

# Integrating count and detection–nondetection data to model population dynamics

ELISE F. ZIPKIN,<sup>1,2,9</sup> SAM ROSSMAN,<sup>1,3</sup> CHARLES B. YACKULIC,<sup>4</sup> J. DAVID WIENS,<sup>5</sup>  
 JAMES T. THORSON,<sup>6</sup> RAYMOND J. DAVIS,<sup>7</sup> AND EVAN H. CAMPBELL GRANT<sup>8</sup>

<sup>1</sup>Department of Integrative Biology, Michigan State University, East Lansing, Michigan 48824 USA

<sup>2</sup>Ecology, Evolutionary Biology, and Behavior Program, Michigan State University, East Lansing, Michigan 48824 USA

<sup>3</sup>Hubbs-Sea World Research Institute, Melbourne Beach, Florida 32951 USA

<sup>4</sup>Southwest Biological Science Center, U.S. Geological Survey, Flagstaff, Arizona 86001 USA

<sup>5</sup>Forest and Rangeland Ecosystem Science Center, U.S. Geological Survey, Corvallis, Oregon 97331 USA

<sup>6</sup>Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington 98112 USA

<sup>7</sup>U.S. Forest Service, Pacific Northwest Region, 3200 SW Jefferson Way, Corvallis, Oregon 97331 USA

<sup>8</sup>USGS Patuxent Wildlife Research Center, SO Conte Anadromous Fish Research Center, 1 Migratory Way, Turners Falls, Washington 01376 USA

**Abstract.** There is increasing need for methods that integrate multiple data types into a single analytical framework as the spatial and temporal scale of ecological research expands. Current work on this topic primarily focuses on combining capture–recapture data from marked individuals with other data types into integrated population models. Yet, studies of species distributions and trends often rely on data from unmarked individuals across broad scales where local abundance and environmental variables may vary. We present a modeling framework for integrating detection–nondetection and count data into a single analysis to estimate population dynamics, abundance, and individual detection probabilities during sampling. Our dynamic population model assumes that site-specific abundance can change over time according to survival of individuals and gains through reproduction and immigration. The observation process for each data type is modeled by assuming that every individual present at a site has an equal probability of being detected during sampling processes. We examine our modeling approach through a series of simulations illustrating the relative value of count vs. detection–nondetection data under a variety of parameter values and survey configurations. We also provide an empirical example of the model by combining long-term detection–nondetection data (1995–2014) with newly collected count data (2015–2016) from a growing population of Barred Owl (*Strix varia*) in the Pacific Northwest to examine the factors influencing population abundance over time. Our model provides a foundation for incorporating unmarked data within a single framework, even in cases where sampling processes yield different detection probabilities. This approach will be useful for survey design and to researchers interested in incorporating historical or citizen science data into analyses focused on understanding how demographic rates drive population abundance.

**Key words:** *Dail-Madsen model*; *detection probability*; *integrated population model*; *N-mixture model*; *occupancy*; *unmarked data*.

## INTRODUCTION

As the focus in ecology and conservation biology shifts toward broader spatial extents (Allen and Hoekstra 2015), making use of data from multiple sources is increasingly necessary as no one data set can adequately characterize a species across its complete geographic range (Marra et al. 2015). This is particularly true when interest lies in assessing population-level consequences of changing demography relative to climate and/or landscape covariates (Robinson et al. 2014). One method for

combining data sources is through the use of integrated population models, which estimate population abundance and demographic rates through the joint analysis of two or more data sets within a single framework (Brooks et al. 2004). Compared to separate analyses of each data type, integrated population models provide inference on a greater number of parameters, increased precision, and more accurate accounting of uncertainty (Schaub and Abadi 2011). To date, integrated population models have focused on combining capture–recapture data with indices of abundance and other data types (e.g., telemetry, dead-recovery, fecundity surveys; Abadi et al. 2012, Wilson et al. 2016). Capture–recapture data are collected by following marked (naturally or with tags) individuals through time, which allows for explicit estimation of population vital rates, and is arguably the

Manuscript received 28 December 2016; revised 6 March 2017; accepted 9 March 2017. Corresponding Editor: Paul Conn.

<sup>9</sup>E-mail: ezipkin@msu.edu

most informative approach for tracking populations (Lebreton et al. 1992). Yet capture–recapture data are expensive and time-intensive to collect and necessarily limited in spatial extent because of practical difficulties. Furthermore, some species/taxa (e.g., invertebrates) do not readily allow for capture–recapture sampling techniques.

Recently developed approaches allow for the estimation of population abundance and demographic rates from “unmarked” data types in which individuals are not identified (Chandler and King 2011, Dail and Madsen 2011, Zipkin et al. 2014b, Rossman et al. 2016). These models, collectively referred to as dynamic N-mixture models, require repeated surveys (over a short time frame when the population is assumed to be closed) across spatial locations to account for detection errors during sampling. This set of surveys is then conducted over successive time periods to estimate annual or seasonal population abundance (e.g., robust design; Dail and Madsen 2011). Abundance changes through time by birth/death and immigration/emigration, which is described through processes of local survival and population gains (recruitment and immigration) within the N-mixture modeling framework. While unmarked data do not provide the same level of detail on demography as capture–recapture data (Zipkin et al. 2014a), they are cheaper and easier to obtain. Count data, along with other less intensive data types such as detection–nondetection data, are thus particularly valuable in projects with large spatial or temporal extents and in cases where it is difficult or impossible to track individuals.

Here we present an integrated modeling approach to combine unmarked data types. We analyze count and detection–nondetection time series data within a single model, assuming abundance changes according to biological processes describing survival and gains under the open-population dynamic N-mixture model (Dail and Madsen 2011, Dorazio 2014). Our approach thus allows for estimation of demographic rates (i.e., survival and recruitment) while explicitly accounting for detection errors during data collection. We show the utility of our modeling approach through a series of simulations illustrating the relative contributions of count vs. detection–nondetection data under a variety of parameter values and survey configurations. We also demonstrate how the model can be used with empirical data through an analysis of a Barred Owl (*Strix varia*) population in the Oregon Coast Ranges, USA. Including count and detection–nondetection data in a single model allows for more accurate and precise estimates of population abundance over time, even in cases where detection probabilities differ by survey type or data are collected at non-overlapping spatial locations or time periods. Our model provides a framework for combining many types of unmarked data into a single analysis and will be useful in investigating the optimal design of future surveys as well as providing capabilities to

incorporate historical or citizen science data with more rigorously collected scientific data.

