Made available through Montana State University's ScholarWorks

Development of optimized methods for unbiased dusky grouse population monitoring using real and simulated data

Elizabeth A. Leipold; Claire Gower; Lance B. McNew

Accessibility Disclaimer:

For a more accessible version of this document, please submit an accessibility request form through the Montana State University Library website.

Check for updates

WILDLIFE BIOLOGY

Research Article

Development of optimized methods for unbiased dusky grouse population monitoring using real and simulated data

Elizabeth A. Leipold[1], **Claire Gower**[2] **and Lance B. McNew**[1]

[1]Department of Animal and Range Sciences, Montana State University, Bozeman, MT, USA
[2]Montana Fish, Wildlife and Parks, Bozeman, MT, USA

Correspondence: Elizabeth A. Leipold (elizabethleipold7@gmail.com)

Wildlife Biology
2026: e01506
doi: 10.1002/wlb3.01506
Subject Editor: Emmanuel Menoni
Editor-in-Chief: Ilse Storch
Accepted 12 January 2026

Rigorous state-wide monitoring programs are lacking for dusky grouse *Dendragapus obscurus*, a North American species of forest grouse with relatively low detectability that is found in coniferous and mountainous areas in the western United States and Canada. Hierarchical models for estimating abundance show promise for overcoming issues associated cryptic forest grouse behavior and low population densities. Our objectives were to evaluate protocols (i.e. number of sites, visits, and route type) and analytical methods for producing annual unbiased and precise indices of abundance (CV < 15%) to inform a statewide monitoring program. During 2019–2022, we designed and implemented multiple survey protocols throughout western Montana, USA, including spring point-counts (> 2200 unique sites) and transect-level (> 390 unique transects) distance sampling. We used an iterative process of field data collection and simulation analyses to evaluate the performance of four different statistical estimators for abundance (N-mixture model, hierarchical time-removal model with distance sampling, detection-naïve model, and hierarchical distance sampling model) for point-counts and transects to produce unbiased and precise estimates of dusky grouse abundance. Simulations demonstrated that increasing the number of sites visited or probability of detection decreased the requisite amount of survey effort for obtaining precise abundance estimates for all estimators. Unbiased and precise estimates of abundance were unachievable under most realistic point-based distance sampling protocols. N-mixture protocols where point-counts conducted during periods of high probability of detection at 80 sites visited four times per area of inference (e.g. study area or region) resulted in unbiased estimates of population size with the highest precision. Our study provides baseline information necessary for the development of state-wide monitoring programs of dusky grouse and more broadly illustrates an approach for developing rigorous and achievable monitoring programs for other species of forest grouse.

Keywords: *Dendragapus obscurus*, hierarchical distance sampling, hierarchical distance sampling with time-removal, N-mixture model, simulation study

NORDIC SOCIETY OIKOS

www.wildlifebiology.org

© 2026 The Author(s). Wildlife Biology published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

Unbiased and precise estimates of wildlife abundance are crucial for effective conservation, management, and understanding ecological relationships (Joseph et al. 2009, Weiser et al. 2019, Neubauer and Sikora 2020). Estimates of abundance are used to define conservation priorities, assess whether population objectives are being met and evaluate the impact of management actions (Wiest et al. 2019, Doser et al. 2021). Without accurate estimates, limited resources can be used inefficiently, and managers may be unable to identify population declines in a timely manner (Sillett et al. 2012). Population monitoring is especially important for game species where changes in abundance may impact hunting quotas and other aspects of species' management (Jakob et al. 2014). Given high costs and the large amount of effort needed to obtain long-term data on abundance, it is important that monitoring programs are well planned and able to make efficient use of limited resources while still achieving target goals (Anderson and Steidl 2019, Kidwai et al. 2019, Weiser et al. 2019).

Effective population monitoring programs should clearly define the purpose of the monitoring program, what population parameters will be estimated, how the results will be used, and the level of accuracy and precision desired (Witmer 2005, Field et al. 2007). Survey designs, sampling methods, and statistical analyses should be evaluated for their ability to work with the species' ecology and account for observation errors in order to produce unbiased and adequately precise estimates of abundance (Pollock et al. 2002, Witmer 2005). Ideally, survey protocols and analytical methods will be continuously reevaluated as more data become available to ensure that recommended methods adequately address management objectives (Gibbs et al. 1999). Lastly, an effective monitoring program must be implementable and achievable regarding number of personnel, cost, accessibility of survey sites, and spatiotemporal survey effort, and be able to provide easily interpretable outcomes to conservationists and managers (Gibbs et al. 1999, Witmer 2005).

Populations of dusky grouse *Dendragapus obscurus*, a species of forest grouse that occur in mountainous habitats of the western USA and Canada, are poorly or inconsistently monitored despite their distribution-wide status as a game species and concerns over the potential influences of environmental disturbances (e.g. logging, wildfire, beetle infestations) on their habitats (Bendell and Elliot 1967, Martinka 1972, Pelren and Crawford 1999, Chan-McLeod and Bunnell 2004, Youtz et al. 2022). Dusky grouse habitats are characterized as partially open conifer forests often dominated by ponderosa pine *Pinus ponderosa* and Douglas fir *Pseudotsuga menziesii*, although the exact composition can vary across their range and includes other fir, spruce *Picea* spp., and pine species (Zwickel and Bendell 2004). Remote and often difficult to access mountainous habitats, low population densities, and low detection rates have made obtaining unbiased and precise estimates of population size of dusky grouse inherently difficult (Rogers 1963) and may limit survey and analytical methodologies.

To date, population monitoring of dusky grouse, like that of other grouse species, has been based on: 1) unstandardized count-based surveys during the spring breeding season, 2) brood counts during the summer, and 3) reported hunter harvests in the fall (Rogers 1963, Bernales et al. 2017, Espinosa et al. 2018, Gates 2019, New Mexico Department of Game and Fish 2020). Survey site placement along roads is common for spring breeding surveys and summer brood counts; however, convenience sampling associated with roads may induce biases in abundance if grouse avoid or select areas near roads due to human disturbance or if roads occur in preferred habitats (Rogers 1963, Hanowski and Niemi 1995, Betts et al. 2007, Buckland et al. 2008, Sands and Pope 2010). Recent research has identified survey protocols that maximize the detectability of grouse (Leipold et al. 2025), but quantitative assessments of the performance of survey protocols (e.g. sample selection, spatial and temporal survey effort) and analytical approaches for estimating or indexing the population size of dusky grouse are lacking.

Count-based surveys of grouse may provide an unbiased index of population size across time and space in the presence of random sampling and consistent observation error (Pollock et al. 2002, Rosenstock et al. 2002). Unfortunately, observation error is rarely constant across space or time and measuring this error to adjust estimates of abundance is critical for proper inference about population size and trends (Rosenstock et al. 2002, Farnsworth et al. 2005). The expansion of analytical methods that estimate and incorporate the observation error (i.e. probability of detection (p)) into abundance estimation with unmarked animals has advanced population monitoring programs for black grouse *Lyrurus tetrix*, red grouse *Lagopus lagopus scoticus*, red-legged partridge *Alectoris rufa*, rock ptarmigan *Lagopus muta*, and greater sage-grouse *Centrocercus urophasianus* (Warren and Baines 2011, Franceschi et al. 2014, Jakob et al. 2014, McCaffery et al. 2016) and may be useful for dusky grouse. Two classes of hierarchical models that can be used to obtain unbiased estimates of abundance for unmarked populations of grouse include: 1) those that estimate detection probabilities through repeated counts during a period of population closure, and 2) those that adjust counts based on estimated distance-decay functions describing detection probability (Kéry and Royle 2016).

Of the first type, N-mixture models (aka binomial-Poisson mixture models) are a class of hierarchical models that use simple counts replicated spatially and temporally within a period of closure to estimate probability of detection (p) and abundance (Royle 2004, Kéry and Royle 2016). The relative simplicity and low costs of count-based survey protocols needed for N-mixture models imply utility over large spatial and temporal scales consistent with statewide and regional application (Kéry and Royle 2016). Additionally, N-mixture models allow for the inclusion of spatial and temporal covariates to evaluate questions related to: 1) survey protocol effects on detection probabilities, 2) site-specific effects of habitat conditions or treatments on local abundance and detection, and 3) spatial and temporal effects of climate and management on population size and trajectories (Romano et al. 2017, Katayama et al. 2020, Dinkins et al. 2021).

19003220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Like all statistical models, N-mixture models have critical assumptions that must be met to ensure abundance estimates are unbiased: 1) counts occur within a period of population closure, 2) all observation errors are negative; no false positives, 3) detection probability is constant for all individuals within a site during a survey, 4) individuals are detected independently of other individuals, and 5) the distributions of abundance and detection probability are adequately described by their chosen parametric distribution (Kéry and Schaub 2012, Kéry and Royle 2016). Estimates of abundance can be sensitive to violations of model assumptions (Barker et al. 2018, Duarte et al. 2018, Knape et al. 2018, Link et al. 2018), but nevertheless, N-mixture models yield accurate estimates of abundance when assumptions are reasonably met (Basile et al. 2016, Kéry 2018, Duarte et al. 2018, Ficetola et al. 2018).

A second class of abundance models, collectively referred to as distance sampling models, uses measured observation distances, rather than repeat surveys, to estimate p. Distance sampling models assume that the probability of detecting an individual decreases with increasing distance from the observer and that a distance–decay detection function can be used to estimate the proportion of individuals unidentified during a survey (Buckland et al. 2001, 2015, Rosenstock et al. 2002). Recent extensions of the distance sampling model (e.g. hierarchical distance sampling) allow models to incorporate covariates for abundance and detection (Kéry and Royle 2016). A major advantage of distance sampling is that sites (points or transects) need only be visited once per survey period. Hierarchical distance sampling (HDS) has been used to estimate density, model relationships between density and habitat associations, and explore relationships between detection and habitat associations and survey conditions for a variety of species including butterflies, endangered bumblebees, lizards, songbirds, and game birds (Hamm 2013, McNeil et al. 2018, Furnas et al. 2019, Sullins et al. 2019, Reidy et al. 2021, Lewis et al. 2022).

Traditional distance sampling models provide estimates of perceptibility, the probability of detection given an individual is available to be detected (p_d). Expansions on HDS, including time-removal HDS models that integrate distance sampling and time-removal models, can provide additional inference on the probability that animals are available to be detected (p_a), and can yield more accurate estimates of abundance or density if temporary movement in and out of the survey area during a survey period is common (Amundson et al. 2014). By jointly considering the distance from the observer and the time of initial detection for individuals, time-removal HDS allows the separate estimation of p_a and p_d, which may improve estimation of abundance. Time-removal HDS also easily allows for the inclusion and evaluation of covariates for both components of detection (p_a, p_d) and abundance, but require the time of individual observations to be recorded during surveys.

Like N-mixture models, HDS models yield unbiased estimates of abundance if assumptions of the model are reasonably met. The assumptions are: 1) animals are distributed uniformly in space and independently of the transects or point locations on which the observations are made, 2) probability of detection is a function of distance and detection is perfect at a distance of zero, 3) individuals are detected at their original locations, and 4) distances are measured without error (Buckland et al. 2001, 2015, Thomas et al. 2010, Kéry and Royle 2016). Additional assumptions for the time-removal HDS model are: 1) individuals are identified correctly in reference to species and double counting of individuals does not occur, 2) availability and perceptibility are independent, 3) surveys occur within a period of closure, and 4) all individuals within a population are present at a site during the survey so that probability of presence equals 1 (Nichols et al. 2009, Amundson et al. 2014).

HDS and time-removal HDS only require one visit per site, which is an advantage over N-mixture models where multiple visits during a period of population closure are needed (Kéry and Royle 2016). On the other hand, the data required for N-mixture models is less intensive as it does not require accurately measuring distances to individuals which can be difficult for auditory signals or defining the time an individual is first detected (Scott et al. 1981, Simons et al. 2007, Amundson et al. 2014, Buckland et al. 2015). In addition, distance sampling requires that a minimum of 60–100 individual detections at a variety of distances are obtained in order to reasonably estimate the distance–decay function, which may decrease their utility in the face of low detection or abundance (Buckland et al. 2001, Rosenstock 2002, Rusk et al. 2007, Franceschi et al. 2014).

Based on discussions with wildlife managers at Montana's state wildlife management agency (Montana Fish, Wildlife, and Parks (FWP)), our goal was to design and implement a population monitoring program for dusky grouse that produces unbiased and precise estimates of annual abundance with a coefficient of variation (CV) of $\leq$ 15% that will allow wildlife managers to obtain precise enough estimates of abundance to detect biologically meaningful changes in population size over timeframes relative to management. We addressed this goal through the following sequential objectives:

1) We used a series of simulations over a 3-year survey period to identify and refine appropriate field survey effort (number of survey sites and visits) for obtaining accurate estimates of abundance;
2) Compared abundance estimates from off-trail, trail, and low-use road-based point-counts to evaluate potential biases associated with convenience-based sampling;
3) Used field survey data and simulations to evaluate the relative performance, measured via estimated bias and precision under varying levels of survey effort, of N-mixture, HDS, time-removal HDS, and detection-naïve models of raw grouse counts. We sought to identify the analytical approach that produced unbiased (i.e. near the truth) and precise (CV $\leq$ 15%) indices of population size with the least amount of required survey effort; and
4) Conducted a power analysis to evaluate whether the final recommended survey protocol was sufficient for accurately estimating population trends of dusky grouse across 3, 5, and 10-year periods.

