The importance of simulation assumptions when evaluating detectability in population models

ADRIAN P. MONROE,1,† GREGORY T. WANN,2 CAMERON L. ALDRIDGE,1 AND PETER S. COATES2

1Natural Resource Ecology Laboratory and Department of Ecosystem Science and Sustainability, Colorado State University, in cooperation with the U.S. Geological Survey, Fort Collins Science Center, Fort Collins, Colorado 80526 USA
2U.S. Geological Survey, Western Ecological Research Center, Dixon Field Station, Dixon, California 95620 USA

Citation: Monroe, A. P., G. T. Wann, C. L. Aldridge, and P. S. Coates. 2019. The importance of simulation assumptions when evaluating detectability in population models. Ecosphere 10(7):e02791. 10.1002/ecs2.2791

Abstract. Population monitoring is important for investigating a variety of ecological questions, and N-mixture models are increasingly used to model population size (N) and trends (λ) while estimating detectability (p) from repeated counts within primary periods (when populations are closed to changes). Extending these models to dynamic processes with serial dependence across primary periods may relax the closure assumption, but simulations to evaluate models and inform effort (e.g., number of repeated counts) typically assume p is constant or random across sites and years. Thus, it is unknown how these models perform under scenarios where trends in p confound inferences on N and λ, and conclusions regarding effort may be overoptimistic. Here, we used global positioning system data from greater sage-grouse (Centrocercus urophasianus) to inform simulations of the detection process for lek counts of this species, and we created scenarios with and without linear annual trends in p. We then compared estimates of N and λ from hierarchical population models either fit with single maximum counts or with detectability estimated from repeated counts (dynamic N-mixture models). We also explored using auxiliary data to correct counts for variation in detectability. Uncorrected count models consistently underestimated N by >50%, whereas N-mixture models without auxiliary data underestimated N to a lesser degree due to unmodeled heterogeneity in p such as age. Nevertheless, estimates of λ from both types of models were unbiased and similar for scenarios without trends in p. When p declined systematically across years, uncorrected count models underestimated λ, whereas N-mixture models estimated λ with little bias when all sites were counted repeatedly. Auxiliary data also reduced bias in parameter estimates. Evaluating population models using scenarios with systematic variation in p may better reveal potential biases and inform effort than simulations that assume p is constant or random. Dynamic N-mixture models can distinguish between trends in p and N, but also require repeated counts within primary periods for accurate estimates. Auxiliary data may be useful when researchers lack repeated counts, wish to monitor more sites less intensively, or require unbiased estimates of N.

Key words: bias; binomial mixture model; Centrocercus urophasianus; detection probability; greater sage-grouse; population trends.

INTRODUCTION

Broad-scale, long-term monitoring is invaluable for basic and applied research, including characterizing population trends and evaluating ecological theories (Nichols 1991, Pollock et al. 2002, Lindenmayer et al. 2012). Many monitoring programs conduct counts of individuals,
where expected count $E(c)$ is the product of detectability ($p$) and the population size of interest ($N$): $E(c) = pN$. As such, counts may yield useful indices of trends in $N$ if detectability is constant (or varies independently of $N$; Johnson 2008, Nichols et al. 2009). However, $p$ can vary over space and time due to factors such as observer differences and vegetation structure, and this potential for error has long invited criticisms of indices (Nichols 1992, Anderson 2001, 2003, Ellingson and Lukacs 2003, White 2005). Specifically, trends in indices may be misleading when trends in $p$ also occur (Link and Sauer 1998, Kéry 2008, Yamaura 2013, Veech et al. 2016), which could confound inferences, frustrate conservation planning, and waste limited resources. Because many monitoring programs involve counts, N-mixture models (Royle 2004) offer the potential to analyze these data while accounting for variation in detectability (Kéry et al. 2005, 2009, McCaffery et al. 2016). Using a robust design where repeated counts are collected from a population that is closed within years or seasons, the observed count ($c_{ijt}$) at site $i$ on survey $j$ in year $t$ can be modeled as the outcome of a binomial distribution, given population size $N_{it}$ and detection probability $p_{ijt}$: $c_{ijt} \sim \text{binomial}(N_{it}, p_{ijt})$. Importantly for modeling population trends, this approach uses information from repeated counts to distinguish annual changes in $N$ from systematic changes in $p$, whereas these processes are confounded in analyses of indices (Kéry 2008, Yamaura 2013).

Simulation studies are useful for evaluating models such as N-mixture, including the level of effort (e.g., repeated counts) needed to distinguish trends in $N$ from $p$ (Kéry and Royle 2016). Such studies are necessary because monitoring data may lack repeated counts from sites for historical reasons or because limited resources for monitoring force a trade-off between temporal replication and spatial representation. If monitoring data occur across a series of years, one could analyze single counts collected each primary period using dynamic N-mixture models (Dail and Madsen 2011), assuming abundance in a given year is a function of population size and change from the previous year. This serial dependence across years may alleviate the need for repeated within-season counts during closed secondary periods to distinguish between trends in $N$ and $p$. However, simulations that evaluate dynamic N-mixture models typically assume that $p$ is constant or varies randomly and independently of $N$ across sites and years (Dail and Madsen 2011, Zipkin et al. 2014, Hostetler and Chandler 2015, Bellier et al. 2016). Under such conditions, detectability is unlikely to confound inferences regarding the latent population trend (Johnson 2008, Nichols et al. 2009), and therefore, it is unknown whether dynamic N-mixture models can distinguish trends in $N$ from $p$, particularly with few or no repeated counts. Furthermore, without considering systematic changes in detectability, inferences may not differ between N-mixture models and models of indices, and researchers could conclude equivalence (e.g., Barker et al. 2018).

To illustrate these implications, we use the greater sage-grouse (Centrocercus urophasianus, hereafter sage-grouse), a species of conservation concern whose populations are monitored across western North America. Each spring, male sage-grouse gather at breeding display sites (leks) to court females. As with many bird species, detectability of sage-grouse during lek counts can be thought of as the product of four distinct components (Table 1; Nichols et al. 2009). Two components relate to the distribution and movement of individuals, including probability that a bird’s home range overlaps the sampling unit ($p_w$ in the case of sage-grouse, attending a lek at least once in a season) and probability of presence at the sampling unit during the count ($p_p$). The other two components characterize the detection process for birds given their presence at the sampling unit: probability of being available for

| Detectability component | Parameter | Examples from lek counts |
|-------------------------|-----------|-------------------------|
| Sampling                | $p_s$     | Individual attending a lek at least once in a season (annual) |
| Presence                | $p_p$     | Individual present at lek during a lek count, within a season |
| Availability            | $p_a$     | Individual visible, given presence at lek during a count |
| Detection               | $p_d$     | Observer able to perceive and count individual, given presence and availability |