## MODEL DESCRIPTION

### *Biological state process*

We incorporate count and detection–nondetection data into a single model by combining dynamic N-mixture (Royle 2004, Dail and Madsen 2011) and occupancy (MacKenzie et al. 2002) modeling frameworks. To do this, we model the latent demographic rates (i.e., the state process) by assuming that population abundance  $N_{j,t}$  (which is observed imperfectly) at a site  $j$  at time step  $t$  is conditional on abundance at  $j$  in the previous time step (Dail and Madsen 2011). We consider an annual cycle but the time step can be modified based on a species' dynamics. The change in  $N_{j,t}$  between  $t - 1$  and  $t$  is modeled by estimating the number of individuals that survive and remain at a site ( $S_{j,t}$ ) and those that are gained to a site  $j$  either by recruitment or immigration ( $G_{j,t}$ ). These quantities are expressed as follows:

$$S_{j,t} \sim \text{Bin}(N_{j,t-1}, \omega),$$

$$G_{j,t} \sim \text{Pois}(\gamma)$$

where  $\omega$  is the apparent annual survival probability of individuals and  $\gamma$  is the expected number of individuals that are gained to  $j$  between  $t - 1$  and  $t$ . Dail and Madsen (2011) present a density independent process of population gains; yet this assumption can be modified to include density dependent recruitment when data are available (Zipkin et al. 2014a, Bellier et al. 2016). The total population abundance at  $j$  in time  $t > 1$  is:

$$N_{j,t} = S_{j,t} + G_{j,t}.$$

The state process is initialized during the first year of sampling ( $t = 1$ ) by modeling abundance at each site,  $N_{j,1}$ , according to a Poisson distribution with an expected count of  $\lambda$ :

$$N_{j,1} \sim \text{Pois}(\lambda).$$

Initial abundance can also be modeled with more flexible distributions (e.g., negative binomial) if site-specific count data do not fit the Poisson assumption of equal mean and variance (Hostetler and Chandler 2015). Covariates and/or spatially correlated random effects can be added to any of the parameters ( $\omega$ ,  $\gamma$ ,  $\lambda$ ) using appropriate link functions to incorporate relevant factors that influence population dynamics across spatial locations or through time.

Many population analyses focus on evaluating the extinction risk and resiliency of local sites. The survival and gains parameters can be used to derive the colonization probability of unoccupied sites and the extinction

probability of occupied sites, quantities that are frequently estimated using dynamic occupancy models. The colonization probability of an unoccupied site,  $\phi_{j,t}$ , is the probability that at least one individual is gained to  $j$  in year  $t$ , i.e.,  $1 - P[G_{j,t} = 0]$ , which can be derived from the probability mass function of the Poisson distribution for the gains equation:

$$\phi_{j,t} = 1 - e^{-\gamma}.$$

The extinction probability,  $\varepsilon_{j,t}$ , of an occupied site is the probability that all individuals die between  $t - 1$  and  $t$  and that no new individuals immigrate to the site, i.e.,  $P[S_{j,t} = 0] \cap P[G_{j,t} = 0]$ :

$$\varepsilon_{j,t} = (1 - \omega)^{N_{j,t}} \cdot e^{-\gamma}.$$

As a result, extinction rates differ among sites dependent upon local abundance and any covariates on  $\omega$  or  $\gamma$ . More generally, metapopulation and colonization/extinction dynamics arise from local demography at the individual level (Ovaskainen and Hanski 2004), processes that cannot readily be accommodated in typical abundance or occupancy models that do not incorporate mechanism (e.g., survival and recruitment) explicitly.

#### Observation process

The demographic parameters and the true underlying abundance  $N_{j,t}$  cannot typically be observed directly. Instead, data are collected on  $N_{j,t}$  at each location  $j$  in each year  $t$  according to one of two sampling processes: (1) counts of individuals (Royle 2004) or (2) detection–nondetection of at least one individual of the species (MacKenzie et al. 2002, Royle and Nichols 2003). In both cases, sites are visited within each year on  $K > 1$  occasions over a timeframe during which the population is assumed to be closed (i.e., abundance is constant). We note that if detection probabilities are the same across all sites, then only a subset of sites need to be sampled repeatedly ( $K_j \geq 1$ ). Count data are collected by enumerating all individuals encountered during a fixed survey time interval while detection–nondetection (occupancy) data are collected by recording simply whether or not (at least one individual of) the species was detected. Count and detection–nondetection data can be collected through point counts, transect walks, or other techniques (MacKenzie et al. 2006, Royle and Dorazio 2008). The key feature of both data types is that the number of individuals counted,  $n_{j,t,k}$ , at a site  $j$  in year  $t$  during sampling replicate  $k$  or the observed occupancy status of a site,  $y_{j,t,k}$ , is subject to incomplete detection. That is, not every individual is detected when collecting count data, such that  $n_{j,t,k} \leq N_{j,t}$ . Similarly, a species that is present at a site  $j$  could incorrectly be recorded as absent if none of the  $N_{j,t}$  individuals are detected during sampling replicate  $k$ . It is therefore necessary to model the relationship of the data to the true

(unobservable) abundance or occupancy status. In the case of the count data,  $n_{j,t,k} = 0, 1, 2, \dots$ , we model the observation process as

$$n_{j,t,k} \sim \text{Bin}(N_{j,t}, p)$$

where  $p$  is the detection probability of each individual at each survey event  $k = 1, 2, \dots, K$  (Royle 2004). In the case of the detection–nondetection data, we assume that the probability of recording a detection ( $y_{j,t,k} = 1$ ) is the probability that at least one individual is observed. Thus the detection process for the detection–nondetection data can be expressed as (Royle and Nichols 2003, Rossman et al. 2016):

$$y_{j,t,k} \sim \text{Bern}(1 - (1 - p)^{N_{j,t}}).$$

In this basic model, the detection probability of individuals ( $p$ ) is assumed to be equal across the count and detection–nondetection data (although we relax this assumption during simulations below). When detection is equal across survey types, repeated sampling of sites with detection–nondetection data is not necessary because  $p$  can be inferred from the count data alone. Thus, our modeling framework could be particularly useful for utilizing historical detection–nondetection data or data collected in remote locations where repeated sampling is challenging or nonexistent (assuming that data are collected under a randomized sampling design, absence data are recorded, and detection probabilities are the same for both survey types [Dorazio 2014, Pagel et al. 2014]). As with the demographic parameters, detection probability can be indexed by site or year to include relevant covariates that account for variation in the sampling process across time or space. An implicit assumption of our modeling framework, and integrated analyses generally, is that the spatial unit of sites is similar across survey types (but we demonstrate an example of how to reconcile data from different spatial units in the application section). Additionally, both detection probability and the demographic parameters are assumed to be equal for all individuals during each survey event (i.e., no individual heterogeneity) and detection of every individual is independent (Royle and Nichols 2003, Royle 2004). This assumption could be modified to account for differences in parameter values across life stages (or other subgroups of the population) if data are available (Zipkin et al. 2014b).