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

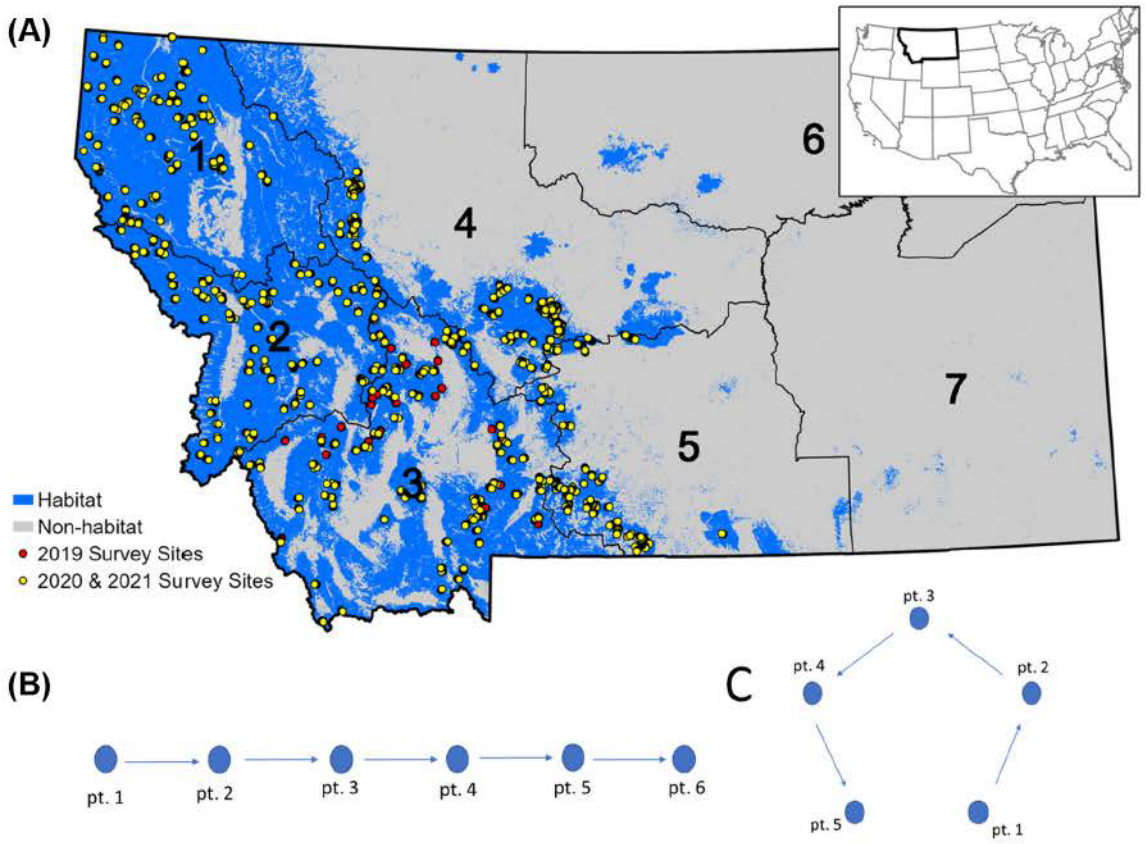


Figure 1. (A) Study area map of Montana, US. Survey sites were located in areas predicted to be dusky grouse *Dendragapus obscurus* habitat (Leipold et al. 2024) within 7 different administrative regions. (B) Walking route used along roads or trails in 2020–2021. (C) Pentagon shaped walking route used offtrail in 2019.

Study area

Our field surveys occurred over three years (2019–2021) and throughout areas predicted to be dusky grouse habitat in western Montana, USA (Fig. 1; Leipold et al. 2024). Montana FWP divides Montana into 7 administrative regions that differ in habitat composition and management objectives. The majority (99%) of dusky grouse habitat occurs in MFWP Regions 1–5. During the pilot year (2019), we restricted grouse surveys to MFWP Region 3. In 2020–2021, we expanded the study area to include all dusky grouse habitat in MFWP Regions 1–5.

Predicted dusky grouse habitat was dominated by coniferous and mountainous forests in west–central (MFWP Regions 4–5) and western portions of the state (MFWP Regions 1–3). Within the study area, average elevation was 1784 m (US Geological Survey 2017). From north to south within Montana, the composition of coniferous forest varied from being co-dominated by Douglas fir and ponderosa pine (MFWP Region 1) to being dominated by lodgepole pine *Pinus contorta* (MFWP Region 2, the north portion of MFWP Region 3, and parts of MFWP Region 4) and Douglas fir (eastern portion of MFWP Region 2, MFWP Regions 3–5; Landfire 2016). Weather conditions were similar during the survey periods in 2019–2021 and fell within the 30-year norm for the study area (Leipold et al. 2025). For precipitation the 30-year average was 76 cm annually (SD = 32), 7 cm in April (SE = 3), 8 cm in May (SD = 2), and for mean spring temperature the 30-year average was 2.4°C (SD = 2.1) in April and 7.2°C (SD = 2.1) in May (PRISM Climate Group 2024).

Material and methods

We used a phased approach to address our study objectives that included iterative reassessment of protocol and analytical performance as additional field data became available (Fig. 2), generally described in the following steps:

Step 1. To inform initial pilot season surveys, use non-empirical estimates (i.e. best guesses) of mean local abundance (λ) and detection probability (p) to simulate a population of dusky grouse and evaluate the performance of several potential survey protocols to recover unbiased estimates of N, λ and p.

Step 2. Implement a pilot study in a test region in 2019 based on the most appropriate protocols identified in initial simulations in Step 1.

Step 3. Analyze survey data from pilot season to obtain more realistic estimates of λ and p.

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

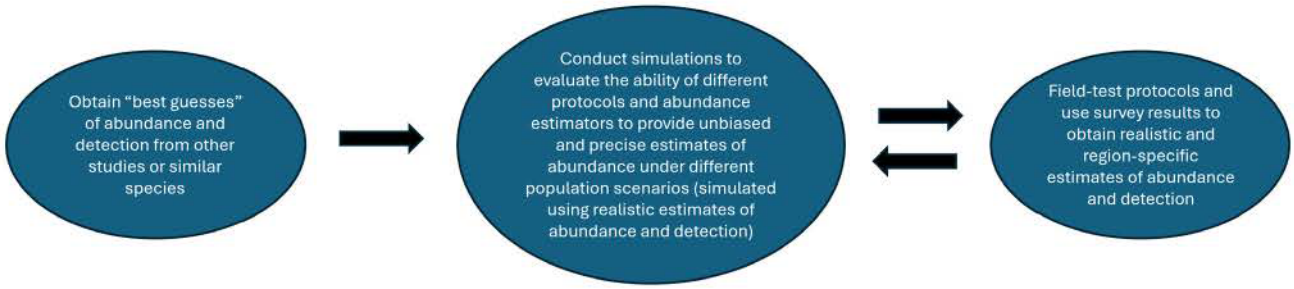


Figure 2. We used an iterative process to evaluate and reassess different protocols and statistical estimators; initially using 'best guesses' of abundance and detection and then refining those guesses with region-specific field data.

Step 4. Repeat simulations and analyses of Step 1 using empirical estimates from Step 2 to better identify most appropriate methods to implement in full survey effort during 2020–2021.

Step 5. Implement most appropriate survey protocols identified in Step 4 during 2020–2021 across the distribution of dusky grouse in Montana to obtain realistic estimates of the variability in λ and p across Montana's management regions.

Step 6. Using statewide survey data collected in Step 5, estimate mean local abundance, total number of grouse, and detection probabilities using 4 statistical estimators: N-mixture models, HDS, time-removal HDS, and detection-naïve count models.

Step 7. Evaluate the performance, via bias and precision of estimates, of the 4 different statistical models under different survey protocols and different measures of local regional abundance and detection probability obtained during Step 6.

Step 8. Conduct additional simulations to identify the power of survey/analytical protocols to accurately identify 1–10% changes in abundance over 3, 5, and 10-year assessment intervals.

Step 9. Make recommendations on most rigorous, accurate, precise, and achievable survey protocols and analytical methods for monitoring dusky grouse.

General field-based survey methods

Field biologists (56 current and former state agency wildlife biologists and 15 university wildlife researchers and technicians) and 7 volunteer naturalists with experience in grouse identification conducted spring point-count surveys for dusky grouse. All observers were trained in auditory and visual grouse identification and survey protocols by the lead author (EL). Surveys were conducted along walking routes consisting of 5–6 survey points during 10 April–21 May in 2019 and 10 April–1 June in 2020 and 2021. During the spring, the average male dusky grouse home range is ~ 0.8 ha, and therefore we separated survey locations by a distance $\geq$ 400 m to ensure independence between survey points (Martinka 1972).

We recorded survey conditions before conducting the pair of point-count surveys at each site, including survey day relative to the first day of the survey period (10 April), start time which was used to calculate minutes since sunrise, temperature, wind speed, precipitation, cloud cover, and background noise level (only in 2020–2021) which mainly consisted of noise from streams and occasionally road traffic. We measured temperature (°C) and wind speed (km hr^{-1}) using a hand-held weather meter (Kestrel model 2000, Kestrel Meters). Precipitation was classified as none, rain, snow, and fog. Cloud cover was classified as 0–15, 16–50, 51–80, and 81–100% of the sky covered. Noise level was classified as none (0), slight background noise but no hearing impairment (1), moderate background noise and some hearing impairment (2), and high levels of background noise and hearing impairment (3). Evaluation and estimation of survey condition effects on dusky grouse detection probability are described in Leipold et al. (2025).

Step 1. Simulations to inform pilot season field surveys

To evaluate potential protocols for the 2019 pilot season, we constructed a series of 16 simulation scenarios and analyzed the simulated data using single-season N-mixture models in a Bayesian framework, which allowed us to estimate all desired parameters with estimates of uncertainty. N-mixture models are composed of two linked hierarchical sub-models where the variation in local abundance is described as a Poisson random process (ecological process), and the variation in counts is described as a binomial random process (observation process):

$$N_i \sim \text{Poisson}(\lambda)$$
$$y_{i,j} \mid N_i \sim \text{Binomial}(N_i, p), \quad (1)$$

where N_i is the true abundance at site i, $y_{i,j}$ is the observed count at site i during replicate survey j and p is the average probability of detecting a grouse during a survey (Royle 2004, Kéry and Schaub 2012, Kéry and Royle 2016). We specified vague prior distributions: a gamma distribution (0.005, 0.005) for λ and a uniform distribution (0, 1) for p (Kéry and Schaub 2012). A description of the simulation data and model in the BUGS language is provided in the Supporting information.

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

For our 16 simulation scenarios, we varied the number of independent survey sites (50, 100, 200, and 500) and number of visits (2–3), simulating data under two mean local abundance (λ) scenarios with the same probability of detection (p). Estimates (i.e. best guesses) of abundance and detection for simulation scenarios were informed by preliminary work in northeastern Utah that observed that average dusky grouse abundance in good to excellent habitat ranged from 0.625 to 1.25 grouse per point-count area and that detection probability during a survey was similar across sites and averaged 0.5 (Dahlgren unpubl.). We sampled local abundance from a Poisson distribution with mean λ and simulated observations of grouse at each site during each survey by drawing randomly from a binomial distribution, where probability of detection (p) had a mean of 0.5. We used the 'R2WinBugs' package to call WinBugs in R (www.r-project.org) to analyze our simulated datasets and used vague priors previously described for all hyperparameters that provided little or no information about the estimated parameters (Supporting information; Kéry and Schaub 2012, Kéry and Royle 2016, Kellner 2019). We ran three chains of length 40 000 after a burn-in period of 10 000 and thinned the posterior chains by 100. We assessed convergence visually using trace plots and by using the Gelman–Rubin ($\hat{R}$) statistic, which examines the variance ratio of the Markov chain Monte Carlo (MCMC) algorithm within and between chains across iterations (Gelman and Rubin 1992). We accepted parameter estimates when they came from Markov chains with $\hat{R}$ between 1.0 and 1.1.

For every simulation scenario, we evaluated bias in the estimates of mean local abundance (λ), local abundance ($\hat{N}_i$) per site, total abundance $\hat{N}$ estimated as the sum of the estimated site-specific abundance (Total $\hat{N} = \sum \hat{N}_i$), and probability of detection (p). To quantify bias for each estimate, we ran 400 iterations of each data simulation and analysis for each survey protocol scenario and calculated the difference between the estimated parameter and true parameter value. We compared the posterior distributions of the mean differences between each estimate and the true values across all 400 simulations and considered an estimate to be clearly biased if the 95% credible interval (CrI) of the differences did not include 0. At each of the MCMC iterations we calculated the coefficient of variation $\left(\text{CV\%} = \frac{\text{estimated standard deviation}}{\text{mean parameter estimate}} \times 100\right)$ for total abundance ($\hat{N}$), evaluating the posterior distributions of the 400 derived CV estimates. We estimated the probability that the average coefficient of variation would meet the manager-determined threshold of CV ≤ 15% by calculating the proportion of the total posterior distribution density greater than 15%. We considered protocols that resulted in a CV ≤ 15%, 90% of the time to be acceptable for meeting the objective of the monitoring program.