Table 1. Detectability components (sensu Nichols et al. 2009) and relevance to lek counts of greater sage-grouse (Centrocercus urophasianus).
detection during the count \((p_d, \text{ e.g., not concealed by vegetation or topography})\) and probability of detection conditional on availability \((p_a; \text{ observer correctly perceives and counts an individual})\). A major source of variation in detectability for sage-grouse occurs because lek attendance \((p_L)\) varies with date and time (Beck and Braun 1980, Emmons and Braun 1984, Walsh et al. 2004, Fremgen et al. 2019, Wann et al. 2019). As a result, trends in sage-grouse populations are typically estimated using the maximum observed number of males from repeated counts each year at each lek (hereafter, peak counts; Connelly et al. 2003). Additionally, detectability of grouse may be influenced by visual obstruction from vegetation, snow, and topography (Fremgen et al. 2016) as well as visibility changes due to weather and time of day (Baumgardt et al. 2017). Nevertheless, factors that influence \(p\) or \(N\) independently are estimable even with indices (Yamamura 2013, Banks-Leite et al. 2014, Barker et al. 2018). More concerning in our view are factors affecting detectability components that also vary with population size (temporally and spatially). For example, trends in vegetation cover following disturbance such as wildfire could affect both detectability (Blomberg et al. 2013, Fremgen et al. 2016) and population size (Blomberg et al. 2012). Changes in observer skill also may coincide with population trends (Link and Sauer 1998). Previously, simulations indicated that N-mixture models recovered population trends despite reduced effort or missing lek counts (McCaffery et al. 2016), but it is unknown how the model performs under scenarios with systematic changes in \(p\).

Using sage-grouse lek counts as a model study system also offers opportunities to identify model limitations and potential solutions. For example, aggregating individual encounter histories into counts results in information loss because \(p\) is assumed constant among individuals (Kéry and Royle 2016, Barker et al. 2018). Variation in detectability among individuals is well recognized (Otis et al. 1978) and may lead to underestimates of population size with N-mixture models (Kéry and Royle 2016, Veech et al. 2016). Individual heterogeneity in lek attendance \((p_s \text{ and } p_p)\) could occur for sage-grouse because yearling males attend leks at different rates than adults \((>1 \text{ yr old}; \text{ Jenni and Hartzler 1978, Walsh et al. 2004, Wann et al. 2019})\), yet age is difficult to ascertain during counts. Yearlings also may remain on the periphery of leks where they are less visible (Jenni and Hartzler 1978). Additionally, some males in a given year skip attending leks altogether (Blomberg et al. 2013, Gibson et al. 2014), and therefore, they may belong to the population but are not sampled during lek counts with probability \(1 - p_L\). In either case, N-mixture models cannot account for these sources of detection bias, yet the consequences for abundance and trend estimates are unclear.

One potential solution to unmodeled heterogeneity in detectability is incorporating auxiliary data in analyses (Martin et al. 2015, Ketz et al. 2018), which can be useful particularly for datasets that lack repeated counts.

Here, we evaluate performance of population assessment models under scenarios with confounded trends in abundance and detectability, using lek counts of sage-grouse as a model study system. We estimate population trends from scenarios with and without declining trends in \(p\) across years, and we compare estimates of population size and trends from dynamic N-mixture models and models where the response is the maximum observed number of males (uncorrected peak counts). For realism, we use empirical estimates of sage-grouse lek attendance and sightability to inform simulations of detectability in count data. Additionally, we evaluate effects of age and sampling probability \((p_s)\) on estimates, and the performance of estimators when auxiliary data are used to correct counts for changes in detectability.

**METHODS**

**Study area and lek attendance data**

To characterize temporal variation in lek attendance of males during spring, we monitored daily movements of sage-grouse at eight sites distributed throughout northern Nevada from 2012 to 2017 (Wann et al. 2019). These sites occurred in sagebrush-steppe and desert ecosystems with average site elevation ranging between 1732 and 2181 m. Each site contained 2–11 known active leks (i.e., with two or more male sage-grouse displaying at a lek during the study period). Additional details on local climate and floristic composition of sites are reported in Wann et al. (2019).
We collected global positioning system (GPS) locations from 79 individually marked male sage-grouse to determine when they were present on leks. All sage-grouse were captured in accordance with the U.S. Geological Survey Western Ecological Research Center Animal Care and Use Protocol (WERC-2015-02), and we describe sage-grouse capture and aging methods in Appendix S1. We fit captured males with a 30-g GPS platform transmitter terminal (North Star Science and Technology, King George, Virginia, USA), and we used a solar array and rechargeable batteries to power the GPS units that had an operational life between two and three years. We programmed the GPS units to collect one to three locations within three hours of sunrise (when birds are typically on leks), and we used the PTT Tracker decoding software to download locations (version 1.0.0.3.7; GeoTrak, Inc. 2011).

To determine whether grouse locations occurred at leks, we mapped lek boundaries in the spring and summer of 2016 and 2017 (Appendix S1). Error in GPS locations was generally <26 m but could be as high as 100 m. To reduce influence of this uncertainty, we restricted analyses to locations with error <26 m and we buffered lek polygons by 25 m. We also restricted analyses to data collected 1 January to 30 June and between 60 min before sunrise through 180 min after sunrise, and we considered males present on a lek if a given location occurred inside a buffered lek polygon during this time window. Several individuals were monitored over multiple years (16 over two years, six over three years, and one over four years), which yielded 110 individual-year attendance records including 85 adults and 25 yearlings.

**Lek attendance model**

We jointly modeled annual lek attendance probability (attending a lek at least once in a season; \( p_a \)) and within-season attendance probability (\( p_p \)) using a zero-inflated Bernoulli model with complete interactions for age (yearling or adult). We began by modeling annual lek attendance of individual-year \( k \) (\( X_k \)) as the outcome of a Bernoulli model with age-specific probability \( p_{a_{ge(k)}} \) where age(\( k \)) denotes a subscript for age (\( 1 = \) adult, \( 2 = \) yearling) depending on the individual-year: \( X_k \sim \text{Bernoulli}(p_{a_{ge(k)}}) \). We then modeled whether each individual was present at a lek during observation \( j \) (\( y_{jk} = 1 \)) with another Bernoulli distribution conditional on \( p_p \) and the latent state of an individual attending a lek at all within a year (\( X_k = 1 \)): \( y_{jk} | X_k \sim \text{Bernoulli}(p_p X_k) \).