#### SIMULATION STUDY DESIGN

We developed a series of simulations to assess the utility of our combined count and occupancy model to estimate demographic rates and population abundance and to determine optimal sampling schemes in cases where parameter values vary across sampling locations (e.g., according to covariates) and/or sampling methodology (e.g., differences in detection). We examined the accuracy

and precision of our model over a range of parameter values and sampling protocols using at least 1,000 simulated data sets for all scenarios and number of surveyed sites. For each analysis and parameter combination, we generated 10 years of latent population abundances at individual sites (using parameter values specific to each scenario), during which we assumed abundance changed according to the process described in the model description section. Each of the sites was then “surveyed” three times annually, assuming independence and closure within intra-annual sampling events, according to either a count- or occupancy-based protocol (number of sites with a particular sampling protocol varied among simulation scenarios). We then analyzed the simulated data with the joint model using a Bayesian analysis with Markov chain Monte Carlo in the programs R and JAGS (Plummer 2003). We specified vague priors for all parameters ( $\omega$ ,  $\gamma$ ,  $\lambda$ ,  $p$  as well as any additional parameters defined within a specific simulation). Model code and implementation details for the simulation studies are provided in Data S1.

#### *Accuracy and precision of the basic model*

We determined the accuracy and precision of the basic model under a wide range of parameter values and across a realistic range of possible count/occupancy site combinations. To that end, we generated data sets by randomly selecting parameter values from the following distributions:  $\lambda \sim U(0.5, 3)$ ,  $\omega \sim U(0, 1)$ ,  $\gamma \sim U(0, 2.5)$ ,  $p \sim U(0, 1)$ . These distributions cover the complete parameter space for survival and detection probabilities and represent conditions for which site abundance and occupancy is likely to vary among sites. For example, we set an upper bound of 2.5 individuals for  $\gamma$  (expected number of recruits/immigrants gained annually per site) because site-specific population abundance becomes very high, leading to no unoccupied sites, when the expected number of individuals gained to sites is large. In such situations, collection of detection–nondetection data would be uninformative. Parameters were drawn independently to guarantee ample coverage across the specified parameter space. We examined the benefits of combining either 0, 25, 75, or 150 sites with detection–nondetection data to either 5, 15, or 30 sites with count data. For each count/occupancy site combination, we generated 5,000 data sets to ensure that a sufficiently wide range of possible parameter combinations was included in the results.

#### *Determining optimal sampling schemes*

Combining count and detection–nondetection data will be particularly useful in cases where it is difficult to obtain sufficient data across a covariate space using a single sampling protocol. For this simulation, we assume that a covariate influences the survival probability of individuals across spatial locations as follows:

$$\text{logit}(\omega_j) = \beta_0 + \beta_1 \text{covariate}_j.$$

For simplicity, we assume that  $\beta_1$  is positive and thus survival increases as the value of the covariate increases. We envision a scenario where either detection–nondetection or count data could be added to existing data to improve precision in parameter estimates and examined four such cases: (1) both count and detection–nondetection data are available over the complete range of the covariate ( $-3 < \text{covariate}_j < 3$ ); (2) count data are collected at sampling locations over a range where survival, and thus abundance, is high ( $1 < \text{covariate}_j < 3$ ) and detection–nondetection data are collected at locations where survival, and thus abundance, is low ( $-3 < \text{covariate}_j < 1$ ); (3) count data are collected at locations with average survival probabilities ( $-1 < \text{covariate}_j < 1$ ) and detection–nondetection data are collected at locations where survival is either high ( $\text{covariate}_j > 1$ ) or low ( $\text{covariate}_j < -1$ ); and (4) both count and detection–nondetection data are available over a subset of the range of the covariate, where survival is high ( $1 < \text{covariate}_j < 3$ ). We generated 1,000 data sets for each of these scenarios using the following parameter values:  $\lambda = 4$ ,  $\beta_0 = 0.5$ ,  $\beta_1 = 0.7$ ,  $\gamma = 2$ ,  $p = 0.5$  and assumed a fixed number of 40 sites with count data and 100 with detection–nondetection data.

#### *Combining data sources when detection probabilities differ*

We have so far considered scenarios where detection probabilities are equal for individuals across both count- and occupancy-based sampling schemes. The degree to which this assumption is reasonable depends on individual survey protocols. For example, the detection probability of individuals may differ by surveys because of the duration of the collection process, the area surveyed, or the manner in which individuals are detected (e.g., auditory vs. visual surveys). The best approach for dealing with differences in detection is to include relevant covariates (MacKenzie et al. 2006, Royle and Dorazio 2008). However, in some situations the baseline detection for individuals may be different enough that each survey type requires independent estimation of the detection probability. We explored both scenarios where this could be true: (1) detection probability of individuals is higher in the count data ( $p_{\text{count}} = 0.5$ ) than in the detection–nondetection data ( $p_{\text{occ}} = 0.3$ ) and (2) detection probability is lower in the count data ( $p_{\text{count}} = 0.3$ ) than in the detection–nondetection data ( $p_{\text{occ}} = 0.5$ ). We also examined the situation in which detection probability differs between sampling protocols but is incorrectly modeled assuming that they are equivalent (i.e.,  $p_{\text{count}} = p_{\text{occ}}$ ). We generated 1,000 data sets for each of these scenarios across a range of count (5, 15, 30) and occupancy (25, 75, 150) site combinations using the following demographic parameter values:  $\lambda = 1$ ,  $\omega = 0.7$ ,  $\gamma = 1.5$ .



## SIMULATION STUDY RESULTS

Simulation results indicate that our model combining unmarked data types can provide accurate estimates of demographic rates, population abundance, and individual detection probabilities across the comprehensive range of parameter values that we examined (Figs. 1–3). Precision in parameter estimates varied by the amount of data included in analyses and not surprisingly, increased with additional data (Fig. 1; Appendix S1, which shows results from the basic simulation with only five years of data).

Count data undoubtedly inform parameter values more efficiently than detection–nondetection data (i.e., Fig. 1, comparison across panel colors). However, the addition of a small number of occupancy sites (e.g., 25) to existing count data improved precision of parameters and abundance estimates in all scenarios, especially when the amount of available count data was relatively low (Fig. 1, left-side panels; Appendix S1).

Combining count and detection–nondetection data was especially useful in simulations with a covariate on survival, particularly if a single data type was not