Step 2. Pilot season surveys in 2019

During 2019, we randomly located initial survey points at each walking survey route ≥ 300 m from a road or trail in areas predicted to be dusky grouse habitat (Montana Spatial Data Infrastructure 2017, US Forest Service 2020a, 2020b, 2020c, 2020d, Leipold et al. 2024). Subsequent survey points were placed 500 m apart in a pentagon shape that was oriented away from anthropogenic structures (e.g. roads and trails) to ensure that the starting and ending points were relatively close to decrease the necessary time for accessing the survey area (Fig. 1). We surveyed each site during 3 different mornings within a 2-week period. We recorded the number and locations of all dusky grouse, while walking between points and during each point-count. Care was taken to not double count grouse. Grouse detected along the route (i.e. transect) and during a point-count were recorded independently for each method. Point-count surveys consisted of two 4-min point-counts conducted consecutively. The first point-count was conducted without the use of electronic grouse calls and the second point-count was conducted with electronic playback of a female sooty grouse call recordings (cantus, whinny, and cackle; Mike Schroeder, Washington Department of Fish and Wildlife, unpubl.; SanDisk 8 GB Clip Jam Mp3 Player manufactured by Western Digital Technologies, Inc., JBL Charge 3 speaker manufactured by JBL Harman) played on max speaker volume. We used a female sooty grouse call because we lacked a high-quality recording of dusky grouse, sooty grouse were classified as the same species as dusky grouse until 2006, and dusky grouse males have been shown to respond to female calls of other grouse species (Harju 1974, Leipold et al. 2025). Playback recordings consisted of alternating 30 seconds of calling and 30 seconds of silence until the entire four minutes of the survey had elapsed. For each initial detection of a dusky grouse, the distance to each grouse, vocalization, behavior, and sex, if known, was recorded, with distance to grouse measured to the nearest meter with a laser rangefinder. For auditory detections, the location of a grouse was narrowed to a specific area and then distance to that area was measured. To control for measurement error in exact distance for auditory detections, we placed distances into four bins: 0–25, 26–50, 51–75, and 76–100 m. In addition, for use with removal models, for point-counts we also recorded the minute of the survey in which the detection first occurred.

Step 3. Analyze pilot season data

To obtain empirical estimates of abundance and detection using our field data, we used single season N-mixture models (Royle 2004) evaluated using the *pcount* function within the R package 'unmarked' (Fiske and Chandler 2011). We analyzed data from the 2019 point-counts conducted with electronic playback using single-season N-mixture models (Kéry and Royle 2016). We estimated mean local abundance (λ) and probability of detection (p) by specifying a model with constant detection and abundance. We evaluated potential overdispersion in our count data by evaluating model support using Akaike's information criterion (AIC) for N-mixture models with different abundance distributions: Poisson, which assumes that variance is proportional to λ, and negative binomial and zero-inflated Poisson models

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that estimate dispersion parameters (Burnham and Anderson 2002, Kéry and Schaub 2012, Kéry and Royle 2016). We found a Poisson error distribution was most supported by the data, indicating no evidence of overdispersion in our count data, and was therefore used for all subsequent N-mixture models (Supporting information).

Step 4. Update simulations based on pilot season results

To inform protocols for 2020–2021 surveys, we used empirical estimates of mean local abundance ($\lambda = 0.36$, 95% CI: 0.18, 0.73) and detection ($p = 0.28$, 95% CI: 0.13, 0.50; see Results) from our 2019 pilot season data (Step 3) to update our simulation scenarios and more fully evaluate potential survey protocols. We first used simulations to evaluate the efficacy of our pilot season survey methods (3 replicated visits at 100 independent survey sites located off-trail) and then after examining the results, we evaluated whether estimator precision could be increased by 1) increasing the number of replicate survey visits per point, and 2) increasing numbers of independent survey points. We evaluated simulated datasets based on 100 independent survey points with increasing numbers of replicate visits and then increasing number of survey points with three replicate visits. Next, we kept the number of visits at 4 and varied the number of survey points from 200–360. To examine the logistical feasibility of the potential protocols that yield relatively precise results, we calculated how many survey mornings would be needed if survey transects included 5 or 6 points per transect, and 3 or 4 replicates occurred in one morning. Overall, we conducted a series of 13 simulations analyzed using single-season N-mixture models in a Bayesian framework using JAGS (Plummer 2003). Our simulation parameters, specification of vague priors, and metrics of bias and precision were the same as previously described simulations.

Step 5. Implement protocols identified from simulations during 2020–2021

Due to manager feedback regarding the prohibitive effort of accessing randomly-located survey routes in Montana's remote and mountainous landscapes in 2019, we restricted survey routes in 2020–2021 to be located on established hiking trails or along low-use roads within areas identified as suitable dusky grouse habitat (Leipold et al. 2024). We randomly located the first survey point of a route to be located $\leq$ 150 m of a trailhead or along an unimproved or low-use road and generated 5 additional survey points along the route (total of 6 points); points were separated by $\geq$ 400 m. Detection probability of dusky grouse did not differ among survey routes located on low-use roads, trails, or off-trail (Leipold et al. 2025). We randomly generated 722 potential survey routes, and among these available randomly-generated survey routes, selected routes each year that were accessible to observers during the spring survey period (April–May); snowpack frequently limited accessibility to mountainous areas resulting in a quasi-random sampling design.

Effects of route type on abundance

To evaluate potential sampling-induced biases in convenience sampling along unimproved or low-use roads and trails, we evaluated the effect of route type (roads, trails, off-trail) on estimated abundance using field data collected during 2019–2021 in MFWP Region 3, by evaluating an N-mixture model in which route type was included as a covariate for λ. To account for potential differences in detection probability among years, we included covariates representing survey conditions (minutes since sunrise, survey day, cloud cover, and temperature) previously found to influence p (Leipold et al. 2025). We evaluated support for the effect of route type on abundance by assessing whether the 95% confidence intervals (CI) for effect size overlapped 0 and by comparing abundance estimates for off-trail, trail, and road point-count surveys.

Step 6. Estimate average and regional abundances from 2020–2021 surveys

We used three models evaluated using functions within the R package 'unmarked' to obtain empirical estimates of abundance and detection: single-season N-mixture models (*pcount* function), hierarchical distance sampling models (*distsamp* function), and HDS time-removal models (*gdistremoval* function; Fiske and Chandler 2011). To obtain estimates of average local abundance across and within MFWP Regions 1–5, we evaluated the point-count survey data with single-season N-mixture models and HDS models (Kéry and Schaub 2012, Kéry and Royle 2016). We evaluated models in which mean p and λ were held constant to obtain average state-wide estimates of abundance. To obtain regional estimates of abundance, we evaluated models in which p was specified to be constant but λ varied by MFWP administrative region.

We evaluated HDS models for grouse observation data collected at both survey points and survey transects. For our point-based HDS models, we used observation data from first visit to each survey point to fit and evaluate a constant abundance model with three different detection functions: half-normal, hazard rate, and uniform (Buckland et al. 2001). For our transect-based HDS models, we evaluated the three different detection functions for replicate visits 1 and 2. We used AIC to rank and select the most appropriate detection function for estimating local abundance (Burnham and Anderson 2002); we found the half-normal to be most supported for both point-counts and transects and subsequently used the half-normal function for all HDS analyses (Supporting information). We then compared regional estimates of local abundance for point-counts for the N-mixture and HDS models; estimated local abundance was similar among models (see Results) and we used estimated λ from the N-mixture model to inform our simulation scenarios. Dusky grouse calls can be heard up to 100 m (Blackford 1958, Mussehl 1963, Zwickel and Bendell 2004), and therefore for our line-transect HDS models we extrapolated the abundance estimate per transect from the point-count abundance estimates from the

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

N-mixture models where the point-count radius for potential detections was estimated to be 100 m.

Estimates of detection probability for the three different hierarchical models are based on different parameters. For the N-mixture model we estimated probability of detection, p, which is described by a binomial random process that represents the variation in counts in the observed data. For HDS models, the multinomial cell probabilities for each detection class depend on the parameter σ, which is used to estimate the half-normal decay function. For the HDS time-removal model, we estimated both σ, for use in estimating the half-normal decay function, and p_a, which represents the per interval probability of detection for each minute and is used to estimate the probability of availability (ϕ; Amundson et al. 2014). Estimates for the N-mixture were obtained from using data from all 4 visits; estimates for the HDS and time-removal HDS models were obtained from the first visit to a site or transect.

Probability of detection for dusky grouse can vary based on survey conditions (Leipold et al. 2025), and therefore we estimated an average probability of detection and a high probability of detection. To obtain an average probability of detection, we evaluated models with constant p and λ. To obtain an optimistic high probability of detection representing surveying under optimal survey conditions, we predicted detection probability (p, σ, p_a) under recommended survey conditions (Leipold et al. 2025). For the point-based N-mixture and HDS models, p and σ, respectively, were modeled as a function of all survey conditions: precipitation, cloud cover, wind speed, survey day, minutes since sunrise, and background noise level. For time-removal HDS models, probability of detection is broken into two components: availability (p_a) and perceptibility (σ). We modeled p_a as a function of precipitation, cloud cover, wind speed, survey day, and minutes since sunrise, while we modeled σ as a function of background noise level. We did not have transect level covariates for survey conditions, so for HDS line-transects we modeled σ as a function of survey day to obtain our high probability of detection.

Before fitting models to predict high probability of detection, we examined the possibility of non-linear relationships (quadratic) between probability of detection and minutes since sunrise and survey day due to known temporal display behaviors of grouse (Bendell and Elliot 1967, Zwickel and Bendell 2004, Leipold et al. 2025). We used $[x + x^2]$ to represent the quadratic form where detection probability may be maximized at some intermediate value of x. We evaluated support for non-linear relationships using AIC to evaluate univariate models for the two different functional responses for the N-mixture, HDS, and time-removal HDS for point-count surveys and HDS for line-transect surveys.

Step 7. Evaluate performance of different abundance models under different survey scenarios

We used results from Step 6 to parameterize final population simulations and evaluations of the performance of four different statistical estimators for monitoring dusky grouse populations in Montana: single season N-mixture models, HDS models, time-removal HDS models, and to verify the importance of incorporating observation error for obtaining unbiased estimates, a detection-naïve Poisson model. To evaluate survey effort required to achieve accurate annual estimates of dusky grouse abundance with CV ≤ 15% from point-count and transect survey protocols using each of the abundance models, we developed simulated datasets based on empirical estimates of abundance and detection probabilities from our 2020–2021 spring survey effort (identified during Step 6) and estimated abundance in a Bayesian framework using the four statistical estimators: single-season N-mixture models (Eq. 1), HDS (Eq. 2), time-removal HDS (Eq. 3), and a naïve model (Eq. 4). We used a Bayesian framework because of increased model flexibility to estimate all the desired parameters of interest ($\lambda, \hat{N}_i$ per site, $\hat{N}, p, \sigma, p_a, \phi$) with estimates of uncertainty.

We developed 6 simulated datasets on which to apply our 4 statistical abundance models; the 6 scenarios represented realistic abundance or detectability situations for dusky grouse recovered from analyses of the 2020–2021 data (Table 1):

1) Average abundance with average detection,
2) High abundance with average detection,
3) Low abundance with average detection,
4) Average abundance with high detection,
5) High abundance with high detection, and
6) Low abundance with high detection.

We defined average, high, and low mean local abundances (λ) for our simulation scenarios based on empirical estimates from our 2020–2021 statewide surveys (Table 1, 2). We used the statewide estimate from all surveys to define the 'average' λ scenario (0.18, 95% CI: 0.17, 0.20; Results), the lowest estimated regional λ to define the 'low' abundance scenario (0.08, 95% CI: 0.06, 0.11; MFWP Region 4), and the highest estimated regional λ to define the 'high' abundance scenario (0.31, 95% CI: 0.27, 0.37; MFWP Region 2). We defined 'average' detection scenarios for each model type using statewide average values of detection parameters, p for N-mixture models (0.37, 95% CI: 0.35, 0.40), σ for HDS point-counts (43, 95% CI: 38, 47), σ for HDS line-transects (43, 95% CI: 38, 47), σ for time-removal HDS (42, 95% CI: 38, 47), and p_a for time-removal HDS (0.23, 95% CI: 0.15, 0.34). Simulated data for 'high' detectability scenarios were parameterized using maximum empirical estimates of p (0.57, 95% CI: 0.52, 0.62), σ for HDS point-counts (58, 95% CI: 38, 86), σ for HDS line-transects (51, 95% CI: 38, 86), σ for time-removal HDS (48, 95% CI: 41, 55), and p_a (0.43, 95% CI: 0.16, 0.69) estimated under optimal survey conditions identified in our preliminary assessments of covariate effects on detectability.

To implement HDS models within a Bayesian framework, we used a three-part multinomial, binomial, Poisson mixture model as a random variable cannot be specified as a multinomial index in BUGS language (Kéry and Royle 2016, p. 453):

$$y_s \mid n_s \sim \text{Multinomial}\left(n_s, \pi_s^c\right) \quad (2)$$

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Empirical parameter estimates for local abundance (λ) and detection parameters of dusky grouse used to parameterize simulation scenarios. Abundance estimates (λ) were used to inform scenarios for all models. σ was used to inform the detection function for HDS models, p was used to inform the probability of detection for the single-season N-mixture models, and ϕ (calculated using p_a) was used to inform the probability of availability for the time-removal HDS models. PC = point-count survey; Line = line-transect survey. Study occurred in western Montana from 2020–2021.