We fit age-specific covariates for \( p_p \) on the logit scale (\( \beta_{a_{ge(k)}} \)) including ordinal date, minutes after sunrise (time), and random terms for year \( (\epsilon_{\text{year}(k)}) \) and site \( (\epsilon_{\text{site}(k)}) \); see Study area and lek attendance data:

\[
\logit(p_{P_k}) = \beta_{0_{a_{ge(k)}}} + \beta_{1_{a_{ge(k)}}}\text{date}_k + \beta_{2_{a_{ge(k)}}}\text{date}_k^2 + \beta_{3_{a_{ge(k)}}}\text{date}_k^3 + \beta_{4_{a_{ge(k)}}}\text{time}_k + \beta_{5_{a_{ge(k)}}}\text{time}_k^2 + \epsilon_{\text{site}(k)} + \epsilon_{\text{year}(k)}.
\]

Random terms were normally distributed with mean 0 and standard deviation \( \sigma \). For comparison and to inform our count corrections, we also analyzed location data without accounting for differences in age (no-age model) because age is typically not recorded during counts. Before analyses, we standardized continuous covariates by subtracting the mean and dividing by the standard deviation. We analyzed both models using a Bayesian approach with JAGS (version 4.2.0; Plummer 2003) and the rjags package (version 4-6; Plummer et al. 2016) in R (version 3.3.2; R Development Core Team 2016). After discarding 15,000 posterior samples during the adaptive phase, we saved 100,000 samples from each of three parallel chains for the age model, and 85,000 samples for the no-age model. We provide details on priors and assessment of convergence in Appendix S2.

**Population trend and lek count simulations**

We simulated scenarios with \( \lambda = 0.9, 1.0, \) and 1.1 for declining, stable, and increasing population trends, respectively. For each population trend, we evaluated effects of trends in detectability by creating three sets of simulations: (1) where \( p \) varied randomly (without trends across years); (2) with annual trends in \( p_p \) (probability of presence); and (3) with annual trends in \( p_p p_d \) (availability and detection, see below). For brevity, we did not consider scenarios with annual trends in both \( p_p \) and \( p_p p_d \) although these may occur in practice. To evaluate the influence...
of repeated counts on parameter estimates, we created scenarios with the same counts and population sizes but where we randomly assigned three repeated counts to a proportion of leks (0.3, 0.5, 0.75, and 1), resulting in 36 scenarios (three population trends × three detectability trends × four repeated count proportions). Leks without repeated counts were counted once each year. Finally, we assessed effects of unmodeled heterogeneity in detectability due to mixed ages of sage-grouse in observed samples by using the same scenarios except that all individuals were adults. Thus, we simulated 72 unique scenarios and we generated 250 random datasets for each scenario.

We begin by distinguishing between the population size of individuals associated with a lek \(N\) and the superpopulation size of individuals that attend a lek at least once in a year \(E(N^*) = N_p\). This distinction is important because the latter population size is typically the parameter of inference for N-mixture models (i.e., \(c_{ijt} | N^*_j \sim \text{binomial}(N^*_j, \pi_{ijt})\) where \(p = p_a p_b p_d\); Nichols et al. 2009), whereas the former includes individuals in the population that are not sampled with counts. We also note that although the number of individuals present can vary among lek counts (secondary periods) because \(p_p < 1\), population sizes \((N\) and \(N^*)\) are constant within a lekking season and vary across years. Each simulated dataset consisted of 100 leks monitored over 10 yr, with initial population size for each lek \(i\) in year \(t = 1\) \((N_{it})\) drawn from a negative binomial distribution with expected abundance \((\Lambda) = 30\) and overdispersion parameter \((\alpha) = 5\), where \(\sigma^2 = \Lambda + (\Lambda^2/\alpha)\). We simulated changes in the latent population size across years with an exponential model where \(N_{it}\) was a function of the previous year’s population size and the rate of population change \((\lambda\); Hostetler and Chandler 2015\): \(N_{it} \sim \text{Poisson}(N_{it−1}\lambda)\). To account for differences in annual and within-season attendance probabilities between yearlings and adults, we assumed that 70% of individuals at each lek and year were adults \((N_{adults})\) and the rest were yearlings \((N_{yearlings})\); Walsh et al. 2004), so that \(N_{it} = N_{adults} + N_{yearlings}\).

For the detection process, the number of adults in the superpopulation \(N^*_{adults}\) was drawn from a binomial distribution given latent population size \(N_{adults}\) and age-specific sampling probability (annual lek attendance; \(p_{s,adult}\)): 

\[
N^*_{adults} | N_{adults} \sim \text{binomial}(N_{adults}, p_{s,adult})
\]

The observed count of adults on occasion \(j\) \((c_{adults,j})\) was then a binomial outcome given \(p_{adults}\) and \(N^*_{adults}\):

\[
c_{adults,j} | N^*_{adults} \sim \text{binomial}(N^*_{adults}, p_{adults,j})
\]

where \(p_{adults}\) is the product of probabilities of presence, availability, and detection \((p_{p,adults}, p_a, p_d)\). We specified age-specific covariates for \(p_p\) (i.e., \(\beta_\lambda\) intercept for adults) on the logit scale including date and time:

\[
\logit(p_{p,adults,j}) = \beta_\lambda + \beta_{1a} \text{date}_{ijt} + \beta_{2a} \text{date}^2_{ijt} + \beta_{3a} \text{date}^3_{ijt} + \beta_{4a} \text{time}_{ijt} + \beta_{5a} \text{time}^2_{ijt}
\]

We used mean posterior estimates from the age model (see Lek attendance model) as coefficients for \(p_p\), date, and time parameters. For availability and detection probabilities, we assumed their product was the mean estimate previously reported from sightability models \((p_a p_d = 0.87\); Fremgen et al. 2016\). We similarly simulated counts of yearlings \((c_{yearlings})\), and the observed count used in analyses was:

\[
c_{ijt} = c_{adults,j} + c_{yearlings,j}.
\]

For scenarios with linear annual trends in detectability, we included a negative effect of year \((\beta_\lambda = −0.1\) on the logit scale and standardized) to either \(p_p\) or \(p_d\).

For each simulated dataset, we randomly assigned three dates (March 15–May 15) and three times (30 min before to 90 min after sunrise) to each lek each year corresponding with count protocols for when lek attendance is highest (Connelly et al. 2003). We then recorded date_{ijt}, time_{ijt}, and \(c_{ijt}\) for analysis, and \(N_{it}\) (the true, simulated population size) and \(N^*_{it}\) (the superpopulation size) for comparisons of estimator performance. We also determined the maximum count for each lek and year \((c_{ijt})\) for analyses with peak count models. We provide R code for simulating datasets via ScienceBase (https://doi.org/10.5066/p91l28pg).

**Model comparison**

We analyzed each dataset with six hierarchical population models where each model had the
same state process but a different observation process depending on whether the response was peak counts or repeated counts. We used an exponential model for the state process, and initial population sizes were estimated from a negative binomial distribution: $N_{ij} \sim \text{NB}(A, \alpha)$. We characterized the observation process with either a Poisson ($P$) or negative binomial (NB) distribution for peak counts (e.g., $c_t \sim \text{P}(N_{it})$), or a binomial observation process for repeated counts (Nmix). We fit N-mixture models with either a linear annual trend on $p$ ($\beta_0$) for scenarios with annual trends or a random term varying by year (offset) for scenarios without annual trends.