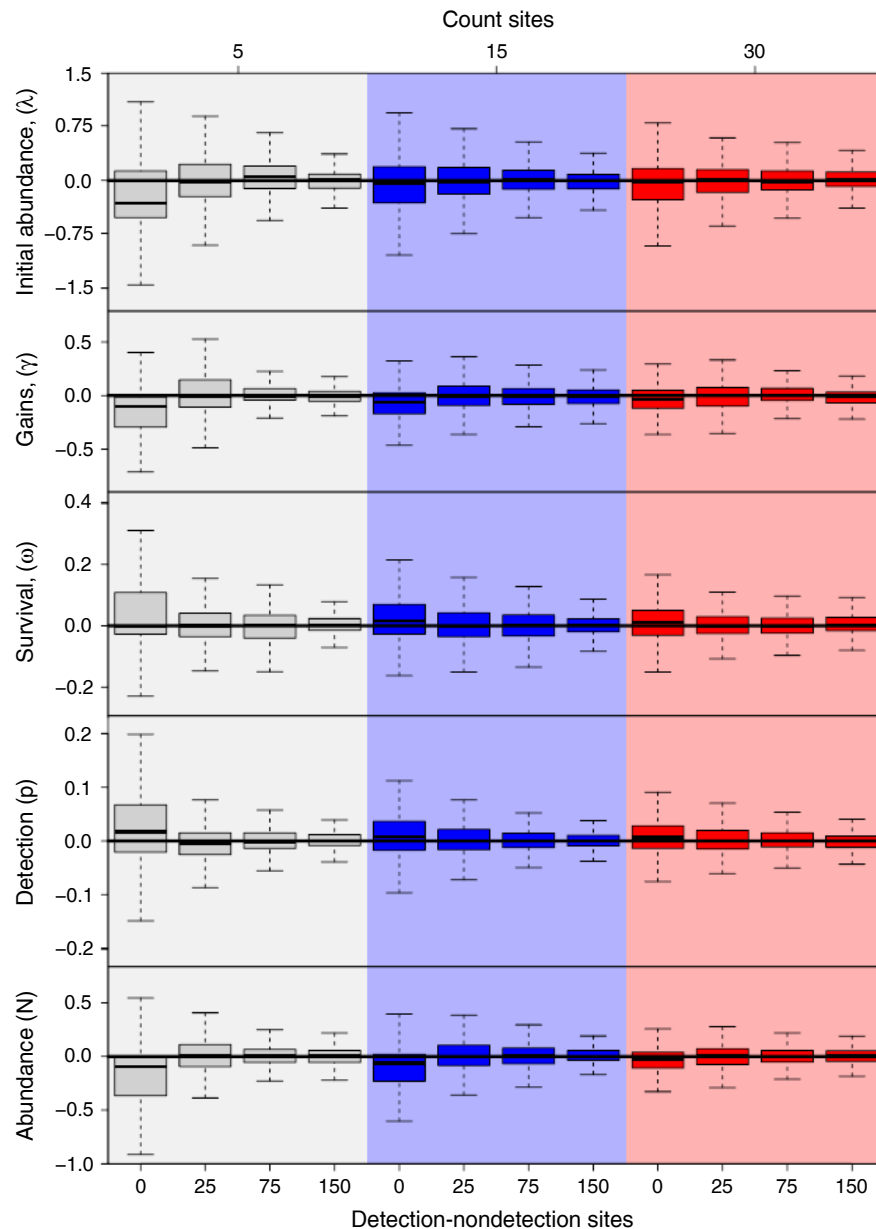


FIG. 1. Boxplots summarizing the accuracy and precision of analyses with simulated data under an array of sites surveyed using count- and occupancy-based protocols. The x-axis indicates the number of detection–nondetection sites for each simulation and the colored panels indicate the number of count sites. Each panel shows the median (thick line within boxes), 50% quantiles (boxes), and  $\pm 1.5$  times the interquartile range (whiskers) for the median estimated value minus the true value of parameters (top four panels) and abundance (bottom panels) for 5,000 simulated data sets with random combinations of the true parameter values. Parameter estimates equal the true values where the y-axis equals zero (black lines).

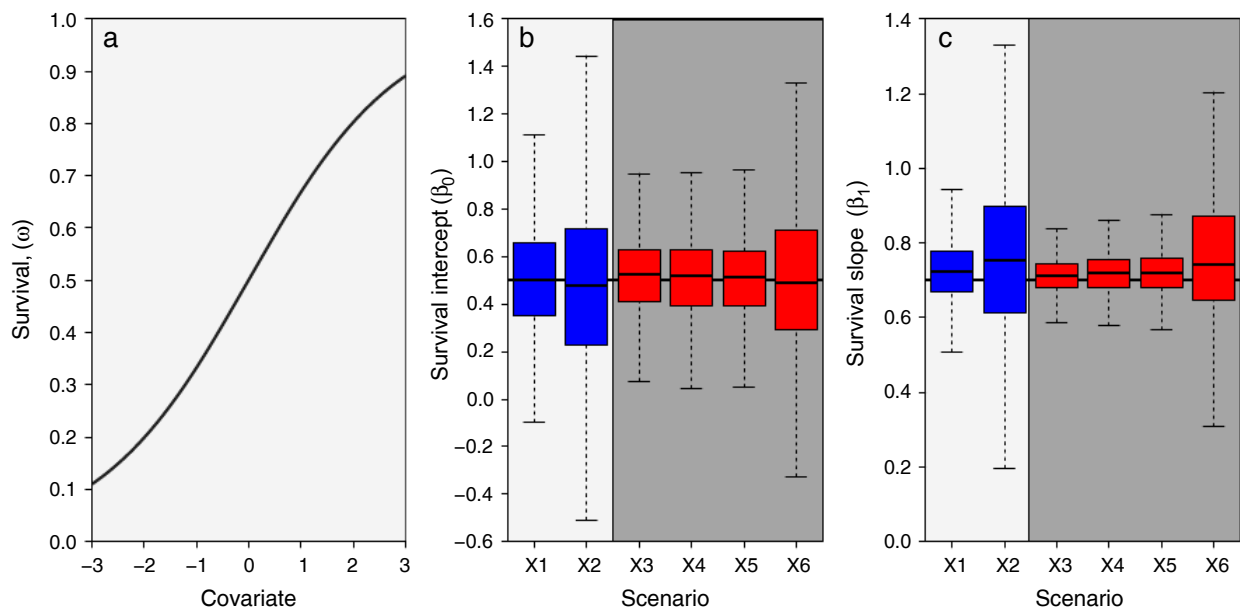


FIG. 2. Estimates of a covariate effect on survival under a number of sampling protocols. (a) The relationship between the covariate and survival. The other two panels show the estimated (b) intercept ( $\beta_0 = 0.5$ ) and (c) slope ( $\beta_1 = 0.7$ ) under six scenarios: count data only (blue boxes), available across the whole range of the covariate (X1) and only where survival is high (X2); a combination of count and detection–nondetection data (red boxes) available across the range of the covariate (X3), from count data where survival is high and detection–nondetection data where survival is low (X4), from count data where survival is average and detection–nondetection data where survival is low or high (X5), and where both count and detection–nondetection data are only available where survival is high (X6). Boxplots show median parameter estimates (thick line within boxes), 50% quantiles (boxes), and  $\pm 1.5$  times the interquartile range (whiskers) for 1,000 simulated data sets. True parameter values are shown with a thick black line.