Model	Survey type	Parameter	Estimate	95% CI
All	–	Low λ	0.08	0.06, 0.11
All	–	Average λ	0.18	0.17, 0.20
All	–	High λ	0.31	0.27, 0.37
N-mixture	PC	Average *p*	0.37	0.35, 0.40
N-mixture	PC	High *p*	0.57	0.52, 0.62
HDS	PC	Average σ	43	38, 47
HDS	PC	High σ	58	38, 86
Time-removal HDS	PC	Average σ	42	38, 47
Time-removal HDS	PC	High σ	48	41, 55
Time-removal HDS	PC	Average ϕ	0.65	–
Time-removal HDS	PC	High ϕ	0.89	–
HDS	Line	Average σ	42	37, 49
HDS	Line	High σ	51	42, 61

Where $\pi_k^c = \pi_k / (1 - \pi_0)$, with index k representing the kth element of the vector π_s^c, and $1 - \pi_0$ represents total detection probability.

$$n_s \mid N_s \sim \text{Binomial}(N_s, 1 - \pi_0)$$

$$N_s \sim \text{Poisson}(\lambda_s)$$

Within the model, we used a conditional multinomial observation model where we conditioned on the observed data n_s, the number of individuals observed at site s, instead of the latent variable, N_s (Kéry and Royle 2016). The multinomial distribution in the first part of the model describes the model for the distance class of n_s, the observed individuals which are the result of imperfect detection of N_s as described in the second part of the model (Kéry and Royle 2016). In the third part of the model, local abundance (N_s) is estimated as a Poisson random variable with a mean λ_S (Kéry and Royle 2016).

We used a four-part model hierarchical model as described by Kéry and Royle (2016, p. 473–474) and Amundson et al. (2014) for the time-removal HDS model:

$$\begin{aligned} M_s &\sim \text{Poisson}(\lambda_s) \\ \log(\lambda_s) &= \alpha_0 + \alpha_1 \times x_s \\ N_s &\sim \text{Binomial}(M_s, \phi) \\ n_s &\sim \text{Binomial}(N_s, \overline{p}_s), \end{aligned} \tag{3}$$

where $\overline{p}_s$ is made of two components:

$$\text{tint} \sim \text{Categorical}(\pi_a^c)$$

$$\text{dclass} \sim \text{Categorical}(\pi_d^c),$$

and M_S represents the local population size at a site (s) and is estimated as a Poisson random variable with a mean λ_S (Kéry and Royle 2016). The second part of the model represents the number of individuals available to be detected during the distance sampling survey, (N_S), with parameters probability of availability (ϕ) and M_S. The probability of availability (ϕ) describes the probability that an individual is available for detection at all during the survey and is related to the per-interval probability of availability, p_a, where $\phi = 1 - (1 - p_a)^{\text{number of intervals}}$ (Kéry and Royle 2016). The number of individuals detected at site s, n_s, is the result of a random binomial draw based on the number available to be detected (N_s) and the net probability of an individual being detected at all in any distance class, $\overline{p}_s$ (Kéry and Royle 2016). Last, conditional on n_s, the distributions for two categorical individual covariates for distance class (*dclass*) and time interval (*tint*) are specified (Kéry and Royle 2016). Cell probabilities for *dclass* are dependent on the distance-based detection model, while cell probabilities for *tint* are dependent on p_a (Kéry and Royle 2016).

The detection-naïve model does not incorporate terms to account for imperfect detectability and is fitted using a simple Poisson regression:

Table 2. Estimates of local abundance (λ) for dusky grouse evaluated using single season N-mixture models and hierarchical distance sampling (HDS). Study occurred in western Montana from 2020–2021.

	λ (95% CI)	
Region	N-mixture	HDS
1	0.13 (0.10–0.17)	0.12 (0.08–0.19)
2	0.31 (0.27–0.37)	0.36 (0.27–0.47)
3	0.19 (0.16–0.23)	0.21 (0.15–0.29)
4	0.08 (0.06–0.11)	0.07 (0.04–0.12)
5	0.21 (0.17–0.26)	0.23 (0.16–0.33)
Average	0.18 (0.17–0.20)	0.20 (0.16–0.24)

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$N_i \sim \text{Poisson}(\lambda) \quad (4)$$

We used JAGS (package 'jagsUI') in R to analyze our simulated datasets and used vague priors for all hyper-parameters that provided little or no information about the estimated parameters: for λ we used a gamma distribution (0.005, 0.005 or 0.001, 0.001); for p in the N-mixture model we used a uniform distribution (0, 1); for σ in the HDS model for point-counts and line-transects, and time-removal HDS we used a uniform distribution (0, 100); for probability of availability (p_a) in time-removal HDS we used a uninformative uniform distribution (0, 1; Kéry and Schaub 2012, Kéry and Royle 2016, Kellner 2019; Supporting information). We varied the length of chains, burn-in period, and thinning based on simulation needs for each model type to ensure that convergence of chains on parameters occurred (Supporting information). We assessed convergence using the Gelman–Rubin ($\hat{R}$) statistic and accepted parameter estimates when they came from Markov chains with $\hat{R}$ between 1.0 and 1.1 (Gelman and Rubin 1992).

We quantified bias for detection parameters p, σ, and ϕ, mean local abundance (λ), estimated number of grouse per site ($\hat{N}_i$, $\hat{N}_{s_i}$ or $\hat{M}_i$), and total abundance $\hat{N}$ estimated as $\sum \hat{N}_i$ or $\sum \hat{N}_s$ (or $\hat{M}$ estimated as $\sum \hat{M}_i$ for time-removal HDS) by simulating and evaluating data 500 times for each survey protocol scenario and calculated the difference between the estimated parameter and the true parameter at each MCMC iteration. We then compared the posterior distributions of the mean differences between each estimate and the true values across all 500 simulations for each scenario and considered an estimate to be biased if the 95% credible interval (CrI) of the differences did not include 0. At each of the iterations we also calculated the coefficient of variation for total abundance ($\hat{N}$ or $\hat{M}$) and evaluated the posterior distribution of the 500 derived estimates of coefficient of variation (CV). We estimated the probability that the mean coefficient of variation would meet the desired level of precision of a CV ≤ 15% by calculating the proportion of the total CV posterior distribution density > 15%. We considered a protocol that resulted in a CV ≤ 15%, 90% of the time to be acceptable for the monitoring program.

For the N-mixture, hierarchical distance sampling and time-removal HDS models we evaluated whether estimator precision could be increased by increasing the number of independent survey sites. For our simulated survey protocols for point-counts, we increased the number of sites visited each time by 100 until we achieved unbiased and relatively precise (CV < 15%, 90% of the time) estimates of population abundance. We then decreased the number of sites by 20, evaluating the different protocols until the model no longer produced estimates with the desired level of precision, after which we increased the number of sites by 10 to evaluate the midpoint between the thresholds to identify the requisite number of sites. For line transects we started with 100 sites, and then as 100 sites was more than sufficient for reaching our desired level of precision, we decreased the number of sites from 100 by 20 until the CV was no longer ≤ 15%, 90% of the time. At that point we then increased the number of sites by 10 and then decreased by 5 to further narrow down the number of sites that need to be visited.

For our N-mixture models we also varied the number of visits to a site, evaluating survey protocols with 2, 3, or 4 visits. Conducting the necessary replicate surveys on different days requires much additional effort. To reduce effort and address the primary assumption that replicate surveys occur within period of population closure (Royle 2004), we conducted all replicate survey visits at each survey point within a route on the same morning. Another assumption, however, is that counts are conditionally independent (Royle 2004), which could potentially be violated under our same-day replicate approach. To evaluate the potential effects of violating this assumption, we 1) estimated correlations in our replicated counts across survey visits, and 2) evaluated whether measures of observed correlations among counts resulted in biased estimates of the p and λ using additional simulations and analyses (Supporting information). We found some evidence of positive correlation ($r = 0.67$) between dusky grouse counts that occurred back-to-back (visits 1 and 2, and visits 3 and 4), but little evidence of correlation between the non-back-to-back visits (visits 1 and 3, 1 and 4, 2 and 3, and 2 and 4; Supporting information). Simulations revealed that when the true probability of detection was ≥ 0.57, the proposed sampling effort and protocol produced unbiased estimates of detection and local abundance despite modest correlation in counts among survey visits (Supporting information). We found modest upward bias in detection probability (approximately +0.10–0.11) and downward bias in local abundance (approximately −0.01 to −0.04 birds per survey area) when detection probability was ≤ 0.37, our average empirical probability of detection (Supporting information).

For the hierarchical distance sampling (HDS) models, we evaluated models for both point-counts and line-transect surveys using a half-normal detection function determined previously to be appropriate for our data (Supporting information). For the line transect surveys we conducted simulations for transect lengths of 2681 m (the average transect length in our study) and 5000 m. For both the line transect and point-count simulations we ran three chains of length 5000 after a burn-in period of 1000 and thinned the posterior chains by 1.

For the time-removal HDS model, we evaluated data only under two simulation scenarios due to computational constraints and preliminary results indicating the number of point-counts needed (> 6000) would be logistically unfeasible even for the high abundance, average detection scenarios. We evaluated the model for both high abundance scenarios, and given previous patterns, we assumed the scenarios with high abundance would require the lowest amounts of survey effort. For these simulations we ran three chains of length 20 000 after a burn-in period of 1000 and thinned the posterior chains by 1.

We evaluated detection-naïve models for point-counts where a site was only visited once and based the number

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of sites (60, 80, 140, 170, 240, 490) visited on the recommended survey protocol out of the other three model types for the different scenarios. For the simulations, we ran three chains of length 3000 after a burn-in period of 100 and thinned the posterior chains by 1.

Step 8. Power of protocols to detect population trends

Finally, we evaluated the ability of our recommended survey protocol and annual abundance estimation method to detect 3, 5, and 10% declines in population size over 3, 5, and 10 years. We generated an initial average local abundance of 0.18 grouse per site for year 0 across 80 sites based on the mean λ estimates from survey data collected during 2020–2021. We then projected the population through annual time steps by modeling the persistence of birds at each site using a binomial random draw where the probability of persistence was 1 – a target annual trend (e.g. 1–0.03 for a 3% annual decline; Steidl et al. 2013, Anderson and Steidl 2019). We simulated dusky grouse counts for each survey visit in each year by randomly drawing from a binomial distribution with parameters $\hat{N}_t$ and p, where $\hat{N}_t$ = total estimated population size in year t, and p = probability of detection for four visits estimated from our 2020–2021 dataset. We then used the N-mixture model in 'unmarked' to estimate total abundance in each year using the simulated count data. We regressed the estimated total true and estimated population sizes onto years to get the slope of the population trends for the true annual abundances and estimated abundances, respectively. We used an alpha level of 0.10 to determine if the slope of the trend was different than 0 and ran this simulation 500 times (Gibbs and Melvin 1997, Steidl et al. 2013). The proportion of the 500 simulation runs where mean slopes were different from 0 represented the power of the protocol to recover 3, 5, and 10% declines in the population over 3, 5, and 10 years; we considered an acceptable protocol to have a power ≥ 80% (Steidl et al. 2013, Weiser et al. 2019). We quantified the percentage of simulated runs in which the slope was predicted to be negative. We also calculated the mean trend for the estimated abundances and the mean difference between the trends estimated using the estimated abundances and the true abundances, calculating a 95% confidence interval as the 0.025 and 0.975 quantiles estimated from the 500 simulation runs.

Results

Simulations to inform field survey protocols

2019 pilot season protocols and results

Results of our initial simulation analyses indicated that 3 replicate surveys at each of 100 independent survey sites within a period of closure yielded unbiased and relatively precise (CV ≤ 15%) indices of regional population abundance when the average site-specific abundance (λ) was ≥ 0.625 grouse and the probability of detecting a grouse (p) averaged 0.5 (Supporting information). We used these results to develop our initial survey protocols for spring surveys of dusky grouse in 2019. We surveyed 90 sites, the majority of which (98%) were surveyed three times; we were unable to survey an additional 10 sites due to inaccessibility, weather, and logistical constraints. Applying the N-mixture model to the pilot seasons survey data resulted in lower estimates of average local abundance (λ = 0.36, 95% CI: 0.18, 0.73) and detection probability (p = 0.28, 95% CI: 0.13, 0.50) than the best guesses (0.625, 0.5, respectively) we used in initial simulations. However, our empirical estimates of λ and p had low precision (CV > 40%) and 95% CIs included original best guesses.

2020–2021 season survey protocols

We used empirical estimates of λ and p, from our analysis of the 2019 dusky grouse survey data to inform simulations of potential statewide survey protocols for 2020–2021. Assessments of the performance of our original survey effort (100 sites visited 3 times) parameterized with empirical estimates from 2019 data yielded relatively accurate but imprecise estimates of λ (CV > 15%; Supporting information). Simulations revealed that reasonable precise estimates of λ were only achievable with 3 replicate visits when the number of survey sites was > 600 (Supporting information). For a sample of 100 survey points, ≥ 7 replicate visits were needed to yield unbiased and suitably precise (CV ≤ 15%) indices of regional population abundance (Supporting information). Simulation results indicated that unbiased and relatively precise (CV ≤ 15%) indices of regional population abundance were achievable under a protocol in which 360 points were surveyed 4 times (Supporting information). Based on these results, we designed survey protocols for 2020–2021 in which ≥ 360 sites (points) were surveyed 4 times per MFWP region.

Effects of route type on abundance

We conducted point-count surveys at 90 off-trail sites, 273 road sites, and 448 trail sites within MFWP Region 3 during 2019–2021. We found that the effects of road and trail surveys, relative to those off-trail, on abundance were not supported (roads, β = −0.08, 95% CI: −0.47, 0.31; trails, β = −0.20, 95% CI: −0.58, 0.18) and estimated local abundance of dusky grouse was similar across survey route types. The estimated local abundance (λ) was 0.26 (95% CI: 0.16, 0.41), 0.24 (95% 0.18, 0.32), and 0.21 (95% CI: 0.16, 0.28) for surveys occurring off-trail, on low-use roads, and on trails, respectively.