To study the use of auxiliary data for peak count models, we used parameter estimates from the no-age lek attendance model (see Lek attendance model) to correct peak counts for annual and within-season lek attendance rates ($P$ and NB+). We used mean posterior estimates of annual attendance ($p_t$), coefficients for within-season attendance ($p_p$), and simulated values for availability and detection ($p_d$) and applied these to peak counts ($c_{it}$), given the date and time associated with each count: $c_{corr} = c_{it} (p_t p_p p_d)$. For scenarios with linear trends in $p_p$ or $p_d$, we assumed that auxiliary data contained information on such trends and these were incorporated in count corrections. For N-mixture models, auxiliary data could help account for variation that is not estimable, such as $p_d$. Here (Nmix+), we used estimates of $p_s$ from the no-age attendance model and moment matching (Hobbs and Hooten 2015) to create an informative beta prior for $p_s$ where $\mu = \lambda/(\alpha + \beta)$, $\sigma^2 = \lambda\beta/((\alpha + \beta)^2(\alpha + \beta + 1))$, and $p_s \sim \text{beta}(\alpha, \beta)$. We then used this prior to estimate $N$ where $N_{it}^\alpha | N_{it}^\beta \sim \text{binomial}(N_{it}, p_s)$ and $c_{it} | N_{it}^\beta \sim \text{binomial}(N_{it}^\beta, p_{it})$. Using informative priors based on other analyses (rather than jointly estimating multiple models) is useful for reducing computation time of analyses (Martin et al. 2015).

For each model, we monitored $\hat{N}$ (estimated mean total population size: $1/T \sum_{t=1}^{T} \sum_{i=1}^{n} \hat{N}_{it}$ for $n = 100$ leks and over $T = 10$ yr) and $\hat{\lambda}$ (mean rate of population change) for comparison with the simulated, true values of $N$, $N^*$, and $\lambda$, and we saved 40,000 posterior samples from each of three chains after discarding 10,000 iterations from the adaptive phase. We summarized simulations with % bias ([est – truth]/truth $\times 100$), root mean square error (RMSE), proportion of 95% credible intervals (CrI) containing the simulated value (coverage), and credible interval length. We report priors in model code available in ScienceBase. Finally, for each model we evaluated fit using posterior predictive checks (Bayesian $P$ values) based on chi-square discrepancy test statistics (Kéry and Royle 2016, Conn et al. 2018), and we considered $P$ values <0.1 and >0.9 indicated lack of fit.

**RESULTS**

**Lek attendance**

Probability of attending a lek at least once in a season (sampling probability $p_s$; Table 1) was higher for adults ($p_{s\text{adult}} = 0.89, 95\%$ credible intervals [CrI] = 0.82–0.95) than yearlings ($p_{s\text{yearling}} = 0.75, 95\%$ CrI = 0.57–0.90). Overall annual lek attendance from the no-age model ($p_t$) was 0.87 (95\% CrI = 0.79–0.93), which is equivalent to beta(88.2, 13.2) that we supplied to Nmix+ models. We estimated substantial variation in within-season attendance probability ($p_p$), with presence peaking in early April and declining after sunrise (Fig. 1). We also estimated differences in attendance by age class, with attendance lower and somewhat later in the season and morning for yearlings than adults. Without accounting for age differences, within-season attendance probabilities were slightly lower than those of adults (Fig. 1). For simulation scenarios with a linear annual trend ($\beta_0 = -0.1$), median probability of presence ($p_p$) declined from 0.43 at $t = 1$ (range = 0.25–0.50) to 0.24 at $t = 10$ (range = 0.12–0.50) for yearlings, and from 0.72 at $t = 1$ (range = 0.45–0.81) to 0.52 at $t = 10$ (range = 0.25–0.81) for adults. For availability and detection ($p_d$), this same trend resulted in a decline from 0.91 to 0.81 over 10 yr.

**Population size ($N$ and $N^*$)**

Overall, bias in population size estimates varied little among scenarios with different annual trends in $N$ or $p$ (Fig. 2; Appendix S3). Estimates of population size were most biased from peak count models without corrections for $p$ (P and NB), and these models underestimated $N^*$ by >40% and $N$ by >50%. Absolute bias and RMSE were lowest, and coverage was highest, for $N$ from models with corrections for detectability.
(P+, NB+, and Nmix+). All peak count models increasingly underestimated \( N \) and \( N^* \) as the proportion of sites with repeated counts decreased, resulting in overestimates of \( N \) from P+ and NB+ models when all sites were counted repeatedly and underestimates when 30% of sites were counted repeatedly. Bayesian \( P \) values for negative binomial models suggested adequate fit (0.2 < \( P \) value < 0.8), whereas Poisson models generally indicated lack of fit (\( P \) value >0.9; Appendix S3: Tables S1, S3 and S5).

N-mixture models without an informative prior for \( p_s \) (annual lek attendance; Nmix) performed relatively well in estimating the number of individuals attending leks at least once (\( N^* \)) when all sites were counted repeatedly, and yielded less bias than peak count models without corrections but still underestimated \( N \) by \( \sim \)18% (Fig. 2). Informative priors for \( p_s \) reduced absolute bias in \( N \) from Nmix+ when all leks were counted repeatedly but still underestimated \( N \) by \( \sim \)6%. N-mixture models did not suggest lack of

Fig. 1. Within-season lek attendance probability (\( p_p \)) of sage-grouse by date (top) and minutes after sunrise (bottom) with differences by age (left) or averaged across age classes (right). Date predictions were based on mean time of 104 min after sunrise, whereas time predictions were based on mean date of 18 April. Shaded regions denote 95% credible intervals. Vertical red dashed lines represent the start and end dates (March 15–May 15) or time intervals (30 min before–90 min after sunrise) of lek counts.
Fig. 2. Violin plots summarizing % bias \((\text{est} - \text{truth})/\text{truth} \times 100\%) of posterior means of mean total population size estimates \(\hat{N}\) (averaged over 10 yr) relative to the true population size \(N\) and superpopulation size \(E(N^*) = Np_s\) for different simulated values of population rate of change \(\lambda\) and proportion of sites with repeated counts. Detectability \(p\) was the product of probabilities of sampling \(p_s\), presence \(p_P\), and availability and detection \(p_a p_d\), and scenarios were created either without a linear annual trend in \(p\) (A), with \(p_P\) declining annually (B), or with \(p_a p_d\) declining annually (C). Each model used either a Poisson (P) or negative binomial (NB) observation process with peak counts, or a binomial observation process with repeated counts (Nmix). We also corrected peak counts for variation in \(p\) (P+ and NB+) or included an informative prior for \(p_s\) (Nmix+). Violin plots denote the density of % bias from 250 simulated datasets for each scenario, black horizontal lines indicate the median from each scenario, and dashed lines represent no bias in parameter estimates.
fit when most sites were counted repeatedly, but as the proportion of sites with repeated counts decreased to 0.3, absolute bias and RMSE increased, coverage decreased, and Bayesian $P$ values indicated lack of fit (Appendix S3: Tables S2, S4 and S6).

