U.S. Fish and Wildlife Service, the Oregon Department of Fish and Wildlife, Oregon State University, and the Wildlife Management Institute. The use of trade or firm names in this publication is for reader information and does not imply endorsement by the U.S. Government of any product or service

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2023.110303.

References

Alonso, R.S., McClintock, B.T., Lyren, L.M., Boydston, E.E., Crooks, K.R., 2015. Mark-recapture and mark-resight methods for estimating abundance with remote cameras: a carnivore case study. PloS One 10 (3), e0123032.
Amstrup, S.C., McDonald, T.L., Manly, B.F., 2005. Handbook of Capture-Recapture Analysis. Princeton University Press.
Barker, R.J., Schofield, M.R., Link, W.A., Sauer, J.R., 2018. On the reliability of N-mixture models for count data. Biometrics 74 (1), 369–377.
Bayley, P.B., Peterson, J.T., 2001. An approach to estimate probability of presence and richness of fish species. Trans. Am. Fish. Soc. 130 (4), 620–633.
Briscoe, N.J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A.K., Guillera-Arroita, G., 2021. Can dynamic occupancy models improve predictions of species' range dynamics? A test using swiss birds. Glob. Chang. Biol. 27 (18), 4269–4282.
Caci, G., Biscaccianti, A.B., Cistrone, L., Bosso, L., Garonna, A.P., Russo, D., 2013. Spotting the right spot: computer-aided individual identification of the threatened cerambycid beetle Rosalia alpina. J. Insect Conserv. 17 (4), 787–795.
Cantor, M., Wedekin, L.L., Daura-Jorge, F.G., Rossi-Santos, M.R., Simões-Lopes, P.C., 2012. Assessing population parameters and trends of Guiana dolphins (Sotalia guianensis): An eight-year mark-recapture study. Mar. Mammal Sci. 28 (1), 63–83.
Clawson, C.M., Falke, J.A., Bailey, L.L., Rose, J., Prakash, A., Martin, A.E., 2022. High-resolution remote sensing and multistate occupancy estimation identify drivers of spawning site selection in fall chum salmon (Oncorhynchus keta) across a sub-Arctic riverscape. Can. J. Fish. Aquat. Sci. 79 (3), 380–394.
Couturier, T., Cheylan, M., Bertolero, A., Astruc, G., Besnard, A., 2013. Estimating abundance and population trends when detection is low and highly variable: a comparison of three methods for the Hermann's tortoise. J. Wildl. Manag. 77 (3), 454–462.
Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random forests. Ensemble Machine Learning. Springer, Boston, MA, pp. 157–175.
Duarte, A., Brown, D.J., Forstner, M.R., 2011. Estimating abundance of the endangered Houston toad on a primary recovery site. J. Fish Wildl. Manag. 2 (2), 207–215.
Duarte, A., Adams, M.J., Peterson, J.T., 2018. Fitting N-mixture models to count data with unmodeled heterogeneity: bias, diagnostics, and alternative approaches. Ecol. Model. 374, 51–59.
Duarte, A., Peterson, J.T., Pearl, C.A., Rowe, J.C., McCreary, B., Galvan, S.K., Adams, M. J., 2020. Estimation of metademographic rates and landscape connectivity for a conservation-reliant anuran. Landsc. Ecol. 35 (6), 1459–1479.
Duarte, A., Peterson, J.T., 2021. Space-for-time is not necessarily a substitution when monitoring the distribution of pelagic fishes in the San Francisco Bay-Delta. Ecol. Evol. 11 (23), 16727–16744.
Ellis, M.M., Ivan, J.S., Schwartz, M.K., 2014. Spatially explicit power analyses for occupancy-based monitoring of wolverine in the US Rocky Mountains. Conserv. Biol. 28 (1), 52–62.
Elmore, S.A., Huyvaert, K.P., Bailey, L.L., Milhous, J., Alisauskas, R.T., Gajadhar, A.A., Jenkins, E.J., 2014. Toxoplasma gondii exposure in arctic-nesting geese: a multistate occupancy framework and comparison of serological assays. Int. J. Parasitol. Parasites Wildl. 3 (2), 147–153.
Fuller, A.K., Linden, D.W., Royle, J.A., 2016. Management decision making for fisher populations informed by occupancy modeling. J. Wildl. Manag. 80 (5), 794–802.
Hammond, E.L., Anthony, R.G., 2006. Mark–recapture estimates of population parameters for selected species of small mammals. J. Mammal. 87 (3), 618–627.