Empirical estimates of abundance and detection from 2020–2021 data

During 2020–2021, we conducted 3292 sets of point-counts across western Montana (each set varying between 1–4 repeat visits) across 2372 sites and walked 551 line-transects two times. Due to incomplete data at some sites, we only used 3123 sets of point-counts (each with 4 repeat visits) for the N-mixture model analysis, 3234 point-counts for the

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

point-count HDS and time-removal HDS analyses, and 514 transects for the transect HDS analysis.

Abundance – Average estimates of local abundance (λ) from point-counts were 0.18 (95% CI: 0.17, 0.20) dusky grouse from the N-mixture model and 0.20 (95% CI: 0.16, 0.24) from the HDS model (Table 2). Regional estimates of λ were similar across the N-mixture and HDS using point-count data (Table 2). Estimated λ was lowest for MFWP Region 4, where the N-mixture model estimated a local abundance of 0.08 (95% CI: 0.06, 0.11) and the hierarchical distance sampling model estimated a local abundance of 0.07 (95% CI: 0.04, 0.12; Table 2). Estimated λ was greatest in MFWP Region 2, where the N-mixture model estimated a local abundance of 0.31 (95% CI: 0.27, 0.37) and the HDS model estimated a local abundance of 0.36 (95% CI: 0.27, 0.47; Table 2). Because, on average, the N-mixture model produced more precise estimates of local abundance (CV = 17 versus 28%), we used estimates from N-mixture models for each region to inform our subsequent simulation scenarios; we used a low local abundance of 0.08, an average local abundance of 0.18, and a high local abundance of 0.31 to parameterize simulation scenarios (Table 1).

N-mixture model probability of detection – The average probability of detection was 0.37 (95% CI: 0.35, 0.40). A quadratic relationship between probability of detection and survey day and minutes from sunrise was most supported (Supporting information). The maximum estimate of probability of detection occurring on 13 May and at 86 min post-sunrise was 0.57 (95% CI: 0.52, 0.62; Supporting information).

HDS model probability of detection for point-counts and line-transects – For point-counts, estimated average σ, which describes the half-normal detection function, was 43 (95% CI: 38, 47). A quadratic effect of survey day on σ was supported (Supporting information), where σ maximized on 10 May (Supporting information); on this date, σ was estimated to be 58 (95% CI: 38, 86). We retained σ = 58 to inform subsequent simulation scenarios for point-based HDS models with 'high' detection probability (Table 1). Average σ for transect-based surveys was estimated to be 42 (95% CI: 37, 49; Table 1). We again found support for survey day to have a nonlinear quadratic relationship with σ, resulting in σ being highest (51; 95% CI: 42, 61) on 9 May (Supporting information). We retained σ = 51 to inform subsequent simulation scenarios for transect-based HDS models with 'high' detection probability (Table 1).

Time-removal HDS probability of detection – Average σ (for a half-normal detection function) was estimated to be 42 (95% CI: 38, 47) and availability (ϕ) was calculated to be 0.65 with p_a estimated to be 0.23 (95% CI: 0.15, 0.34; Table 1). We found support for a quadratic relationship between survey day and availability, with availability highest on 12 May (Supporting information). Under ideal survey conditions, our high estimate of probability of availability (ϕ) was 0.89, p_a was estimated to be 0.43 (95% CI: 0.16, 0.69), and σ for the half-normal detection function was 48 (95% CI: 41, 55; Table 1).

Simulations evaluating different statistical estimators and survey effort

Overall, N-mixture models, HDS for point-counts and transects, and time-removal HDS produced unbiased estimates for both abundance and detection parameters with sufficient survey effort; 95% CrI of the bias included 0 for both λ (Supporting information), and detection parameters p, p_a, and σ (Fig. 3). As a result, estimates of total population abundance (total N) were also unbiased for all hierarchical models (Fig. 4), whereas predictions of λ and total N were always downward biased for the naïve model (Fig. 4, Supporting information). Precision of estimates of total N, however, varied substantially by survey protocol for each analytical method. Generally, estimated precision of total N increased with the number of sites and the number of replicate surveys per site (Fig. 5, 6). For all models, the number of sites needing to be surveyed and total number of point-count surveys needing to be conducted decreased as p or λ increased, (Fig. 5, 6, Supporting information). The required number of survey sites to produce acceptable precision (CV ≤ 15%) was generally logistically unachievable (≥ 800 survey points in the best-case scenario and ≥ 4200 in the worse case scenario considered) for point-count based HDS (Table 3, Fig. 5).

The survey protocol that required the fewest total point-counts (number of sites × number of visits) conducted for all scenarios was a protocol for visiting sites 4 times and analyzing the count data using N-mixture models (Table 3, Fig. 5). Depending on the abundance–detection scenario, the number of sites necessary varied between 60 and 490 (Table 3). The least number of surveys required occurred for scenarios with both high estimates of detection parameters and high λ (Table 3, Fig. 5, 6). For example, for point-counts, under simulated conditions of high λ (λ = 0.31) and average detection (for N-mixture $\bar{p} = 0.37$, for HDS σ = 43, for time-removal HDS σ = 42, availability = 0.65), 170 survey sites would need to be surveyed 4 times to obtain acceptably precise (CV ≤ 15%) estimates of population size evaluated using an N-mixture model, 1090 sites using hierarchical distance sampling and > 6000 sites for time-removal HDS (Table 3). In contrast, for point-counts, under a high abundance λ (λ = 0.31) and high detection (for N-mixture $\bar{p} = 0.57$, for HDS model σ = 58, for time-removal HDS σ = 48, availability = 0.89), only 60 survey sites would need to be surveyed 4 times using an N-mixture model, 800 sites for HDS model for point-counts, and 1390 sites for time-removal HDS for recovery of accurate and suitably precise estimates of abundance (Table 3).

For line-transects survey protocols, the number of transects that needed to be visited was lower for longer transects (5000 m) than shorter transects (2681 m; Fig. 6, Supporting information). Under average abundance (5.73 dusky grouse per km^2) and high detection probability (σ = 51), for example, 20 5000-m long transects were required to yield CV ≤ 15% for abundance, whereas 35 transects were needed when their length was 2681 m (Table 3). The number of required transects declined as the probability of detection increased (Fig. 6, Supporting information).

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

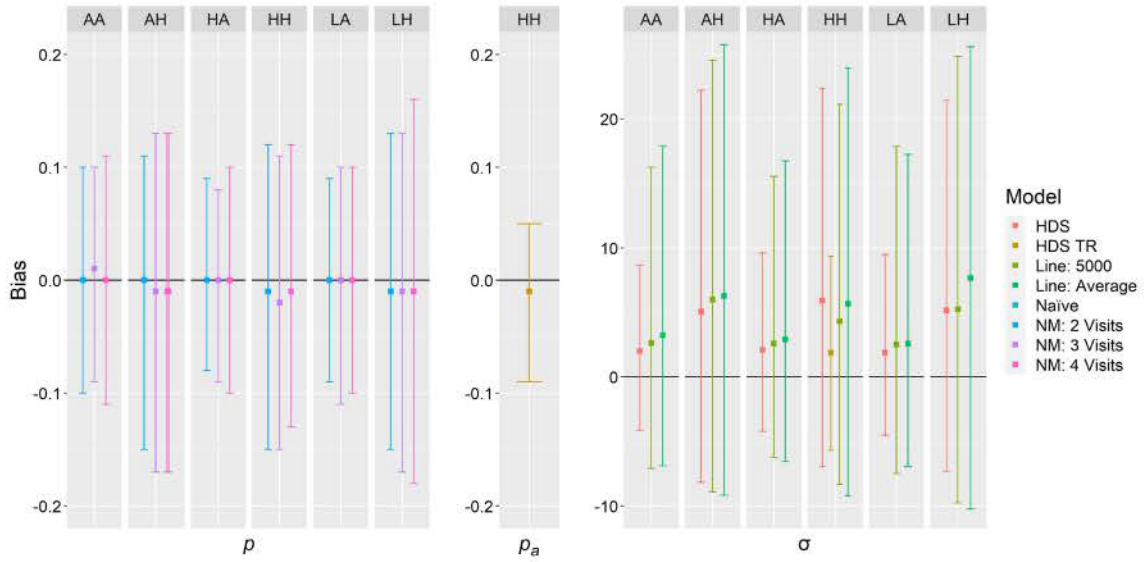


Figure 3. Bias with 95% credible intervals (CrI) for p (probability of detection; left), p_a (probability of availability; middle), σ (sigma; right) over 500 simulations under 6 scenarios for protocols that had the desired level of precisions (CV < 15%) for 8 different models (Table 3). If the 95% CrI overlap 0, the estimates are considered unbiased. Scenarios are HA (high abundance, average detection), AA (average abundance, average detection), LA (low abundance, average detection), HH (high abundance, high detection), AH (average abundance, high detection), and LH (low abundance, high detection). Models are HDS (hierarchical distance sampling for point-counts), HDS TR (time-removal HDS), Line: 5000 (hierarchical distance sampling for line transects 5000 m in length), Line: average (hierarchical distance sampling for line transects of average (2681 m) length), and NM with varying visits (N-mixture model with either 2, 3, or 4 visits).

In all cases, point-count protocols that included 4 replicated counts conducted in a single morning at 6 points per route resulted in lower necessary sampling effort than transect-based protocols. For example, under the average abundance – high detection scenario, 14 6-point survey routes (~ 2500 m in length) are necessary to produce unbiased and precise estimates using the point-count protocol, while 35 and 20 transects are needed for transect lengths of 2681 and 5000

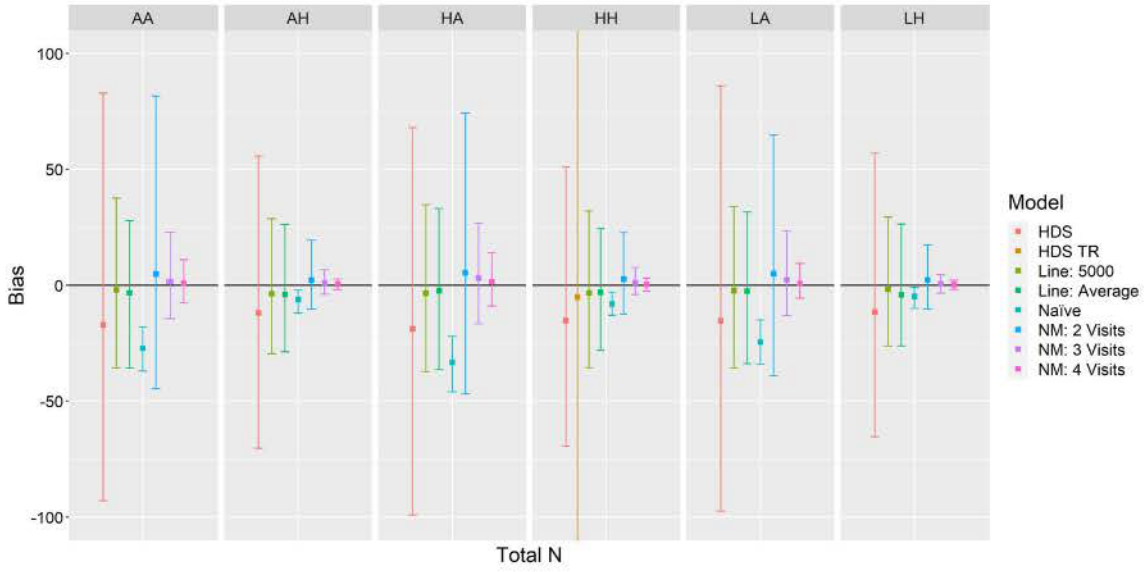


Figure 4. Bias with 95% credible intervals (CrI) for total population size (total N) over 500 simulations under 6 scenarios for protocols that had the desired level of precisions (CV < 15%) for 8 different models (Table 3). If the 95% CrI overlap 0, the estimates are considered unbiased. Scenarios are HA (high abundance, average detection), AA (average abundance, average detection), LA (low abundance, average detection), HH (high abundance, high detection), AH (average abundance, high detection), and LH (low abundance, high detection). Models are HDS (hierarchical distance sampling for point-counts), HDS TR (time-removal HDS), Line: 5000 (hierarchical distance sampling for line transects 5000 m in length), Line: average (hierarchical distance sampling for line transects of average (2681 m) length), and NM with varying visits (N-mixture model with either 2, 3, or 4 visits).