**Population trends ($\lambda$)**

All models generally performed well in estimating $\lambda$ for scenarios without annual trends in $p$, and all but one model yielded similarly unbiased estimates of $\lambda$ (Fig. 3A) and low RMSE (Appendix S3: Tables S7 and S8). The exception was for NB+ models that tended to underestimate $\lambda$ slightly. Estimates of $\lambda$ were more precise (shorter credible interval lengths) from peak count models than N-mixture models, but this precision often excluded the true simulated value despite overall lack of bias in mean trend estimates. Coverage of $\lambda$ tended to be greater for N-mixture models (0.94–0.99; Appendix S3: Table S8) than models for peak counts (0.05–0.64; Appendix S3: Table S7). Coverage also was somewhat less for P+ and NB+ than P and NB models.

Most models performed worse under scenarios with annual trends in probability of presence ($p_p$). Peak count models consistently underestimated $\lambda$ when $p_p$ declined annually, and absolute bias in $\lambda$ increased with fewer repeated counts and as $\lambda$ increased (Fig. 3B). Correcting peak counts for detectability improved % bias, RMSE, and coverage for Poisson and negative binomial models (Appendix S3: Table S9). N-mixture models also produced more biased estimates when annual trends in $p_p$ occurred but performed well when all sites were counted repeatedly each year. As the proportion of leks with repeated counts decreased, N-mixture models increasingly underestimated $\lambda$, and specifying an informative prior for $p_s$ had little effect on $\lambda$ from Nmix+ (Fig. 3B; Appendix S3: Table S10). Similar patterns emerged for scenarios with trends in $p_p$ (Fig. 3C), but estimates tended to be less biased, RMSE smaller, and coverage greater (Appendix S3: Tables S11 and S12) than for scenarios with trends in $p_p$.

**Effects of age on model accuracy**

When we removed the effect of age on attendance by simulating all individuals as adults, all models yielded somewhat higher estimates of $N$ and $N^*$ (Appendix S4: Fig. S1) than scenarios with age differences (Fig. 2). However, this resulted in overestimates of $N$ and $N^*$ when we corrected peak counts for detectability ($P^*$ and NB+; Appendix S4: Fig. S1, Tables S1, S3 and S5). N-mixture models with and without an informative prior for $p_s$ showed greater coverage, reduced RMSE, and increased precision (shorter credible interval length; Appendix S4: Tables S2, S4 and S6) compared to scenarios with age differences (Appendix S3: Tables S2, S4 and S6), but again coverage decreased and absolute bias and RMSE increased with fewer repeated counts (Appendix S4: Fig. S1). Measures of fit followed similar patterns as scenarios with age differences. Removing age differences had little effect on $\lambda$. (Appendix S4: Fig. S2, Tables S7–S12).

**DISCUSSION**

Accounting for variation in detectability when estimating population size and trends is an active field of research, yet researchers seldom evaluate how models respond to systematic changes in detectability that could confound inferences on latent parameters of interest, and instead use simulation scenarios that assume $p$ is constant or random over space and time. Despite substantial variation in lek attendance by date and time, population trend estimates were similarly unbiased from dynamic N-mixture and peak count models when detectability did not vary systematically across years. Recovery of simulated population trends also was not affected by reduced effort (fewer repeated counts). However, we found that these scenarios offered an incomplete assessment of models, and this was apparent when we simulated trends in both population size and probability of presence ($p_p$) or probability of availability ($p_a$) and detection ($p_{ad}$), and therefore trends in $p$. In these cases, dynamic N-mixture models estimated $\lambda$ with little bias when all leks were counted repeatedly, peak count models without corrections underestimated population trends, and performance of peak count models improved when auxiliary information was incorporated into analyses. Absolute bias from N-mixture models also increased as the proportion of leks with repeated counts decreased. These results illustrate the importance of fully exploring the response of
Fig. 3. Violin plots summarizing % bias ([est – truth]/truth × 100%) of posterior means of mean population rate of change (λ, averaged over 10 yr) for different simulated values of λ and proportion of sites with repeated counts. Detectability (p) was the product of probability of sampling (pₛ), probability of presence (pₚ), and probability of availability and detection (pₚₚₚₑₚₚ), and scenarios either lacked a linear annual trend in p (A), pₚ declined annually (B), or pₚₚₑₚₚ declined annually (C). See Fig. 2 caption for model name definitions. Violin plots denote the density of % bias from 250 simulated datasets for each scenario, black horizontal lines indicate the median from each scenario, and dashed lines represent no bias in parameter estimates.
population models such as dynamic N-mixture to systematic changes in $p$ when evaluating model performance. Scenarios with trends in $p$ also may better identify situations where incorporating auxiliary data in analyses is useful for inferences.

Our results agree with previous studies indicating that N-mixture models are robust to reduced survey effort when estimating trends in lek populations while $p$ is constant or varies randomly (McCaffery et al. 2016). Similarities between N-mixture and uncorrected count models also were observed for datasets where $p$ was constant or random (Yamaura 2013, Barker et al. 2018). Our study suggests that these previous results likely did not indicate a particular attribute of the N-mixture model, but instead reflected the absence of trends in $p$ within these study systems (or the lack of covariates to estimate such trends). When $p$ is constant, or variation in $p$ is random and therefore the expectation (on average) is constant, detectability is less likely to confound inferences on population trends (Monroe et al. 2016). Simulation scenarios with constant (or random) $p$ are still useful for certain objectives, for example, to evaluate sensitivity of models to different levels of detectability (e.g., low vs. high $p$; Zipkin et al. 2014). However, without also simulating scenarios with trends in $p$, researchers assume, paradoxically, the same conditions under which indices can yield unbiased estimates of $\lambda$ and therefore obviate the need for methods that account for detectability when estimating population trends. Simulations with constant or random $p$ also may encourage use of models and study designs that cannot distinguish between potential trends in $p$ and $N$, such as uncorrected peak counts or dynamic N-mixture models with few or no repeated counts.