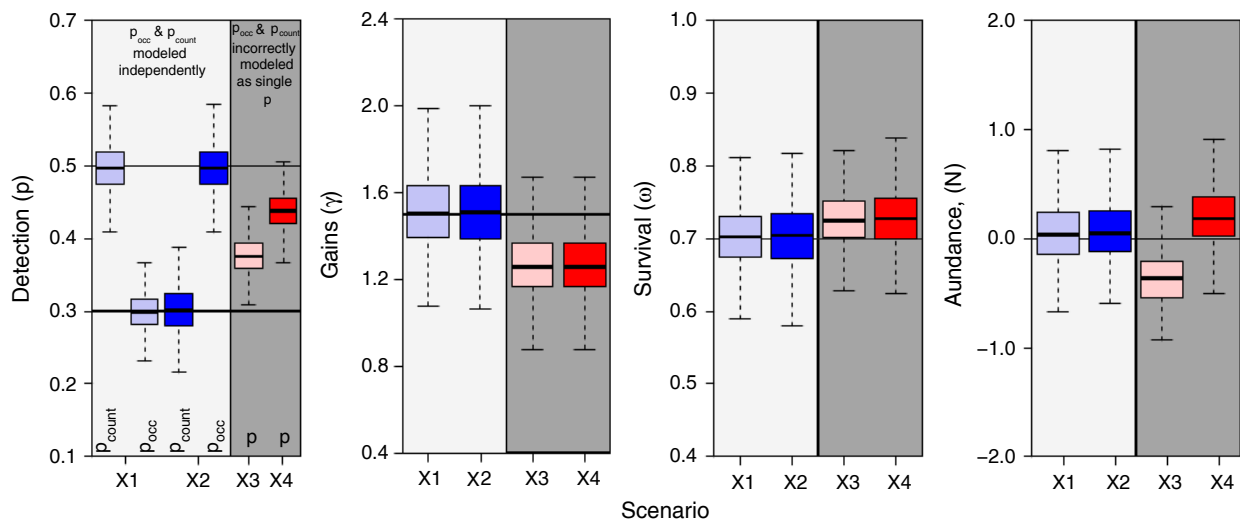


FIG. 3. Accuracy and precision of parameter values under four scenarios for 15 count and 75 detection–nondetection sites. The first two (blue) assume data are modeled according to the data generating process where individual detection probability is higher in the count than in the detection–nondetection data (X1) and detection is higher in the detection–nondetection than in the count data (X2). Scenarios X3 and X4 (red) model data generated in X1 and X2 using the standard model, which assumes that detection probability is equal across both sampling protocols. Black lines show the true values of the data generation process. Boxplots show the median (dark lines), 50% quantiles (boxes), and  $\pm 1.5$  times the interquartile range (whiskers) for 1,000 simulations.

available throughout the complete range of the covariate (Fig. 2; Appendix S2). Accurate and precise estimation of a covariate effect depends on whether the available data span the complete range of the covariate value and not on the data type, whether generated from count or

occupancy protocols (Fig. 2, blue boxes compared to red boxes). These results demonstrate that the inclusion of detection–nondetection data in addition to count data (or count data in addition to detection–nondetection data) allows for estimates of demographic rates and

abundance in locations with only detection–nondetection data while simultaneously improving precision on estimates of the covariate effect in areas with count data.

Our model produces accurate estimates of demographic rates and abundance even in cases where detection varies by the data collection method (Fig. 3, blue boxes in light gray panels; Appendix S3). This is true regardless of whether individual detection probability is higher with either count- or occupancy-based protocols. The precision of parameter estimates, however, depends on the amount of available data; increasing the number of parameters requires more data for comparable precision (Appendix S3 compared to Fig. 1). For this parameter combination, population gains are underestimated while survival probabilities are overestimated when we incorrectly assume that detection is equal across sampling methods when in fact it is different (Fig. 3, red boxes in dark gray panels; Appendix S3). Our simulation results indicate that abundance is overestimated (Fig. 3, dark red boxes) when detection in the detection–nondetection data is higher than that in the count data and underestimated in the reverse situation (Fig. 3, light red boxes), likely due to the inclusion of more occupancy than count sites. Thus, when detection is underestimated at the majority of sites, abundance is naturally overestimated (with the reverse also being true; Royle and Nichols 2003). The degree to which parameter biases, caused by mis-specifying the detection process, are significant will depend on the magnitude of the differences in detection probabilities among sampling protocols and the relative amount of sites surveyed for each data type. Additional simulations across a wider parameter space would allow for a more nuanced understanding of the consequences of mis-specifying the detection process.

#### APPLICATION TO EMPIRICAL DATA

We applied our modeling framework to survey data collected on an expanding population of Owls in a 1,692-km<sup>2</sup> region in the central Oregon Coast Ranges, over a period of two decades. Barred Owls were historically limited to eastern North American forests, but their range has expanded into the Pacific Northwest over the last century with local densities increasing dramatically over the last decade (Yackulic et al. 2012, Dugger et al. 2016). There is considerable interest in understanding the population dynamics of Barred Owls because of their potential negative impact on threatened Northern Spotted Owls (*Strix occidentalis caurina*) and other native wildlife (Wiens et al. 2014, Yackulic et al. 2014, Holm et al. 2016). Detection–nondetection data on Barred Owls were collected incidentally within Spotted Owl surveys from 1995 to 2014 (Lint et al. 1999). Spotted Owl surveys followed a standardized protocol (Lint et al. 1999) and were focused on 106 historical breeding territories (e.g., sites), which averaged 9.9 km<sup>2</sup> in size (Fig. 4a). During annual surveys of Spotted Owls,

observers visited each site up to eight times during the breeding season (March–August) and additionally recorded whether territorial Barred Owls (individuals or pairs) were detected.

A new count-based survey protocol, targeting Barred Owls, was initiated in 2015 as part of a broader study to improve estimation of Barred Owl abundances and examine the effects of experimental removals on the population demography of Northern Spotted Owls (Wiens et al. 2011, Diller et al. 2016). The experiment included locations where Barred Owls were either removed (treatments, about a third of the study area) or not (controls), but for the purposes of this study we restricted estimates to pre-treatment (2015) survey data collected on both areas, and post-treatment survey data on the control area only (i.e., to avoid confounding effects in our analysis of Barred Owl removals in treatment areas). The Barred Owl surveys employed a standard design in which a grid of 5-km<sup>2</sup> hexagons were overlaid to include historical breeding territories of Spotted Owls (Fig. 4a). Each of these hexagonal sites were surveyed up to three times during the breeding season. During each survey, observers used an amplified megaphone (Wildlife Technologies, Manchester, New Hampshire, USA) to broadcast digitally recorded Barred Owl calls at established call points that provided complete coverage of the site. All territorial pairs and single owls were recorded. Barred Owl individuals were assumed to be part of a territorial pair when (1) both sexes were observed within 400 m of each other on the same visits or (2) at least one adult was observed with young (Wiens et al. 2011).