Gaston, K.J., Blackburn, T.M., Greenwood, J.J., Gregory, R.D., Quinn, R.M., Lawton, J. H., 2000. Abundance–occupancy relationships. J. Appl. Ecol. 37, 39–59.
Gibbs, J.P., Droege, S., Eagle, P., 1998. Monitoring populations of plants and animals. BioScience 48 (11), 935–940.
Green, A.M., Chynoweth, M.W., Şekercioğlu, Ç.H., 2020. Spatially explicit capture-recapture through camera trapping: a review of benchmark analyses for wildlife density estimation. Front. Ecol. Evol. 473.
Howell, P.E., Hossack, B.R., Muths, E., Sigafus, B.H., Chandler, R.B., 2016. Survival estimates for reintroduced populations of the Chiricahua leopard frog (Lithobates chiricahuensis). Copeia 104 (4), 824–830.
Jensen, T., Vokoun, J.C., 2013. Using multistate occupancy estimation to model habitat use in difficult-to-sample watersheds: bridle shiner in a low-gradient swampy stream. Can. J. Fish. Aquat. Sci. 70 (10), 1429–1437.
Kendall, W.L., 1999. Robustness of closed capture–recapture methods to violations of the closure assumption. Ecology 80 (8), 2517–2525.
Kéry, M., Dorazio, R.M., Soldaat, L., Van Strien, A., Zuiderwijk, A., Royle, J.A., 2009. Trend estimation in populations with imperfect detection. J. Appl. Ecol. 46 (6), 1163–1172.
Kéry, M., Schaub, M., 2011. Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective. Academic Press.
Kéry, M., Royle, J.A., 2020. Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS: Volume 2: Dynamic and Advanced Models. Academic Press.
Knape, J., Arlt, D., Barraquand, F., Berg, Å., Chevalier, M., Pärt, T., Żmihorski, M., 2018. Sensitivity of binomial N-mixture models to overdispersion: The importance of assessing model fit. Methods Ecol. Evol. 9 (10), 2102–2114.
Laake J.L. (2013). RMark: an R interface for analysis of capture-recapture data with MARK.
Larsen, D.P., Kincaid, T.M., Jacobs, S.E., Urquhart, N.S., 2001. Designs for evaluating local and regional scale trends: we describe a framework for evaluating the effects of spatial and temporal variation on the sensitivity of alternative ecological survey designs to detect regional temporal trends. BioScience 51 (12), 1069–1078.
Link, W.A., Schofield, M.R., Barker, R.J., Sauer, J.R., 2018. On the robustness of N-mixture models. Ecology 99 (7), 1547–1551.
Lituma, C.M., Buehler, D.A., Tanner, E.P., Tanner, A.M., Keyser, P.D., Harper, C.A., 2017. Factors affecting availability for detection: An example using radio-collared Northern Bobwhite (Colinus virginianus). PloS One 12 (12), e0190376.
MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83 (8), 2248–2255.
Mackenzie, D.I., 2006. Modeling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly. J. Wildl. Manag. 70 (2), 367–374.
MacKenzie, D.I., Nichols, J.D., Seamans, M.E., Gutiérrez, R.J., 2009. Modeling species occurrence dynamics with multiple states and imperfect detection. Ecology 90 (3), 823–835.
MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2017. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier.
Martin, J., Chamaillé-Jammes, S., Nichols, J.D., Fritz, H., Hines, J.E., Fonnesbeck, C.J., Bailey, L.L., 2010. Simultaneous modeling of habitat suitability, occupancy, and relative abundance: African elephants in Zimbabwe. Ecol. Appl. 20 (4), 1173–1182.
McGrady, M.J., Hines, J.E., Rollie, C.J., Smith, G.D., Morton, E.R., Moore, J.F., Oli, M.K., 2017. Territory occupancy and breeding success of Peregrine Falcons Falco peregrinus at various stages of population recovery. Ibis 159 (2), 285–296.
Meyer, C.F., Aguiar, L.M., Aguirre, L.F., Baumgarten, J., Clarke, F.M., Cosson, J.F., Kalko, E.K., 2010. Long-term monitoring of tropical bats for anthropogenic impact

assessment: gauging the statistical power to detect population change. Biol. Conserv. 143 (11), 2797–2807.
Morin, D.J., Boulanger, J., Bischof, R., Lee, D.C., Ngoprasert, D., Fuller, A.K., Karanth, U., 2022. Comparison of methods for estimating density and population trends for low-density Asian bears. Glob. Ecol. Conserv. 35, e02058.