Most models under consideration here produced biased estimates of $N$. This was expected for uncorrected peak count models, and N-mixture models performed well in estimating the number of individuals that attended leks ($N^*$) when all leks were counted repeatedly. As with population trend estimates, survey effort had a strong influence on $N$, and absolute bias from uncorrected peak count and N-mixture models increased as the proportion of sites with repeated counts decreased. Additionally, both annual ($p_a$) and within-season ($p_p$) lek attendance of sage-grouse varied with age, and unmodeled age structure contributed to bias in $\hat{N}$. Previous studies found adult male sage-grouse attended leks earlier and more frequently than yearlings (Emmons and Braun 1984, Walsh et al. 2010, Wann et al. 2019). Yearlings also may attend alternative leks more often than adults (Emmons and Braun 1984, Schroeder and Robb 2003, but see Gibson et al. 2014, Fremgen et al. 2017, Wann et al. 2019), contributing to lower lek attendance rates. In addition, we estimated that some males did not attend a lek at all in a given year, and therefore could not be detected by lek counts. This phenomenon may be attributed to the hourly timing of our GPS recordings, although attendance rates in our study were similar to previous estimates from Nevada (Blomberg et al. 2013). Males not sampled also could be attending alternative leks (especially those unknown to observers), skipping reproduction, or mating off-lek (Blomberg et al. 2013, Gibson et al. 2014). We found dynamic N-mixture models underestimated $N$ by $\geq 18\%$, and absolute bias in $\hat{N}$ decreased when we assumed all individuals were adults because yearlings attended leks less frequently than adults and therefore were less likely to be counted. Dynamic N-mixture models still underestimated $N$ by $\geq 9\%$ in scenarios without yearlings due to individuals that did not attend leks, and additional data and models to estimate annual lek attendance ($p_a$) and yearling abundance could be useful when estimates of absolute population size are needed (Barker et al. 2018, Knape et al. 2018, Link et al. 2018). In contrast, unmodeled heterogeneity from age and attendance had little effect on $\hat{\lambda}$ from N-mixture models and so benefits from auxiliary data were more limited. Still, we assumed age ratios and annual attendance rates were constant across space and time, and thus, our results do not preclude the use of auxiliary data to account for trends in either factor.

Adjusting peak counts with auxiliary data generally yielded less biased estimates of $\hat{\lambda}$ and $\hat{N}$ for scenarios with declining trends in $p$ and often outperformed N-mixture models for scenarios where $\leq 75\%$ of leks were counted repeatedly. In the Great Basin of western North America, sage-grouse populations are increasingly threatened by wildfire and exotic grass spread, and counts conducted at $>1000$ leks are used for monitoring and to inform management (Coates et al. 2016).
Given that resources for monitoring are finite, detectability corrections could increase spatial representation by studying more leks less intensively. However, this approach also assumes that auxiliary data are comparable and appropriate to the study system (e.g., reflecting the population’s age structure), and we found that corrected count models overestimated \( N \) for scenarios with only adults, likely because attendance rates were based on data from adults and yearlings. Considering the extent of sage-grouse distribution and monitoring (Schroeder et al. 2004), additional covariates and replication over space and time may be needed for auxiliary data to be useful. Weighing the costs and benefits of collecting GPS data or conducting lek counts presents opportunities for a formal optimal sampling framework (e.g., Sanderlin et al. 2014, Reich et al. 2018, Williams et al. 2018), which could inform monitoring design while considering the resources and effort required for adequate inference.

Goodness of fit based on posterior predictive values indicated N-mixture models fit the data adequately when all leks were counted repeatedly but suggested increasing lack of fit as the proportion of sites with repeated counts declined. This occurred similarly across scenarios for both Nmix and Nmix+ and corresponded with increasing absolute bias in \( N \) but not in \( \lambda \). Additionally, posterior predictive checks were conservative, often indicating lack of fit (P value >0.9) when \( N \) was underestimated by >40% with Nmix and >20% with Nmix+. Lack of power to detect bias with posterior predictive checks and other goodness-of-fit tests also was noted in recent studies (Conn et al. 2018, Duarte et al. 2018, Link et al. 2018). Posterior predictive checks were even less informative for peak count models and corresponded with neither \( N \) nor \( \lambda \). These results may reflect the limitations of posterior predictive checks when evaluating bias in estimates of latent variables and processes (Kéry and Royle 2016). Furthermore, goodness-of-fit tests may perform worse when detectability is low and sample sizes are small (Duarte et al. 2018), although we did not evaluate this directly in our study. Additional simulation studies and more targeted discrepancy measures, graphical methods, and cross-validation could provide further insights into model adequacy (Kéry and Royle 2016, Conn et al. 2018, Knappe et al. 2018).

Although we simulated constant rates of annual lek attendance (\( p_a \)) across leks and years, attendance may vary systematically and could bias trend estimates from lek counts (Blomberg et al. 2013). For example, density dependence was previously estimated using peak counts (Garton et al. 2011, Monroe et al. 2017, Edmunds et al. 2018), but attendance also may be density-dependent (Blomberg et al. 2013) and could confound estimates of density dependence in \( \lambda \). Similarly, cumulative winter precipitation influences soil moisture and productivity of semi-arid ecosystems dominated by sagebrush (Artemisia spp.; Schlaepfer et al. 2012) and was positively correlated with peak attendance rates in Nevada (Wann et al. 2019). We also assumed the population age structure was constant across years, but the proportion of yearlings (with lower attendance rates than adults) will likely vary with reproductive success and recruitment rates. In addition, the relative contribution of individual detectability components to trends in \( p \) warrants further study, and we found scenarios with negative annual trends in \( p_a \) produced more bias in \( \lambda \) than scenarios with similar trends in \( p_a p_d \). Probability of presence also was generally lower than \( p_a p_d \), and lower detectability can contribute to greater bias in parameter estimates (Yamaura 2013, Zipkin et al. 2014, Kéry and Royle 2016, Veech et al. 2016, Duarte et al. 2018). Finally, we assumed no within-season mortality, but this too could influence \( p_s \) and \( p_a \) (Fremgen et al. 2017), and the implications for monitoring and population trend analyses are worthy of additional study.

We conclude that dynamic N-mixture models can distinguish between trends in \( p \) and \( N \) with a robust design where all or most sites are counted repeatedly each year, whereas these processes are confounded in models of indices and can lead to biased estimates of population trends. Use of auxiliary data also may be helpful for estimating \( N \) and \( \lambda \), particularly when repeated counts are not obtainable. Although variation in within-season lek attendance (\( p_a \)) was the greatest contributor to imperfect detection in our study system, this should not constrain our inferences to study systems with lekking species. Indeed, temporary emigration can bias estimates of population size among other bird species (Chandler et al. 2011). Our results can be extended to other systems where N-mixture...
models are applied, including single-season studies with closed populations that vary across space rather than time (Kéry 2008, Yamaura 2013, Veech et al. 2016), and more broadly to other hierarchical models that account for detectability. For example, occupancy models fit with constant \( \hat{p} \) overestimated occupancy trends because detection (which varied annually with abundance) and occupancy were confounded (Latif et al. 2018). Occurrence or prevalence of systematic changes in \( p \) is difficult to generalize because this likely varies by study system and research question. Increasing spatial and temporal extent of monitoring and randomization during study design may reduce, but not eliminate, the risk of systematic biases in detection. We therefore encourage researchers to investigate the response of models to trends in detectability and suggest caution when using dynamic N-mixture models with few or no repeated counts because these models may be susceptible to the same biases that arise from models of indices. Similarly, a previous study found that with sufficient spatial replication (>50 leks), only one count per lek each year was needed for adequate inferences on population trends (Fedy and Aldridge 2011), but our results suggest this may require revisiting in the context of trends in detectability. We also recommend additional consideration of potential biases from unmodeled variation in detectability, such as individual heterogeneity and sampling probability (\( p_{s} \)).