While our simulation study focused on instances in which detection–nondetection and count data come from spatially distinct sites, our modeling framework can also be used in cases where the two data types are collected in the same locations in different time periods. Sites can be alternatively sampled using either occupancy- or count-based protocols as long as they are independent and the basic assumptions outlined in the model description section are met. In the case of this study, we needed to standardize the data from the two survey methods in which sites overlapped, but where Barred Owls were sampled at different spatial scales (Fig. 4a) in order to combine the historical detection–nondetection data with the newer count data. To do this, we reassigned each of the counts of territorial pairs detected within the 5-km<sup>2</sup> hexagonal sites (collected during Barred Owl-specific surveys in 2015 and 2016) to the larger Spotted Owl survey sites (i.e., historical territories) using the GPS coordinates of each pair observation. As a result, the 106 sites used for our analysis were defined according to the historical survey design, from which most of the data originate, and the finer-resolution count data were reconfigured to fit within that framework. Within our model, we allowed detection probabilities to differ between sampling schemes because of the very different spatial scales and protocols used for the surveys. In the case of the occupancy data, we assumed that detection of individuals ( $p_{occ}$ ) was constant

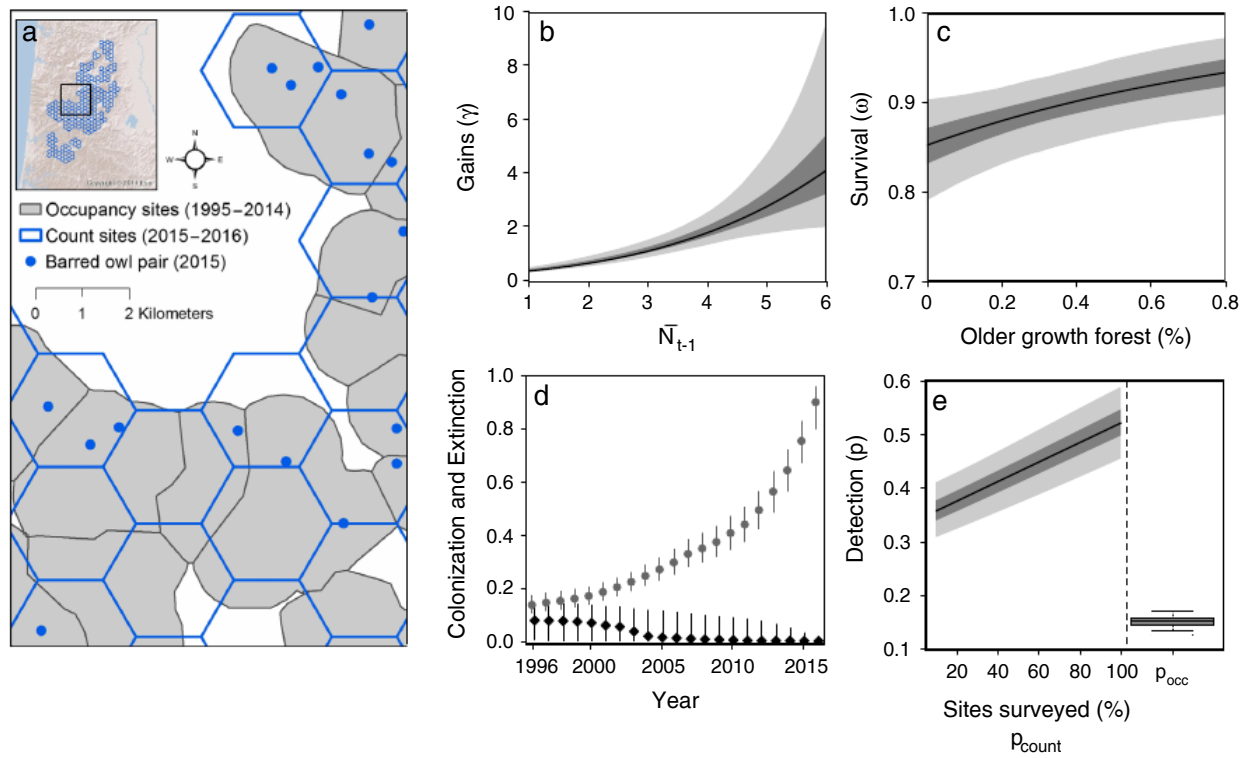


FIG. 4. Study area and results from the Barred Owl application. (a) Map of the study area in the central Oregon Coast Ranges, USA. The gray areas with black outlines depict breeding territories of Northern Spotted Owls (i.e., detection–nondetection sites) where Barred Owls were detected incidentally during surveys of Spotted Owls from 1995 to 2014. Blue hexagons (i.e., count sites) indicate where Barred Owl-specific count surveys were completed in 2015 and 2016. Blue dots demonstrate the GPS locations of Barred Owl counts that we used in reconciling detections of territorial pairs between the different spatial scales of the survey sites. (b) Expected site-specific gains,  $\gamma$ , relative to average regional abundance in the previous year. (c) Apparent annual survival,  $\omega$ , relative to the amount of older growth forest cover within sites. (d) Mean annual colonization ( $\phi$ , gray circles) and extinction ( $\epsilon$ , black diamonds) probabilities over the study period shown with 95% CI. (e) Detection probabilities for the count (left panel) and detection–nondetection (right panel) data. In panels b, c, and e, black lines indicate mean values, plotted with 50% CI (dark gray region) and 95% CI (light gray region). In panel e, the boxplot for  $p_{occ}$  shows the mean (black lines in box), 50% CI (box), and 95% CI (whiskers).

across sites and years as data were all collected by trained observers in the early morning. However, we specified the detection process for the count data ( $p_{count}$ ) using the proportion of the total area of the historical site  $j$  that was surveyed during replicate  $k$  in year  $t$ ,  $area_{j,k,t}$ , as an offset with the cloglog link function:

$$\text{cloglog}(p_{count,j,k,t}) = \alpha_0 + \log(area_{j,k,t}).$$

Thus, if a given count-based sampling event only covered half the area of the larger occupancy site, we recorded 0.5 for the offset on detection. Similarly, if a hexagon count site overlapped more than one historical occupancy site, only the proportion of that hexagon that overlapped the focal occupancy site was used. The cloglog link function is designed for encounter–nonencounter data given a Poisson intensity function, which arises in our model due to a Poisson recruitment process and a Bernoulli survival process. It has the useful property that, given low area swept, a doubling of area-swept results in a doubling of encounter probability, and was consequently a natural choice for our analysis.

We assumed that the Barred Owl population was closed to changes within years but that local site-level abundance could change annually through survival and gains. We included a covariate on the annual apparent survival probability of individuals based on area of older (approximately  $\geq 80$  yr) coniferous forest patches (Davis et al. 2015) within each site (using the same approach as in the simulation study). This covariate was calculated annually and, due to low levels of recent older forest disturbance and the slow rate of forest succession within the study's time frame, was fairly constant at most sites. Finally, recent evidence suggests that site-level gains in abundance may be dependent on the total regional population size as Barred Owls are exceptionally good at colonizing new sites (Yackulic et al. 2012, 2014). As such, we included a covariate on the gains parameter,  $\gamma$ , to account for a potential effect of regional population size:

$$\log(\gamma_t) = \delta_0 + \delta_1 \cdot \bar{N}_{t-1} + \delta_2 \cdot N_{t-1}^2$$

where  $\bar{N}_{t-1}$  is the average abundance of all sites in year  $t - 1$ , which we normalized by subtracting 1 (a value that



was close to the average site-specific abundance over the two decades of the study). We standardized all of the covariate data (e.g., forest cover) and analyzed the model using the programs R and JAGS, assuming uninformative prior distributions for each of the parameters (see Appendix S4 for model code and implementation details).