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

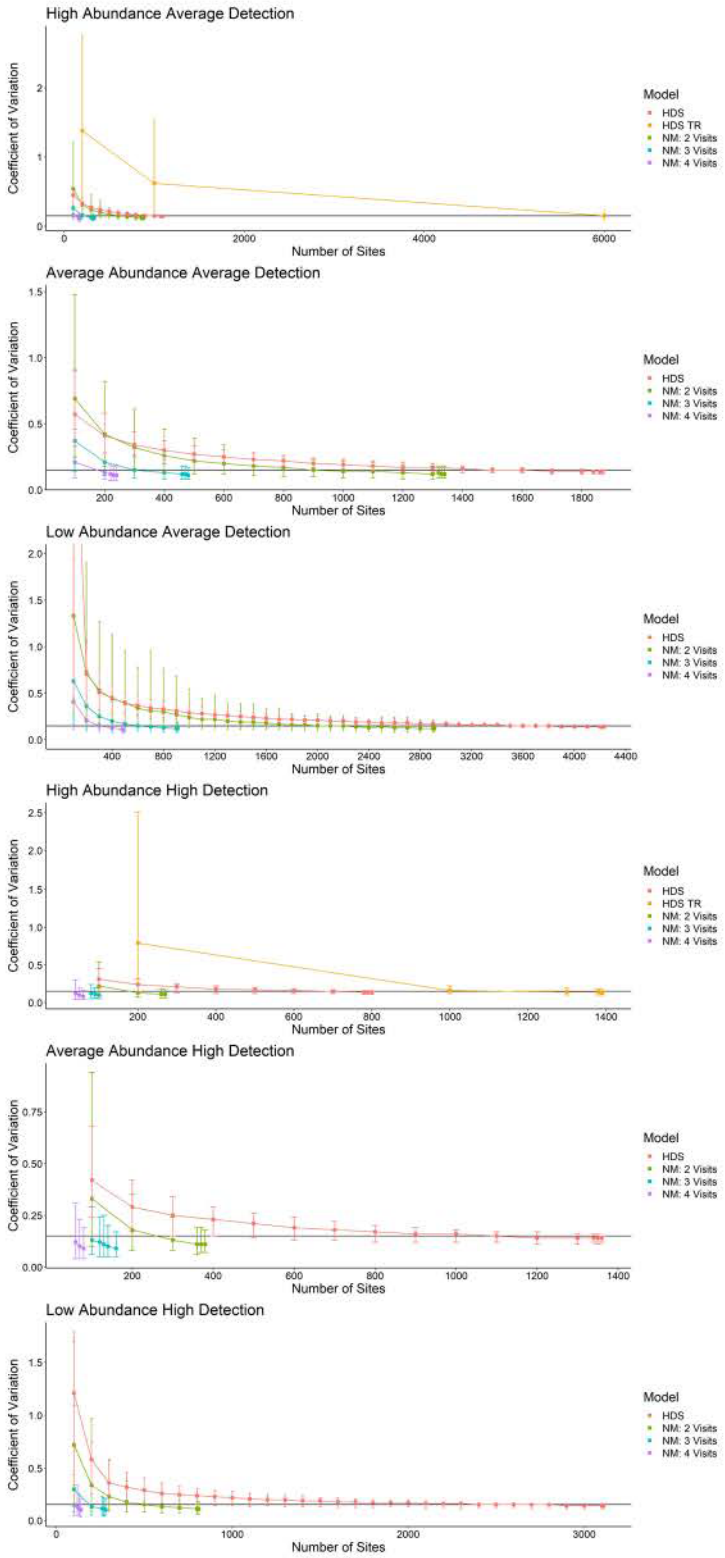


Figure 5. Coefficient of variation (CV) with 95% credible intervals (CrI) for estimates of population size (total N) over 500 simulations under different point-count protocols (i.e. number of sites, visits) under 6 different scenarios with varying abundance (high, average, low) and detection (high, average) for 3 different model types: HDS (hierarchical distance sampling), HDS TR (time-removal HDS), and NM (N-mixture model) where protocols with 2, 3, or 4 visits are evaluated. The horizontal line represents a CV of 0.15, under which a protocol/model combination meets the desired level of precision for estimates of total N.

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

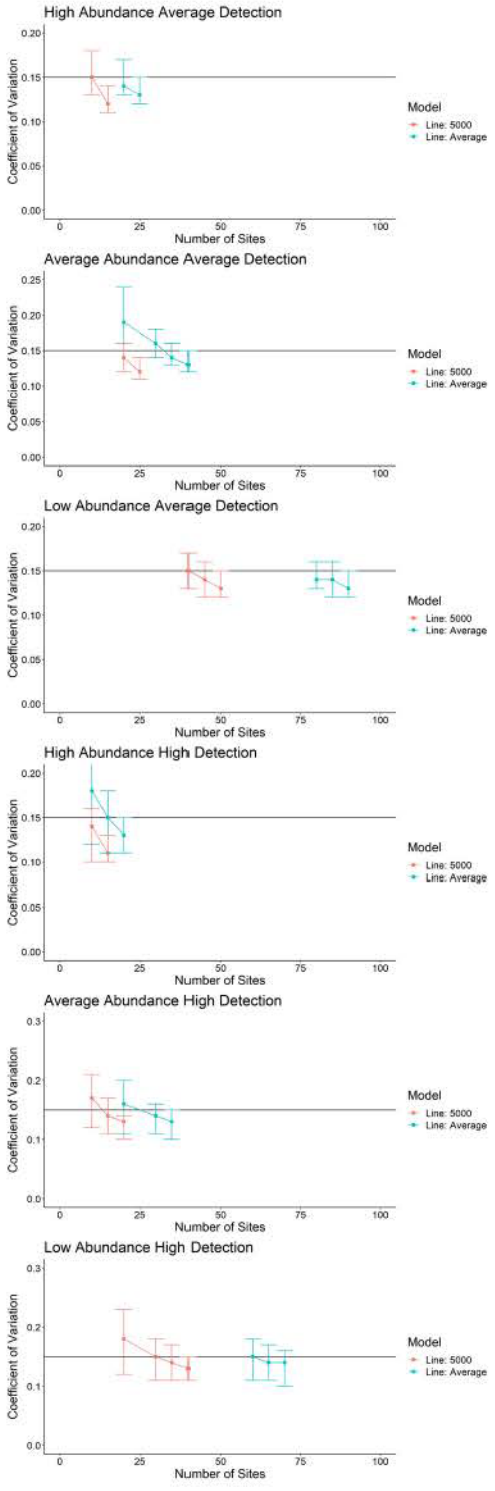


Figure 6. Coefficient of variation (CV) with 95% credible intervals (CrI) for estimates of population size (total N) over 500 simulations for different transect protocols (i.e. number of sites) evaluated using hierarchical distance sampling for an average transect length 2681 m and 5000 m transect length under 6 different scenarios where probability of detection (average, high) and abundance varied (high, average, low). The horizontal line represents a CV of 0.15, under which a protocol/model combination meets the desired level of precision for estimates of total N.

m, respectively (Table 3). Under average detectability scenarios, when conducting point-count protocols that included 4 replicated counts with 6 points located along a transect, the number of required transects was lower or equal to that of the line-transect method when line transects were 2681 m in length, and more than that when line transects were 5000 m in length. For example, under the average abundance, average detection scenario, 40 survey transects are required for the point-count N-mixture protocol and transects 2681 m in length while only 25 are needed for transects 5000 m in length (Table 3).

Power of protocols to detect population trends

Survey protocols that produced precise (CV ≤ 15%) annual estimates of abundance did not have power to detect a 1% annual decline in abundance over 10 years but did have power (≥ 80%) to detect a 3 and 5% annual decline over the same period (Table 4A). Over 5 years, our top performing sampling protocol and N-mixture model had power (≥ 80%) to detect a 10% annual decline (Table 4A). Estimated slopes (β) describing the change in mean annual abundance over time accurately tracked simulated rates of population declines over all time periods (Table 4B). However, 95% CI of slopes included 0 (i.e. was not statistically different than 0) when the true annual decline was 1–5% over a 3-year time period (Table 4B). Using our best performing survey protocol, estimated rates of population change were significantly < 0 (i.e. accurately predicted population declines) over a 5-year assessment period when the true rate of population declines > 5% year (Table 4B). A negative trend in abundance was predicted ≥ 70% of the time for all combinations of 3, 5, and 10% annual declines, and after 5 years for a 1% annual decline (Table 4C). On average, the difference between the annual estimated population trend and the true trend was small; difference between real and estimated slopes was < 0.003 on average (Table 4D).

Discussion

Our study provides the foundation for a rigorous population monitoring program for dusky grouse in Montana and, more broadly, illustrates an intuitive adaptive framework for the development of unbiased monitoring programs for other species of forest grouse. By iteratively evaluating various sampling designs, survey protocols, and analytical methods through simulations, we identified survey protocols and analytical approaches that are both logistically feasible and yield unbiased, precise (CV ≤ 15%) abundance estimates, despite variable and imperfect detection probabilities. Our rigorous approach involved assessing hundreds of potential survey protocols across six realistic scenarios based on empirical estimates of abundance and detection probabilities from three years of field surveys of dusky grouse. We determined that for achievable levels of survey effort, point-count protocols with four site visits analyzed using N-mixture models and line-transect surveys analyzed using hierarchical distance sampling

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3. Summarized from the simulation results, the number of visits, sites to be surveyed, total number of point-counts to be conducted, and the potential number of transects (6 points on each transect) for providing robust population estimates using N-mixture models (point-counts), HDS (point-counts and transects), and time-removal HDS (point-counts) under 6 different scenarios: HA = high abundance, average detection, AA = average abundance, average detection, LA = low abundance, average detection, HH = high abundance, high detection, AH = average abundance, high detection, LH = low abundance, high detection. Simulations were only conducted under 2 scenarios for the time-removal HDS. The transect length evaluated was 2681m (the average transect length for 2020 and 2021).

Scenario	Model	No. of visits	No. of sites	No. of point counts	Transect (6 pts)
HA	N-mixture point count	4	170	680	29
	HDS point count	1	1090	1090	182
	Time-removal HDS point count	1	> 6000	> 6000	> 1000
	HDS transect: 2681 m	1	25	NA	25
	HDS transect: 5000 m	1	15	NA	15
AA	N-mixture point count	4	240	960	40
	HDS point count	1	1870	1870	312
	Time-removal HDS point count	1	NA	NA	NA
	HDS transect: 2681 m	1	40	NA	40
	HDS transect: 5000 m	1	25	NA	25
LA	N-mixture point count	4	490	1960	82
	HDS point count	1	4230	4230	705
	Time-removal HDS point count	1	NA	NA	NA
	HDS transect: 2681 m	1	90	NA	90
	HDS transect: 5000 m	1	50	NA	50
HH	N-mixture point count	4	60	240	10
	HDS point count	1	800	800	134
	Time-removal HDS point count	1	1390	1390	232
	HDS transect: 2681 m	1	20	NA	20
	HDS transect: 5000 m	1	15	NA	15
AH	N-mixture point count	4	80	320	14
	HDS point count	1	1360	1360	227
	Time-removal HDS point count	1	NA	NA	NA
	HDS transect: 2681 m	1	35	NA	35
	HDS transect: 5000 m	1	20	NA	20
LH	N-mixture point count	4	140	560	24
	HDS point count	1	3110	3110	519
	Time-removal HDS point count	1	NA	NA	NA
	HDS transect: 2681 m	1	70	NA	70
	HDS transect: 5000 m	1	40	NA	40

(HDS) offer the necessary accuracy and precision for effective population monitoring of dusky grouse in Montana.

Our results agree with other recent research demonstrating that hierarchical statistical models that incorporate imperfect and variable detection probabilities can produce unbiased and reasonable precise estimates of abundance for unmarked populations of grouse (McCaffery et al. 2016, Coates et al. 2019). However, results of our simulations revealed that, under logistically feasible survey effort and realistic conditions, only N-mixture models fitted to point-count data, and hierarchical distance sampling (HDS) fitted to transect-level data were suitable for unbiased monitoring of dusky grouse populations. A detection-naïve model of simple counts always resulted in downward-biased estimates of abundance and could yield incorrect inference about population trends because the probability of detecting a grouse can vary over space and time due to site- and survey-specific conditions (Thompson 2002, Leipold et al. 2025). Time-removal HDS and HDS fitted to point-count data produced accurate and precise estimates of abundance, but only under unachievable levels of survey effort (under a high abundance, average detection scenario > 6000 surveys and > 1000 surveys, respectively).

Performance of N-mixture models

Generally, results of our simulation scenarios indicated that N-mixture models outperformed HDS and time-removal HDS models under logistically-feasible survey effort. Estimates of local and total abundance from N-mixture models with ≥ 2 replicate visits were always substantially less biased and more precise than those from distance sampling models applied to point-counts. Moreover, the amount of survey effort necessary to achieve acceptable levels of precision (CV ≤ 15%) was significantly lower for N-mixture models with achievable survey effort, regardless of abundance or expected detection probability. Under moderate expectations of average local abundance ($\lambda = 0.18$ grouse per survey area) and detection probability ($p = 0.37$), surveying 260 survey points (40 survey routes) 4 times each resulted in accurate and precise estimates of abundance.

Our results were generally consistent with simulation results of N-mixture models for other species and simulation studies in which increasing the number of plots and visits increased the precision of estimates (McIntyre et al. 2012, Yamaura 2013). As the number of visits increased from 2 to 4, for example, the requisite number of sites required to reach

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 4. Predicted (A) power, (B) slopes of predicted annual population trend, (C) percent of simulation trials where estimates slopes indicated a declining population ($\beta < 0$), and (D) differences between estimated and true slopes (β) for a dusky grouse protocol where 80 sites are surveyed 4 times per year and abundances are estimated using an N-mixture model. Sufficient power to detect a population decline was 80%. A 10% decline over 10 years was not evaluated as sufficient power was reached after a period of 5 years. CI: confidence interval.