**Acknowledgments**

We thank Daniel Gibson, Adam Green, Kira Monroe, and two anonymous reviewers for providing helpful comments that improved this manuscript. This research was supported by funding from the U.S. Geological Survey (USGS) and the Wyoming Landscape Conservation Initiative. This research also used resources provided by the Core Science Analytics, Synthesis, and Libraries (CSASL) Advanced Research Computing (ARC) group at USGS. Collection of field data was supported by multiple federal and state agencies, including USGS, Bureau of Land Management, U.S. Fish and Wildlife Service, U.S. Forest Service, and Nevada Department of Wildlife (NDOW). We are particularly grateful to Mark Ricca (USGS), John Severson (USGS), Brian Prochazka (USGS), Michael Casazza (USGS), and Shawn Espinosa (NDOW) for their contributions to field study logistics, data, and conceptual design. We thank numerous technicians who helped locate, capture, and mark sage-grouse with GPS transmitters, particularly Katie Andrle and Steven Matthews. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

**Literature Cited**

Anderson, D. R. 2001. The need to get the basics right in wildlife studies. Wildlife Society Bulletin 29:1294–1297.
Anderson, D. R. 2003. Response to Engeman: Index values rarely constitute reliable information. Wildlife Society Bulletin 31:288–291.
Banks-Leite, C., R. Pardini, D. Boscolo, C. R. Cassano, T. Puttker, C. S. Barros, and J. Barlow. 2014. Assessing the utility of statistical adjustments for imperfect detection in tropical conservation science. Journal of Applied Ecology 51:849–859.
Barker, R. J., M. R. Schofield, W. A. Link, and J. R. Sauer. 2018. On the reliability of N-mixture models for count data. Biometrics 74:369–377.
Baumgardt, J. A., K. P. Reese, J. W. Connelly, and E. O. Garton. 2017. Visibility bias for sage-grouse lek counts. Wildlife Society Bulletin 41:461–470.
Beck, T. D. I., and C. E. Braun. 1980. The strutting ground count: variation, traditionalism, management needs. Proceedings of the Western Association of Fish and Wildlife Agencies 60:558–566.
Bellier, E., M. Kéry, and M. Schaub. 2016. Simulation-based assessment of dynamic N-mixture models in the presence of density dependence and environmental stochasticity. Methods in Ecology and Evolution 7:1029–1040.
Blomberg, E. J., J. S. Sedinger, M. T. Atamian, and D. V. Nonne. 2012. Characteristics of climate and landscape disturbance influence the dynamics of greater sage-grouse populations. Ecosphere 3:55.
Blomberg, E. J., J. S. Sedinger, D. V. Nonne, and M. T. Atamian. 2013. Annual male lek attendance influences count-based population indices of greater sage-grouse. Journal of Wildlife Management 77:1583–1592.
Chandler, R. B., J. A. Royle, and D. I. King. 2011. Inference about density and temporary emigration in unmarked populations. Ecology 92:1429–1435.
Coates, P. S., M. A. Ricca, B. G. Prochazka, M. L. Brooks, K. E. Doherty, T. Kroger, E. J. Blomberg, C. A. Hagen, and M. L. Casazza. 2016. Wildfire, climate, and invasive grass interactions negatively impact an indicator species by reshaping sagebrush ecosystems. Proceedings of the National Academy of Sciences 113:12745–12750.
Conn, P. B., D. S. Johnson, P. J. Williams, S. R. Melin, and M. B. Hooten. 2018. A guide to Bayesian model checking for ecologists. Ecological Monographs 88:526–542.

Connelly, J. W., K. P. Reese, and M. A. Schroeder. 2003. Monitoring of greater sage-grouse habitats and populations. College of Natural Resources Experiment Station Bulletin, Moscow, Idaho, USA.

Dail, D., and L. Madsen. 2011. Models for estimating abundance from repeated counts of an open metapopulation. Biometrics 67:577–587.

Duarte, A., M. J. Adams, and J. T. Peterson. 2018. Fitting N-mixture models to count data with unmodeled heterogeneity: bias, diagnostics, and alternative approaches. Ecological Modelling 374:51–59.

Edmunds, D. R., C. L. Aldridge, M. S. O’Donnell, and A. P. Monroe. 2018. Greater sage-grouse population trends across Wyoming. Journal of Wildlife Management 82:397–412.

Ellingson, A. R., and P. M. Lukacs. 2003. Improving methods for regional landbird monitoring: a reply to Hutto and Young. Wildlife Society Bulletin 31:896–902.

Emmons, S. R., and C. E. Braun. 1984. Lek attendance of male sage grouse. Journal of Wildlife Management 48:1023–1028.

Fedya, B. C., and C. L. Aldridge. 2011. The importance of within-year repeated counts and the influence of scale on long-term monitoring of sage-grouse. Journal of Wildlife Management 75:1022–1033.

Fremgen, A. L., C. P. Hansen, M. A. Rumble, R. S. Gamo, and J. J. Millspaugh. 2016. Male greater sage-grouse detectability on leks. Journal of Wildlife Management 80:266–274.

Fremgen, A. L., C. P. Hansen, M. A. Rumble, R. S. Gamo, and J. J. Millspaugh. 2019. Weather conditions and date influence male sage grouse attendance rates at leks. Ibis 161:35–49.

Fremgen, A. L., C. T. Rota, C. P. Hansen, M. A. Rumble, R. S. Gamo, and J. J. Millspaugh. 2017. Male greater sage-grouse movements among leks. Journal of Wildlife Management 81:498–508.

Garton, E. O., J. W. Connelly, J. S. Horne, C. A. Hagen, A. Moser, and M. A. Schroeder. 2011. Greater sage-grouse population dynamics and probability of persistence. Studies in Avian Biology 38:293–381.

GeoTrak, Inc. 2011. PTT tracker, version 1.0.0.3.7. North Star Science and Technology, LLC, King George, Virginia, USA.