Model results show that the Barred Owl population grew substantially over the course of the survey period from a mean site-specific value of 0.13 (95% CI [0.06, 0.48]) territorial owls (individuals and pairs) in 1995 to 7.5 (95% CI [4.26, 11.53]) in 2016 (see Appendix S4 for a complete list of parameter estimates). This increase can be largely attributed to a positive density dependent effect on population gains,  $\gamma$  (Fig. 4b). We estimated a significant positive effect of mean regional abundance on the expected number of territorial owls gained to sites annually (mean  $\delta_1$ , 0.59; 95% CI [0.41, 0.78]) that did not decline when abundance was high (mean  $\delta_2$ , -0.02; 95% CI [-0.06, 0.02]; Fig. 4b), suggesting that the population has not yet saturated the study region. Annual survival probabilities were quite high (average range: 0.86–0.93) and increased with the amount of older coniferous forest cover available within a site (Fig. 4c). The intercepts for the  $\gamma$  and  $\omega$  parameters were negatively correlated (-0.55), although this is not unexpected as survival and gains are the only processes by which abundance can change within the model structure. Estimates of annual survival, and relationships with forest conditions, were strikingly similar to those derived from more intensive (and costly) studies of radio-marked individuals conducted in the region (Wiens et al. 2014). We used the parameter estimates and our derived equations to calculate annual colonization and extinction probabilities (Fig. 4d). Colonization, or the probability that an unoccupied site becomes occupied, increased steadily over the time frame of the survey from a low of 0.14 (95% CI [0.10, 0.17]) in 1996 to a high of 0.90 (95% CI [0.81, 0.96]) in 2016. Site extinction probabilities were fairly low throughout the two decade period, averaging 0.07 in 1996 (95% CI [0.00, 0.14]) and declining to practically zero by 2016. Not surprisingly, Barred Owl detection probabilities were much higher during the count surveys as compared to the detection–nondetection surveys and increased with the area sampled (Fig. 4e).

## DISCUSSION

Estimating demographic rates, population abundance, and trends is a universal objective in ecology and is necessary to inform population management. Capture–recapture data of marked individuals is the gold standard because such data allow for detailed demographic analyses. However, many pressing questions related to population dynamics are difficult to answer using capture–recapture data, particularly in the case of invasions that are ongoing or have already occurred, and because capture–recapture data tend to be spatially limited. Successive surveys of spatially replicated counts and

occurrences can provide similar, although less detailed, information on population abundance, demographic rates, and/or colonization and extinction dynamics (MacKenzie et al. 2003, Royle 2004, Dail and Madsen 2011). Combining count and detection–nondetection data into a single integrated model can lead to a more accurate understanding of population demography and changes over time than is possible with independent analyses (Fig. 1).

Integrated population models have typically focused on approaches to augment capture–recapture data with other data types (Schaub and Abadi 2011; However, we show how combining only unmarked data types can provide increased accuracy and precision in estimates of population abundance and spatially varying demographic rates, even in cases where the sampling process leads to different detection probabilities among data types. As with other integrated analyses, this is because the different data are assumed to derive from the same underlying biological processes (Dorazio 2014, Pagel et al. 2014). As a result, combining the data in a single model leads to a more efficient analysis. In some cases, such as in our Barred Owl example, researchers may switch from collecting one unmarked data type to another (e.g., from detection–nondetection to count) within a specific study area. Our modeling approach provides a framework to include the entire time series of data in a single analysis, regardless of this type of change. Zipkin et al. (2014b) found that the length of the time series of data had a greater contribution to parameter precision than the number of sites surveyed in a stage-structured N-mixture model. We anticipate a similar result for the combined detection–nondetection/count model based on estimates from our simulation study (Fig. 1; Appendix S1): longer time series seem to lead to disproportionate parameter precision for a fixed number of total sampling events. Our results further suggest that a site with count data is approximately equivalent to three sites with detection–nondetection data in a model with no covariates; yet the exact information trade-off is dependent on variation in site-level abundance and detection probabilities and will naturally be case specific.

Studies of species distributions, abundances, and dynamics over broad spatial extents often rely on either detection–nondetection data or counts of unmarked individuals. The potential to combine count and detection–nondetection data into a unified analysis lays the foundation for a number of analysis possibilities, particularly in terms of survey design. For example, monitoring invasive species typically involves detection–nondetection surveys combined with detailed count surveys at sites that are known to be occupied. In many such cases, it will not be feasible to conduct counts at every location where the species is encountered; simulations can help determine the optimal placement of count sites relative to detection–nondetection surveys. In general, researchers may want to target count-based protocols at locations with high quality habitat (i.e., with covariates in which survival and/or gains are expected to be high) and save less intensive detection–nondetection

protocols for locations in marginal habitats. Such survey methodologies could provide high quality inferences as long as sites span the complete range of covariate space (Fig. 2). We envision that future work could include presence-only data in combination with other unmarked protocols (Dorazio 2014, Pagel et al. 2014). This may be particularly useful for monitoring emerging species, where reports of detections (e.g., of the salamander chytrid fungus *Batrachochytrium salamandrivorans*) could then trigger cluster count samples in nearby areas. Although presence-only data are often associated with the analysis of historical and archival data sets, they may also arise in citizen-science data sets or other survey protocols.

Population closure is not a reasonable assumption for some sampling protocols and integrating such data may involve adding alternative observation models including those that allow for false positives, double counting, or species misidentification (Miller et al. 2011, Thorson et al. 2014, Chambert et al. 2016). Our results suggest that these efforts can provide accurate parameter estimates if the detection process is modeled correctly, but may still provide useful, if somewhat biased, estimates otherwise (Fig. 3). Parameter identifiability and/or accuracy can be a problem in analyses that estimate demographic rates from unmarked data (Zipkin et al. 2014a, Bellier et al. 2016). Although we did not have this issue in our application of the model (Appendix S4), analyses using comparatively sparser data sets may have difficulties with convergence or identifiability. The incorporation of auxiliary information can increase the accuracy and precision of parameter estimates through the use of informative priors (Morris et al. 2015) or by explicitly integrating available demographic data into the modeling framework. This may be particularly advantageous in cases where model assumptions are not strictly met (Bellier et al. 2016). We anticipate a growing importance for studies that combine data from multiple sampling protocols and thus encourage additional research regarding optimal data collection and analysis methods on integrated model structures.

#### ACKNOWLEDGMENTS

We thank the many field technicians who collected the Barred Owl data that we used in the empirical application of our model. We also thank the associate editor, Paul Conn, and anonymous reviewers for the many useful suggestions that improved the quality of our manuscript. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This is contribution number 577 of the Amphibian Research and Monitoring Initiative (ARMI) of the U.S. Geological Survey.