Annual Decline	3 years	5 years	10 years
(A) Power to detect specified annual population decline over specified time period			
1%	8.2	14.4	39.2
3%	19.0	40.6	80.8
5%	26.8	61.0	94.6
10%	49.6	87.2	
(B) Mean estimated slopes (95% CI) for annual trend in specified population declines			
1%	−1.13 (−8.29, 5.94)	−1.18 (−6.13, 2.78)	−1.08 (−3.75, 0.84)
3%	−3.58 (−12.82, 4.38)	−3.17 (−9.32, 1.63)	−3.21 (−7.81, −0.1)
5%	−5.30 (−16.25, 3.68)	−5.39 (−13.89, −0.08)	−5.24 (−11.1, −1.16)
10%	−11.08 (−26.6, −0.11)	−10.73 (−21.34, −0.07)	
(C) Percent of estimated slopes $\beta < 0$, indicating proper inference of population declines			
1%	60.6	71.0	80.8
3%	78.8	88.4	97.6
5%	84.4	97.8	99.0
10%	97.4	98.2	
(D) Difference (95% CI) between slopes (β) of estimated population decline and the true simulated population decline. The 95% CIs that overlap 0 indicate similar slopes.			
1%	−0.0005 (−0.06, 0.07)	−0.0012 (−0.04, 0.04)	−0.0003 (−0.01, 0.01)
3%	−0.0030 (−0.07, 0.06)	−0.0001 (−0.04, 0.04)	0.0007 (−0.02, 0.017)
5%	0.0018 (−0.06, 0.07)	0.0009 (−0.03, 0.37)	0.0011 (−0.02, 0.02)
10%	−0.0007 (−0.08, 0.07)	0.0014 (−0.04, 0.04)	

the desired level of precision decreased from 900 to 170 under the high abundance average detection scenario (Supporting information). Similar to another study, we found that large improvements in estimator precision occurred as probability of detection increased (McIntyre et al. 2012). We also found that our models performed well at low local abundances, with estimates more variable under lower local abundance and probability of detection (Yamaura 2013).

Performance of HDS models

Our simulations revealed that hierarchical distance sampling (HDS) only provided unbiased and reasonably precise estimates under feasible levels of survey effort when the observational unit was the entire survey transect. HDS applied to point-count data, even for areas with high densities of dusky grouse, did not produce acceptable levels of precision in abundance estimates unless the survey effort was prohibitive (e.g. > 1800 survey points for average detection, average abundance and > 1300 for high detection, average abundance).

Results for the HDS analyses of simulated data were similar to N-mixture analyses; the requisite number of sites visited to achieve our desired level of precision decreased as local abundance or probability of detection increased. Despite this, point-count based HDS and time-removal HDS models required a prohibitively large number of survey sites that were not logistically feasible, even for point-count scenarios with relatively high abundance and detection probability. At least 60 to 100 grouse observations are needed for accurate estimates of abundance using distance sampling, suggesting limited utility of distance sampling for point-count survey protocols of dusky grouse. When mean local abundance was average (e.g. 0.18 dusky grouse per point-count) and detection probability was high, the number of sites necessary to obtain the necessary sample size to be able to obtain a precise abundance estimate is prohibitively large (1360 survey points). Other simulations have shown that the precision for abundance estimates from time-removal HDS is low when probability of availability is low (≤ 0.4; Amundson et al. 2014). While our average probability of availability (0.65) was higher than 0.4, low grouse densities required prohibitively large number of survey points to be visited (> 6000) to achieve our desired level of precision. In short, our simulation results indicate that distance sampling models applied to dusky grouse counts collected at point-counts will not provide suitably accurate and precise estimates of relative population size for management.

Line transects analyzed using HDS required more achievable survey effort than did point-counts. We examined the utility of transects of two different lengths and found that longer transects required fewer sites be visited than the shorter transects, assuming dusky grouse are distributed randomly in relation to transect length. Nevertheless, longer transects take more time to survey and may have higher likelihood that parts of the transect may be inaccessible, especially early in the season due to snowpack. In comparison to point-counts analyzed using N-mixture models, the shorter line-transects analyzed using HDS required a similar or greater number of transects visited. Due to the large credible intervals when examining bias for the abundance and detection parameters for line-transects analyzed using HDS in addition to needing greater survey effort, we recommend point-counts analyzed using N-mixture models over line-transects analyzed using HDS.

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Consequences of convenience sampling and quasi-independent count data

Convenience-based sampling along low-use roads and hiking trails yielded similar estimates of dusky grouse abundance as randomly located off-trail survey points, but with significantly less effort and time investment. Moreover, it was often impractical to complete all off-trail surveys within the morning timeframe that maximized detection probabilities needed for precise abundance estimates. Surveys conducted along low-use roads and trails could be completed within a 4-hour period after sunrise, when detection probability was highest (Leipold et al. 2025). However, the time and effort required to access randomly selected off-trail locations were prohibitive in the mountainous dusky grouse habitats of Montana. Increased snowpack and downed trees made navigating off-trail terrain substantially more challenging than navigating roads and trails. Additionally, reaching the off-trail survey areas often involved 1–2 hours of pre-dawn hiking, increasing safety risks in rugged terrain inhabited by large predators, such as grizzly bears *Ursus arctos*. Consequently, we recommend limiting dusky grouse population monitoring to roads and trails with low traffic.

Our recommended analytical method for estimating dusky grouse abundance, N-mixture models, requires 4 replicated surveys be conducted at each survey point within a period of population closure. Accessing at least 360 survey points in mountainous habitats on 4 different occasions presents logistical challenges that substantially reduce either the number survey points or the number of replicate surveys that can be conducted, both of which result in decreases in accuracy and precision of abundance estimates. Our recommended approach to conduct four replicated surveys at points along a route in the same morning directly addresses the critical population closure assumption of N-mixture models while minimizing the time and effort necessary to access survey points. However, conducting replicate visits within the same morning may violate an implicit assumption that visits represent independent samples. We evaluated the effects of violating this assumption through additional simulation analyses (Supporting information). We found that estimates of abundance were: 1) unbiased when the average probability of detection was ≥ 0.57, and 2) were only modestly downward biased (≤ 0.04 fewer grouse per survey area) under scenarios where detection probability was low ($p = 0.37$). Hence, when surveys are conducted under optimal conditions that maximize the probability of detection (Leipold et al. 2025), replicate visits can be conducted within a single morning without biasing estimates of abundance, negating logistical constraints that may limit the use of N-mixture models.

Power to detect changes in population size

Our recommended protocol derived from the average abundance scenario (point-counts where 80 sites or 14 transects with 6 points per transect, are visited four times and evaluated using N-mixture models) had high power ($\geq 80\%$) to detect average population declines of 3, 5, and 10% over 5–10-year periods, which was lower than expected given CV of annual abundance estimates were $< 15\%$. Nevertheless, we found that the estimated abundance trends were similar to the target trends, and close to the real trends estimated using the true simulated population size, suggesting that while there may be some uncertainty associated with the estimated trends, our protocols may be sufficient for long-term monitoring and able to detect small changes in population size in as little as 3 years. In addition, as the monitoring period increases (> 5 years), the power to detect small changes increases indicating that our protocols are appropriate for long-term monitoring of dusky grouse populations.

Conclusions for dusky grouse monitoring

A useful population monitoring program will produce unbiased estimates or indices of annual population abundance while minimizing survey efforts and associated costs. We used simulations to evaluate the performance (i.e. bias and precision) of dozens of combinations of potential survey protocols and statistical estimators under a variety of realistic survey situations (e.g. low–high abundance; low–high detectability) to identify combinations of logistically-feasible survey and analytical methods that produce unbiased and precise estimates of abundance of dusky grouse. We recommend conducting spring point-count surveys during the month of May using electronic playback of female calls to increase detection probability (Leipold et al. 2025). Survey routes along existing low-use roads and hiking trails produced similar inferences about population size as randomly located off-trail survey routes at much lower effort, saving observers much time and increasing the number of surveys that can be conducted. Because estimates of abundance were not influenced by route type, we recommend N-mixture models be applied to count data collected at survey points located along low-use or seasonal-use roads and trails in dusky grouse habitat. Each point along a survey route should be spaced ≥ 400 apart and surveyed 4 times (e.g. 4-min observations) within a single morning, preferably within 4 hours on days with no precipitation, limited cloud cover, no wind, and limited background noise (Leipold et al. 2025). To achieve adequate precision for population management (CV $\leq 15\%$), at least 14 survey routes composed of 6 points (80 points and 320 total point-counts) should be surveyed for any jurisdiction requiring population inference (e.g. MFWP region). With additional survey effort (21 additional survey transects), HDS applied to count data summed over ≥ 2681-m survey transects also produced accurate and reasonably precise estimates of dusky grouse abundance.

Recommendations for creating population monitoring programs

Precise and unbiased estimates of population size are important for effective conservation and management of all species. For cryptic forest grouse species that often have low densities and frequently low probability of detection, simulations are an important tool for 1) evaluating and comparing the performance of statistical estimators of population size, and 2) determining an appropriate and feasible protocol for

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

population monitoring that leads to effective management. In addition to logistical feasibility of using a particular field survey protocol, we must also evaluate whether the assumptions for different statistical estimators can be met for a given species. For example, distance sampling is often used for estimating abundance of passerines surveyed with point-counts, but we found it to not be an effective method for monitoring dusky grouse populations due to the effort needed to obtain the necessary sample size of dusky grouse observations. For other forest grouse, such as ruffed grouse *Bonasa umbellus* whose display calls can be heard over much greater distances than dusky grouse, obtaining the necessary number of observations for HDS may be more achievable, although accurate distance measurements may be more difficult to obtain.

Population densities and detection also often vary across a species' distribution, implying that a population monitoring protocol (i.e. number of visits and sites) appropriate for one area may not be appropriate for monitoring the species in a different part of its range. Using an iterative process where simulation scenarios are refined as more region-specific information becomes available is an important part of developing a monitoring program for specific areas. Our initial protocols were developed based on abundance and detection data from Utah, but as we obtained empirical estimates of abundance and detection for our study population, our recommended protocols changed. Our study demonstrates the importance of baseline data specific to the study area for informing simulation scenarios.

Ensuring that population monitoring programs are well planned, able to make efficient use of limited resources, and achieve target goals, is critical, and knowing what statistical estimator, protocol, and survey conditions that are most efficient for a particular study are best informed by analyses of simulated datasets. Simulations allow for 1) the comparison in performance of different survey protocols and statistical estimators under variable abundance or observation error, and 2) evaluating the potential effects of model violations on proper population inference. Time, funding, and personnel are always limiting to effective population monitoring programs and employing simulations can help identify most appropriate methods without wasting field resources. Our study emphasizes the importance of simulations and illustrates an effective process for developing an effective population monitoring program for forest grouse species.

Acknowledgements – We thank all the technicians who assisted in collecting data for this project: Teagan Byroth, Lara Macon, David Forbes, Rob Francisco, Mikaela Howie, Chris Smith, Luke Johnson, Lulu McMahon, Thomas Hull, Kellan Karch, Haley Burger, and Alisha Gill. In addition we thank all the volunteers and Montana Fish, Wildlife and Parks biologists who also assisted with data collection.

Funding – This work was funded by Montana Fish, Wildlife and Parks, Montana Agricultural Experiment Station at Montana State University, and Federal Aid in Wildlife Restoration (Pittman-Robertson) grant no. W-162-R-0.

Conflict of interest – The authors declare no conflict of interest.

Author contributions

Elizabeth A. Leipold: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal). **Claire Gower**: Conceptualization (equal); Funding acquisition (equal); Resources (equal); Writing – review and editing (equal). **Lance B. McNew**: Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/wlb3.01506.

Data availability statement

Data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.m63xsj4dm (Leipold et al. 2026).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Amundson, C. L., Royle, J. A. and Handel, C. M. 2014. A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. – Auk 131: 476–494.

Andersen, E. M. and Steidl, R. J. 2019. Power to detect trends in abundance within a distance sampling framework. – J. Appl. Ecol. 57: 344–353.

Barker, R. J., Schofield, M. R., Link, W. A. and Sauer, J. R. 2018. On the reliability of N-mixture models for count data. – Biometrics 74: 369–377.

Basile, M., Balestrieri, R., Posillico, M., Mancinelli, A., Altea, T. and Matteucci, G. 2016. Measuring bird abundance: a comparison of methodologies between capture/recapture and audio-visual surveys. – Avocetta 40: 55–61.

Bendell, J. F. and Elliot, P. W. 1967, Behavior and the regulation of numbers in blue grouse. Canadian Wildlife Service Report Series: 4. – Department of Indian Affairs and Northern Development.

Bernales, H. H., Robinson, J. D. and Stringham, B. 2017. Utah upland game annual report 2016–17. – State of Utah Department of Natural Resources.

Betts, M. G., Mitchell, D., Diamond, A. W. and Bêty, J. 2007. Uneven rates of landscape change as a source of bias in roadside wildlife surveys. – J. Wildl. Manage. 71: 2266–2273.

Blackford, J. L. 1958. Territoriality and breeding behavior of a population of blue grouse in Montana. – Condor 60: 145–158.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. and Thomas, L. 2001. Introduction to distance sampling: estimating abundance of biological populations. – Oxford Univ. Press.

19033220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Buckland, S. T., Marsden, S. J. and Green, R. E. 2008. Estimating bird abundance: making methods work. – Bird Conserv. Int. 18: S91–S108.

Buckland, S. T., Rexstad, E. A., Marques, T. A. and Oedekoven, C. S. 2015. Distance sampling: methods and applications. – Springer.

Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. – Springer.

Chan-McLeod, A. and Bunnell, F. L. 2004. Potential approaches to integrating silvicultural control of mountain pine beetle with wildlife and sustainable managment objectives. – In: Shore, T. L., Brooks, J. E. and Stone, J. E. (eds), Mountain pine beetle symposium: challenges and solutions, Keowna, British Columbia, Canada, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Information report BC-X-399, pp. 267–277.

Coates, P. S., Wann, G. T., Gillette, G. L., Ricca, M. A., Prochazka, B. G., Severson, J. P., Andrle, K. M., Espinosa, S. P., Casazza, M. L. and Delehanty, D. J. 2019. Estimating sightability of greater sage-grouse at leks using an aerial infrared system and N-mixture models. – Wildl. Biol. 2019: wlb.00552.

Dinkins, J. B., Lawson, K. J. and Beck, J. L. 2021. Influence of environmental change, harvest exposure, and human disturbance on population trends of greater sage-grouse. – PLoS One 16: e0257198.