Gibson, D., E. J. Blomberg, M. T. Atamian, and J. S. Sedinger. 2014. Lek fidelity and movement among leks by male greater sage-grouse Centrocercus urophasianus: a capture-mark-recapture approach. Ibis 156:729–740.

Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, New Jersey, USA.

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.

Jenni, D. A., and J. E. Hartzler. 1978. Attendance at a sage grouse lek: implications for spring censuses. Journal of Wildlife Management 42:46–52.

Johnson, D. H. 2008. In defense of indices: the case of bird surveys. Journal of Wildlife Management 72:857–868.

Kéry, M. 2008. Estimating abundance from bird counts: Binomial mixture models uncover complex covariate relationships. Auk 125:336–345.

Kéry, M., R. M. Dorazio, L. Soldaat, A. Van Strien, A. Zuiderwijk, and J. A. Royle. 2009. Trend estimation in populations with imperfect detection. Journal of Applied Ecology 46:1163–1172.

Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. Volume I: prelude and static models. Academic Press, San Diego, California, USA.

Kéry, M., J. A. Royle, and H. Schmid. 2005. Modeling avian abundance from replicated counts using binomial mixture models. Ecological Applications 15:1450–1461.

Ketz, A. C., T. L. Johnson, R. J. Monello, J. A. Mack, J. L. George, B. R. Kraft, M. A. Wild, M. B. Hooten, and N. T. Hobbs. 2018. Estimating abundance of an open population with an N-mixture model using auxiliary data on animal movements. Ecological Applications 28:816–825.

Knappe, J., D. Arlt, F. Barraquand, A. Berg, M. Chevalier, T. Pirtt, A. Ruete, and M. Zmihorski. 2018. Sensitivity of binomial N-mixture models to overdispersion: the importance of assessing model fit. Methods in Ecology and Evolution 9:2102–2114.

Latif, Q. S., M. M. Ellis, V. A. Saab, and K. Mellen-McLean. 2018. Simulations inform design of regional occupancy-based monitoring for a sparsely distributed, territorial species. Ecology and Evolution 8:1171–1185.

Lindenmayer, D. B., et al. 2012. Value of long-term ecological studies. Austral Ecology 37:745–757.

Link, W. A., and J. R. Sauer. 1998. Estimating population change from count data: application to the North American Breeding Bird Survey. Ecological Applications 8:258–268.

Link, W. A., M. R. Schofield, R. J. Barker, and J. R. Sauer. 2018. On the robustness of N-mixture models. Ecology 99:1547–1551.
Martin, J., H. H. Edwards, C. J. Fonnesbeck, S. M. Koslofsky, C. W. Harmak, and T. M. Dane. 2015. Combining information for monitoring at large spatial scales: first statewide abundance estimate of the Florida manatee. Biological Conservation 186:44–51.

McCaffery, R., J. J. Nowak, and P. M. Lukacs. 2016. Improved analysis of lek count data using N-mixture models. Journal of Wildlife Management 80:1011–1021.

Monroe, A. P., C. L. Aldridge, T. J. Assal, K. E. Veblen, D. A. Pyke, and M. L. Casazza. 2017. Patterns in greater sage-grouse population trends estimates. Journal of Wildlife Management 80:667–678.

Nichols, J. D. 1991. Extensive monitoring programmes viewed as long-term population studies: the case of North American waterfowl. Ibis 133:89–98.

Nichols, J. D. 1992. Capture-recapture models: using marked animals to study population dynamics. BioScience 42:94–102.

Nichols, J. D., L. Thomas, and P. B. Conn. 2009. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201–235 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. Modeling demographic processes in marked populations. Springer-Verlag, New York, New York, USA.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62:3–135.

Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Pages 1–10 in K. Hornik, F. Leisch, and A. Zeileis, editors. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, Austria, March 20–22, 2003.

Plummer, M., A. Stukalov, and M. Denwood. 2016. rjags: Bayesian graphical models using MCMC. R package version 4-6. http://cran.r-project.org/package=rjags

Pollock, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. Environmetrics 13:105–119.

R Development Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B. J., K. Pacifici, and J. W. Stallings. 2018. Integrating auxiliary data in optimal spatial design for species distribution modelling. Methods in Ecology and Evolution 9:1626–1637.

Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Sanderlin, J. S., W. M. Block, and J. L. Ganey. 2014. Optimizing study design for multi-species avian monitoring programmes. Journal of Applied Ecology 51:860–870.

Schlaepfer, D. R., W. K. Lauenroth, and J. B. Bradford. 2012. Consequences of declining snow accumulation for water balance of mid-latitude dry regions. Global Change Biology 18:1988–1997.

Schroeder, M. A., and L. A. Robb. 2003. Fidelity of greater sage-grouse Centrocercus urophasianus to breeding areas in a fragmented landscape. Wildlife Biology 9:291–299.

Schroeder, M. A., et al. 2004. Distribution of sage-grouse in North America. Condor 106:363–376.

Veech, J. A., J. R. Ott, and J. R. Troy. 2016. Intrinsic heterogeneity in detection probability and its effect on N-mixture models. Methods in Ecology and Evolution 7:1019–1028.

Walsh, D. P., J. R. Stiver, G. C. White, T. E. Remington, and A. D. Apa. 2010. Population estimation techniques for lekking species. Journal of Wildlife Management 74:1607–1613.

Walsh, D. P., G. C. White, T. E. Remington, and D. C. Bowden. 2004. Evaluation of the lek-count index for greater sage-grouse. Wildlife Society Bulletin 32:56–68.

Wann, G. T., P. S. Coates, B. G. Prochazka, J. S. Severson, A. P. Monroe, and C. L. Aldridge. 2019. Assessing lek attendance of male greater sage-grouse using fine-resolution GPS data: implications for population monitoring of lek mating grouse. Population Ecology 61:183–197.

White, G. C. 2005. Correcting wildlife counts using detection probabilities. Wildlife Research 32:211–216.

Williams, P. J., M. B. Hooten, J. N. Womble, G. G. Esslinger, and M. R. Bower. 2018. Monitoring dynamic spatio-temporal ecological processes optimally. Ecology 99:524–535.

Yamaura, Y. 2013. Confronting imperfect detection: behavior of binomial mixture models under varying circumstances of visits, sampling sites, detectability, and abundance, in small-sample situations. Ornithological Science 12:73–88.

Zipkin, E. F., J. T. Thorson, K. See, H. J. Lynch, E. H. Campbell Grant, Y. Kanno, R. B. Chandler, B. H. Letcher, and J. A. Royle. 2014. Modeling structured population dynamics using data from unmarked individuals. Ecology 95:22–29.
DATA AVAILABILITY

R scripts, JAGS code, and data needed to run simulations and fit models are available via ScienceBase (https://doi.org/10.5066/p91l28pg).

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2791/full