#### LITERATURE CITED

- Abadi, F., O. Gimenez, H. Jakober, W. Stauber, R. Arlettaz, and M. Schaub. 2012. Estimating the strength of density dependence in the presence of observation errors using integrated population models. *Ecological Modelling* 242:1–9.
- Allen, T. F., and T. W. Hoekstra. 2015. *Toward a unified ecology*. Columbia University Press, New York, New York, USA.
- Bellier, E., M. Kéry, and M. Schaub. 2016. Simulation-based assessment of dynamic N-mixture models with density-dependence and environmental stochasticity in vital rates. *Methods in Ecology and Evolution* 7:1029–1040.
- Brooks, S. P., R. King, and B. J. T. Morgan. 2004. A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation* 27: 515–529.
- Chambert, T., B. R. Hossack, L. Fishback, and J. M. Davenport. 2016. Estimating abundance in the presence of species uncertainty. *Methods in Ecology and Evolution* 7:1041–1049.
- Chandler, R. B., and D. I. King. 2011. Habitat quality and habitat selection of golden-winged warblers in Costa Rica: an application of hierarchical models for open populations. *Journal of Applied Ecology* 48:1038–1047.
- Dail, D., and L. Madsen. 2011. Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* 67:577–587.
- Davis, R. J., J. L. Ohmann, R. E. Kennedy, W. B. Cohen, M. J. Gregory, Z. Yang, H. M. Roberts, A. N. Gray, and T. A. Spies. 2015. Northwest forest plan—the first 20 years (1994–2013): status and trends of late-successional and old-growth forests. General Technical Report PNW-GTR-911. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.
- Diller, L. V., K. A. Hamm, D. A. Early, D. W. Lamphear, K. M. Dugger, C. B. Yackulic, C. J. Schwarz, P. C. Carlson, and T. L. McDonald. 2016. Demographic response of northern spotted owls to barred owl removal. *Journal of Wildlife Management* 80:691–707.
- Dorazio, R. M. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography* 23:1472–1484.
- Dugger, K. M., et al. 2016. The effects of habitat, climate, and Barred Owls on long-term demography of Northern Spotted Owls. *Condor* 118:57–116.
- Holm, S. R., B. R. Noon, J. D. Wiens, and W. J. Ripple. 2016. Potential trophic cascades triggered by the barred owl range expansion. *Wildlife Society Bulletin* 999:1–10.
- Hostetler, J. A., and R. B. Chandler. 2015. Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology* 96:1713–1723.
- Lebreton, J. D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* 62:67–118.
- Lint, J., B. Noon, R. Anthony, E. Forsman, M. Raphael, M. Collopy, and E. Starkey. 1999. Northern spotted owl effectiveness monitoring plan for the Northwest Forest Plan. General Technical Report PNW-GTR-440. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, Oregon, USA.
- MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* 84:2200–2207.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83:2248–2255.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollack, L. L. Bailey, and J. E. Hines. 2006. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press, Oxford, UK.

- Marra, P. P., E. B. Cohen, S. R. Loss, J. E. Rutter, and C. M. Tonra. 2015. A call for full annual cycle research in animal ecology. *Biology Letters* 11:20150552.
- Miller, D. A., J. D. Nichols, B. T. McClintock, E. H. C. Grant, and C. M. Tonra. 2015. A call for full annual cycle research in animal ecology. *Biology Letters* 11:20150552.
- Morris, W. K., P. A. Vesk, M. A. McCarthy, S. Bunyavechewin, and P. J. Baker. 2015. The neglected tool in the Bayesian ecologist's shed: a case study testing informative priors' effect on model accuracy. *Ecology and Evolution* 5:2045–7758.
- Ovaskainen, O., and I. Hanski. 2004. From individual behavior to metapopulation dynamics: unifying the patchy population and classic metapopulation models. *American Naturalist* 164:364–377.
- Pagel, J., B. J. Anderson, R. B. O'Hara, W. Cramer, R. Fox, F. Jeltsch, D. B. Roy, C. D. Thomas and F. M. Schurr. 2014. Quantifying rangewide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution* 5:751–760.
- Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the Third International Workshop on Distributed Statistical Computing*. R Project for Statistical Computing, Vienna, Austria.
- Robinson, R. A., C. A. Morrison, and S. R. Baillie. 2014. Integrating demographic data: towards a framework for monitoring wildlife populations at large spatial scales. *Methods in Ecology and Evolution* 5:1361–1372.
- Rossman, S., C. Yackulic, S. Saunders, J. Reid, and E. F. Zipkin. 2016. Dynamic *N*-occupancy models: estimating demographic rates and local abundance from detection-non-detection data. *Ecology* 97:3300–3307.
- Royle, J. A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Royle, J. A., and R. M. Dorazio. 2008. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press, Oxford, UK.
- Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* 84:777–790.
- Schaub, M., and F. Abadi. 2011. Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology* 152:227–237.
- Thorson, J. T., M. D. Scheuerell, B. X. Semmens, and C. V. Pattengill-Semmens. 2014. Demographic modeling of citizen science data informs habitat preferences and population dynamics of recovering fishes. *Ecology* 95:3251–3258.
- Wiens, J. D., R. G. Anthony, and E. D. Forsman. 2011. Barred owl occupancy surveys within the range of the northern spotted owl. *Journal of Wildlife Management* 75:531–538.
- Wiens, J. D., R. G. Anthony, and E. D. Forsman. 2014. Competitive interactions and resource partitioning between northern spotted owls and barred owls in western Oregon. *Wildlife Monographs* 185:1–51.
- Wilson, S., K. C. Gil-Weir, R. G. Clark, G. J. Robertson, and M. T. Bidwell. 2016. Integrated population modeling to assess demographic variation and contributions to population growth for endangered whooping cranes. *Biological Conservation* 197:1–7.
- Yackulic, C. B., J. Reid, R. Davis, J. E. Hines, J. D. Nichols, and E. Forsman. 2012. Neighborhood and habitat effects on vital rates: expansion of the Barred Owl in the Oregon Coast Ranges. *Ecology* 93:1953–1966.
- Yackulic, C. B., J. Reid, J. D. Nichols, J. E. Hines, R. Davis, and E. Forsman. 2014. The roles of competition and habitat in the dynamics of populations and species distributions. *Ecology* 95:265–279.
- Zipkin, E. F., T. S. Sillett, E. H. C. Grant, R. B. Chandler, and J. A. Royle. 2014a. Inferences about population dynamics from count data using multistate models: a comparison to capture-recapture approaches. *Ecology and Evolution* 4:417–426.
- Zipkin, E. F., J. T. Thorson, K. See, H. J. Lynch, E. H. C. Grant, Y. Kanno, R. B. Chandler, B. H. Letcher, and J. A. Royle. 2014b. Modeling structured population dynamics using data from unmarked individuals. *Ecology* 95:22–29.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.1831/supinfo>