Mosher, B.A., Bailey, L.L., Hubbard, B.A., Huyvaert, K.P., 2018. Inferential biases linked to unobservable states in complex occupancy models. Ecography 41 (1), 32–39.
Nichols, J.D., 1992. Capture-recapture models. BioScience 42 (2), 94–102.
Nichols, J.D., Williams, B.K., 2006. Monitoring for conservation. Trends Ecol. Evol. 21 (12), 668–673.
Nichols, J.D., Hines, J.E., Mackenzie, D.I., Seamans, M.E., Gutierrez, R.J., 2007. Occupancy estimation and modeling with multiple states and state uncertainty. Ecology 88 (6), 1395–1400.
Otis, D.L., Burnham, K.P., White, G.C., Anderson, D.R., 1978. Statistical inference from capture data on closed animal populations. Wildl. Monogr. (62), 3–135.
Peterson, J.T., Barajas, M.F., 2018. An Evaluation of three fish surveys in the San Francisco Estuary, 1995–2015. San Franc. Estuary Watershed Sci. 16 (4), 2.
Peterson, J.T., Shea, C.P., 2015. An evaluation of the relations between flow regime components, stream characteristics, species traits, and meta-demographic rates of warm-water-stream fishes: implications for aquatic resource management. River Res. Appl. 31 (10), 1227–1241.
Pollock, K.H., Nichols, J.D., Brownie, C., Hines, J.E., 1990. Statistical inference for capture-recapture experiments. Wildl. Monogr. 107, 3–97.
Price, A.L., Peterson, J.T., 2010. Estimation and modeling of electrofishing capture efficiency for fishes in wadeable warmwater streams. N. Am. J. Fish. Manag. 30 (2), 481–498.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.
Reddell, C.D., Abadi, F., Delaney, D.K., Cain, J.W., Roemer, G.W., 2021. Urbanization's influence on the distribution of mange in a carnivore revealed with multistate occupancy models. Oecologia 195 (1), 105–116.
Rodrigues, R.A., Massara, R.L., Bailey, L.L., Pichorim, M., Moreira, P.A., Braga, É.M., 2020. Using a multistate occupancy approach to determine molecular diagnostic accuracy and factors affecting avian haemosporidian infections. Sci. Rep. 10 (1), 1–10.
Royle, J.A., Nichols, J.D., 2003. Estimating abundance from repeated presence–absence data or point counts. Ecology 84 (3), 777–790.
Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60 (1), 108–115.
Royle, J.A., Link, W.A., 2005. A general class of multinomial mixture models for anuran calling survey data. Ecology 86 (9), 2505–2512.
Royle, J.A., Kéry, M., Gautier, R., Schmid, H., 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. Ecol. Monogr. 77 (3), 465–481.
Schwarz, C.J., Arnason, A.N., 1996. A general methodology for the analysis of capture-recapture experiments in open populations. Biometrics 860–873.
Southwell, D.M., Einoder, L.D., Lahoz-Monfort, J.J., Fisher, A., Gillespie, G.R., Wintle, B. A., 2019. Spatially explicit power analysis for detecting occupancy trends for multiple species. Ecol. Appl. 29 (6), e01950.
Stauffer, G.E., Roberts, N.M., Macfarland, D.M., Van Deelen, T.R., 2021. Scaling occupancy estimates up to abundance for wolves. J. Wildl. Manag. 85 (7), 1410–1422.
Taylor, B.L., Martinez, M., Gerrodette, T., Barlow, J., Hrovat, Y.N., 2007. Lessons from monitoring trends in abundance of marine mammals. Mar. Mammal Sci. 23 (1), 157–175.
Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., Possingham, H.P., 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. Ecol. Appl. 13 (6), 1790–1801.
White, G.C., Burnham, K.P., 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46, S120–S139 sup1.
Whitlock, S.L., Lewis, T.M., Peterson, J.T., 2020. Using a bayesian multistate occupancy model to assess seabird and shorebird status in Glacier Bay, Alaska. Wildl. Soc. Bull. 44 (3), 451–467.
Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and Management of Animal Populations. Academic Press.
Williams, K.A., Frederick, P.C., Nichols, J.D., 2011. Use of the superpopulation approach to estimate breeding population size: an example in asynchronously breeding birds. Ecology 92 (4), 821–828.
Zylstra, E.R., Steidl, R.J., Swann, D.E., 2010. Evaluating survey methods for monitoring a rare vertebrate, the Sonoran desert tortoise. J. Wildl. Manag. 74 (6), 1311–1313.