Doser, J. W., Weed, A. S., Zipkin, E. F., Miller, K. M. and Finley, A. O. 2021. Trends in bird abundance differ among protected forests but not bird guilds. – Ecol. Appl. 31: e02377.

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

Espinosa, S. P., Scott, M., Donham, T. and Kimble, S. 2018. Small game status: 2018: harvest data and population status report. – Nevada Department of Wildlife.

Farnsworth, G. L., Nichols, J. D., Sauer, J. R., Fancy, S. G., Pollock, K. H., Shriner, S. A. and Simons, T. R. 2005. Statistical approaches to the analysis of point count data: a little extra information can go a long way. Technical Report PSW-GRT-191. – USDA Forest Service.

Fiske, I. and Chandler, R. 2011. Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. – J. Stat. Software 43: 1–23.

Ficetola, G. F., Barzaghi, B., Melotto, A., Muraro, M., Lunghi, E., Canedoli, C., Lo Parrino, E. L., Nanni, V., Silva-Rocha, I., Urso, A., Carretero, M. A., Salvi, D., Scali, S., Scarì, G., Pennati, R., Andreone, F. and Manenti, R. 2018. N-mixture models reliably estimate the abundance of small vertebrates. – Sci. Rep. 8: 10357.

Field, S. A., O'Connor, P. J., Tyre, A. J. and Possingham, H. P. 2007. Making monitoring meaningful. – Austral Ecol. 32: 485–491.

Franceschi, S., Nelli, L., Pisani, C., Franzoi, A., Fattorini, L. and Meriggi, A. 2014. A Monte Carlo appraisal of plot and distance sampling for surveys of black grouse and rock ptarmigan populations in alpine protected areas. – J. Wildl. Manage. 78: 359–368.

Furnas, B. J., Newton, D. S., Capehart, G. D. and Barrows, C. W. 2019. Hierarchical distance sampling to estimate population sizes of common lizards across a desert ecoregion. – Ecol. Evol. 9: 3046–3058.

Gates, E. 2019. Annual report of small game, migratory birds, furbearer, wild turkey, and falconry harvest for the biological year 1 June 2018–31 May 2019. – Wyoming Game and Fish Deparment.

Gelman, A. and Rubin, D. B. 1992. Inference from iterative simulation using multiple sequences. – Stat. Sci. 7: 457–511.

Gibbs, J. P. and Melvin, S. M. 1997. Power to detect trends in waterbird abundance with call-response surveys. – J. Wildl. Manage. 61: 1262–1267.

Gibbs, J. P., Snell, H. L. and Causton, C. E. 1999. Effective monitoring for adaptive wildlife management: lessons from the Galápagos Islands. – J. Wildl. Manage. 63: 1055–1065.

Hamm, C. A. 2013. Estimating abundance of the federally endangered Mitchell's satyr butterfly using hierarchical distance sampling. – Insect Conserv. Diversity 6: 619–626.

Hanowski, J. M. and Niemi, G. J. 1995. A comparison of on- and off-road bird counts: do you need to go off road to count birds accurately? – J. Field Ornithol. 66: 469–483.

Harju, H. J. 1974. An analysis of some aspects of the ecology of dusky grouse. – PhD thesis, University of Wyoming, USA.

Jakob, C., Ponce-Boutin, F. and Besnard, A. 2014. Coping with heterogeneity to detect species on a large scale: N-mixture modeling applied to red-legged partridge abundance. – J. Wildl. Manage. 78: 540–549.

Joseph, L. N., Elkin, C., Martin, T. G. and Possinghami, H. P. 2009. Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. – Ecol. Appl. 19: 631–642.

Katayama, N., Odaya, Y., Amano, T. and Yoshida, H. 2020. Spatial and temporal associations between fallow fields and greater painted snipe density in Japanese rice paddy landscapes. – Agricult. Ecosyst. Environ. 295: 106892.

Kellner, K. 2019. jagsUI: a wrapper around 'rjags' to streamline 'JAGS' analyses. – R package ver. 1.5.1, https://CRAN.R-project.org/package=jagsUI.

Kéry, M. 2018. Identifiability in N-mixture models: a large-scale screening test with bird data. – Ecology 99: 281–288.

Kéry, M. and Royle, J. A. 2016. Applied hierarchical modeling in ecology: analysis of distribution, abundance, and species richness in R and BUGS. – Academic Press.

Kéry, M. and Schaub, M. 2012. Bayesian population analysis using WinBUGS a hierarchical perspective. – Elsevier.

Kidwai, Z., Jimenez, J., Louw, C. J., Nel, H. P. and Marshal, J. P. 2019. Using N-mixture models to estimate abundance and temporal trends of black rhinoceros (*Diceros bicornis* L.) populations from aerial counts. – Global Ecol. Conserv. 19: e00687.

Knape, J., Arlt, D., Barraquand, F., Berg, Å., Chevalier, M., Pärt, T., Ruete, A. and Żmihorski, M. 2018. Sensitivity of binomial N-mixture models to overdispersion: the importance of assessing model fit. – Methods Ecol. Evol. 9: 2102–2114.

LandFire 2016. Data from: LANDFIRE Exisiting Vegetation Type Layer – LANDFIRE 2016 Remap (LF 2.0.0). – http://landfire.cr.usgs.gov/viewer.

Leipold, E. A., Gower, C. N. and McNew, L. B. 2024. Using an ensemble approach to predict habitat of dusky grouse (*Dendragapus obscurus*) in Montana, USA. – Avian Conserv. Ecol. 19: 7.

Leipold, E. A., Gower, C. N. and McNew, L. B. 2025. Maximizing the detection probabilities of dusky grouse for population monitoring. – Wildl. Biol. 2025: e01411.

Leipold, E. A., Gower, C. and McNew, L. B. 2026. Data from: Development of optimized methods for unbiased dusky grouse

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

population monitoring using real and simulated data. – Dryad Digital Repository, https://doi.org/10.5061/dryad.m63xsj4dm.
Lewis, W. B., Chandler, R. B., Delancey, C. D., Rushton, E., Wann, G. T., McConnell, M. D. and Martin, J. A. 2022. Abundance and distribution of ruffed grouse *Bonasa umbellus* at the southern periphery of the range. – Wildl. Biol. 2022: e01017.
Link, W. A., Schofield, M. R., Barker, R. J. and Sauer, J. R. 2018. On the robustness of N-mixture models. – Ecology 99: 1547–1551.
Martinka, R. R. 1972. Structural characteristics of blue grouse territories in southwestern Montana. – J. Wildl. Manage. 36: 498–510.
McCaffery, R., Nowak, J. J. and Lukacs, P. M. 2016. Improved analysis of lek count data using N-mixture models. – J. Wildl. Manage. 80: 1011–1021.
McIntyre, A. P., Jones, J. E., Lund, E. M., Waterstrat, F. T., Giovanini, J. N., Duke, S. D., Hayes, M. P., Quinn, T. and Kroll, A. J. 2012. Empirical and simulation evaluations of an abundance estimator using unmarked individuals of cryptic forest-dwelling taxa. – For. Ecol. Manage. 286: 129–136.
McNeil, D. J., Otto, C. R., Moser, E. L., Urban-Mead, K. R., King, D. E., Rodewald, A. D. and Larkin, J. L. 2018. Distance models as a tool for modeling detection probability and density of native bumblebees. – J. Appl. Entomol. 143: 225–235.
Montana Spatial Data Infrastructure (MSDI). 2017. Data from: Transportation framework. – Montana State Library, http://geoinfo.msl.mt.gov/msdi/transportation.
Mussehl, T. W. 1963. Small game research project, blue grouse population study report W-91-R-5. – Montana Fish Wildlife and Parks.
Neubauer, G. and Sikora, A. 2020. Abundance estimation from point counts when replication is spatially-intensive but temporally limited: comparing binomial N-mixture and hierarchical distance sampling models. – Ornis Fenn. 97: 131–148.
New Mexico Game and Fish 2020. 2019–2020 New Mexico small game voluntary harvest reports. – New Mexico Game and Fish.
Nichols, J. D., Thomas, L. and Conn, P. B. 2009. Inferences about landbird abundance from count data: recent advances and future directions. – In: Thomson, D. L., Cooch, E. G. and Conroy, M. J. (eds), Modeling demographic processes in marked populations. – Springer.
Pelren, E. C. and Crawford, J. A. 1999. Blue grouse nesting parameters and habitat associations in northeastern Oregon. – Great Basin Nat. 59: 368–373.
Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. – Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria, 20–22 March 2003, pp. 1–10.
Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L. and Sauer, J. R. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. – Environmetrics 13: 105–119.
Prism Climate Group 2024. Data from: PRISM Gridded Climate Data. – Oregon State Univ., https://prism.oregonstate.edu.
Reidy, J. L. and Thompson, F. R. 2021. Density and habitat associations of the woodhouse's scrub jay in central Texas. – Southwest. Nat. 66: 5–12.
Rogers, G. E. 1963. Blue grouse census and harvest in the United States and Canada. – J. Wildl. Manage. 27: 579–585.
Romano, A., Costa, A., Basile, M., Raimondi, R., Posillico, M., Scinti Roger, D., Crisci, A., Piraccini, R., Raia, P., Matteucci, G. and De Cinti, B. 2017. Conservation of salamanders in managed forests: methods and costs of monitoring abundance and habitat selection. – For. Ecol. Manage. 400: 12–18.
Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T. and Carter, M. F. 2002. Landbird counting techniques: current practices and an alternative. – Auk 119: 46–53.
Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated counts. – Biometrics 60: 108–115.
Rusk, J. P., Hernández, F., Arredondo, J. A., Hernández, F., Bryant, F. C., Hewitt, D. G., Redeker, E. J., Brennan, L. A. and Bingham, R. L. 2007. An evaluation of survey methods for estimating northern bobwhite abundance in southern Texas. – J. Wildl. Manage. 71: 1336–1343.
Sands, J. P. and Pope, M. D. 2010. A survey of galliform monitoring programs and methods in the United States and Canada. – Wildl. Biol. 16: 342–356.
Scott, M. J., Ramsey, F. L. and Kepler, C. B. 1981. Distance estimation as a variable in estimating bird numbers from vocalizations. – Stud. Avian Biol. 6: 334–340.
Sillett, T. S., Chandler, R. B., Royle, J. A., Kéry, M. and Morrison, S. A. 2012. Hierarchical distance-sampling models to estimate population size and habitat-specific abundance of an island endemic. – Ecol. Appl. 22: 1997–2006.
Simons, T. R., Alldredge, M. W., Pollock, K. H. and Wettroth, J. M. 2007. Experimental analysis of the auditory detection process on avian point counts. – Auk 124: 986–999.
Steidl, R. J., Conway, C. J. and Litt, A. R. 2013. Power to detect trends in abundance of secretive marsh birds: effects of species trais and sampling effort. – J. Wildl. Manage. 77: 445–453.
Sullins, D., Conway, W., Haukos, D. and Comer, C. 2019. Using pointing dogs and hierarchical models to evaluate American woodcock winter occupancy and densities. – Proceedings of the American woodcock symposium. 11: 154–167, https://doi.org/10.24926/AWS.0122.
Thompson, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. – Auk 119: 18–25.
Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R., Marques, T. A. and Burnham, K. P. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. – J. Appl. Ecol. 47: 5–14.
US Forest Service 2020a. Data from: Motor vehicle use map: roads. – FSGeodata Clearinghouse, https://data.fs.usda.gov/geodata/edw/datasets.php.
US Forest Service 2020b. Data from: Motor vehicle use map: trails. – FSGeodata Clearinghouse, https://data.fs.usda.gov/geodata/edw/datasets.php.
US Forest Service 2020c. Data from: National Forest System roads. – FSGeodata Clearinghouse, https://data.fs.usda.gov/geodata/edw/datasets.php.
US Forest Service 2020d. Data from: National Forest System trails. – FSGeodata Clearinghouse, https://data.fs.usda.gov/geodata/edw/datasets.php.
US Geological Survey 2017. Data from: 3DEP products and services: The National Map, 3D Elevation Program. – The National Map, https://nationalmap.gov/3DEP/3dep_prodserv.hmtl.
Warren, P. and Baines, D. 2011. Evaluation of the distance sampling technique to survey red grouse *Lagapus lagopus scoticus* on moors in northern England. – Wildl. Biol. 17: 135–142.
Weiser, E. L., Diffendorfer, J. E., López-Hoffman, L., Semmens, D. and Thogmartin, W. E. 2019. Consequences of ignoring spatial variation in population trend when conducting a power analysis. – Ecography 42: 836–844.

1903220x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wiest, W. A., Correll, M. D., Marcot, B. G., Olsen, B. J., Elphick, C. S., Hodgman, T. P., Guntenspergen, G. R. and Shriver, W. G. 2019. Estimates of tidal-marsh bird densities using Bayesian networks. – J. Wildl. Manage. 83: 109–120.

Witmer, G. W. 2005. Wildlife population monitoring: some practical considerations. – Wildl. Res. 32: 259–263.

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. – Ornithol. Sci. 12: 73–88.

Youtz, J., Goljani Amirkhiz, R. and Frey, J. K. 2022. Modeling the impact of climate change and wildfire on the dusky grouse (*Dendragapus obscurus*) in the American Southwest: implications for conservation. – Avian Conserv. Ecol. 17: 35.

Zwickel, F. C. and Bendell, J. F. 2004. Blue grouse: their biology and natural history. – NRC Research Press.

19032200x, 0, Downloaded from https://nsojournals.onlinelibrary.wiley.com/doi/10.1002/wlb3.01506 by Montana State University Library, Wiley Online Library on [19/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License