Received: 9 March 2021 | Accepted: 21 June 2021
DOI: 10.1002/2688-8319.12091

Ecological Solutions and Evidence | BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Using spatial distance sampling models to optimize survey effort and address violations of the design assumption

Joshua H. Schmidt[1] | William W. Deacy[2]

[1] Central Alaska Network, U.S. National Park Service, Fairbanks, Alaska, USA

[2] Arctic Network, U.S. National Park Service, Fairbanks, Alaska, USA

Correspondence
Joshua H. Schmidt, Central Alaska Network, U.S. National Park Service, 4175 Geist Road, Fairbanks, AK 99709, USA.
Email: Joshua_Schmidt@nps.gov

Funding information
U.S. National Park Service

Handling Editor: Steve Hill

Abstract

1. Conventional distance sampling approaches rely on the design assumption (i.e. uniform distribution of individuals in relation to transects) to ensure unbiased inference to the population of interest. However, randomized design recommendations are not always followed or may be impractical to implement for some survey types, particularly in cases where transects must be placed perpendicular to the habitat gradient. Full-likelihood spatial distance sampling models provide a potential solution to violations of the design assumption by jointly modelling the detection and occurrence processes using spatially indexed habitat covariates.
2. Through simulation and an applied example based on a survey for Dall's sheep in Alaska, USA, we used a full-likelihood distance sampling approach to investigate the potential for bias in cases where transects placed perpendicular to the habitat gradient (e.g. elevational contours) are non-randomly sampled. We also assessed the utility of spatial approaches in cases where transects are placed along linear features, such as roads or ridgelines, where habitat may be unrepresentative of the overall study area.
3. Our results showed that the full-likelihood approach was generally unbiased, even in extreme scenarios where habitat was inversely related to distance from the transect. For the Dall's sheep example, our results showed that more efficient designs with reduced sampling effort in low-quality habitats are a practical solution for reducing logistical costs when the data are analysed in a spatial modelling framework.
4. Together, our findings confirm and extend existing work suggesting that spatial distance sampling can be a useful solution when non-random designs are employed. Given the high cost of survey implementation in many cases, the development of valid alternatives to design-based inference will aid in the amount of information available for a variety of species. The results of our work will be useful for practitioners in assessing alternative designs relative to particular survey applications.

KEYWORDS
Dall's sheep, density gradient, non-random sampling, road-based sampling, spatial distance sampling, survey design, transect placement

1 | INTRODUCTION

Much of the wildlife population estimation literature has been dedicated to addressing the problem of imperfect detection of individuals to produce unbiased estimators of population parameters (Williams et al., 2002). Although there are a variety of survey and analytical tools available, conventional distance sampling (CDS) methods are widely used for estimating population abundance or density when the probability of detecting individuals is <1.0. CDS methods rely on four key assumptions (Buckland et al., 2001, 2015; Thomas et al., 2010): (1) animals are distributed independently of transect lines or points (i.e. the design assumption); (2) animals on the line are detected with certainty; (3) distance measurements are exact; and (4) animals are detected at their initial location. There are numerous extensions to basic distance sampling theory that address failures of the latter three assumptions, while proper survey design is generally required to meet the design assumption. However, there are circumstances where the design assumption is violated and transect layout cannot be assumed to be random with respect to animal distribution.

The problem of non-random distribution of animals within the study area is typically addressed through random or systematic (with a random start) line placement (Buckland et al., 2001). In cases where transects are not located independently of animal locations, animal density and detection probability are confounded and cannot be estimated from distance data alone under the basic CDS framework (Buckland et al., 2004). For this reason, survey designs employing transects along roads, trails or other linear features, where animal density may not be representative, limit inference. Although less serious, placement of transects parallel to habitat features (e.g. coastlines, elevation contours) and perpendicular to the density gradient is a related case that can lead to similar problems because animal distribution may be non-random relative to the habitat, and therefore density, gradient.

The general recommendation in a design-based framework is to orient transects parallel to density gradients whenever possible to reduce variation in occurrence probability among transects (figure 1.3 in Buckland et al., 2001), but this may not be feasible in some circumstances due to logistical limitations. For example cetaceans may occur at higher densities at certain distances from the coast, so transects oriented perpendicular to the coast (i.e. parallel to the density gradient) would better represent the entire population of interest and minimize bias. However, out of necessity some designs rely on habitat features whereby transects are generated perpendicular to the habitat and, presumably, density gradients. These designs represent a special case of the problem of linear features than can be appropriately addressed by incorporating additional sources of information (e.g. Marques et al., 2013) or through further modifications to survey design to insure inference to the population of interest (e.g. Becker & Quang, 2009; Obbard et al., 2015; Schmidt et al., 2012; Stapleton et al., 2016).

In contrast to a typical design where randomization is assumed in relation to the study area, when transects follow habitat features, designs instead rely on random distribution relative to the habitat gradient (e.g. elevation, distance from coast) to insure that on average the transects are distributed randomly with respect to individuals (e.g. Becker & Quang, 2009; Schmidt et al., 2012). This is a subtle shift from random (or systematic) distribution of transects that are not tied to habitat features and requires the assumption of randomness in relation to the particular habitat gradient. For example if cetaceans tend to be located near to shore, a transect located far from shore would likely record most observations at large distances. Conversely, those near to shore would record higher numbers primarily at short distances. If transects are placed uniformly with respect to distance from shore, the estimator should be unbiased, but if sampling is non-uniform then the detection function is impacted. In a more typical design where transects run parallel to the habitat/density gradient (e.g. from low to high density), random placement throughout the study area means transects can be added or removed more simply as long as the process is random. When transects are placed perpendicular to the density gradient, the design must ensure that any changes to sampling effort are uniformly distributed across the gradient. These requirements can lead to inefficient designs that require many transects to be surveyed in what may be low-quality habitat where few detections will occur. This can be a significant factor for survey feasibility when the expense of the survey platform is high (e.g. ship, aircraft). In some cases, stratification may be a simple solution, although sample sizes must then be divided among strata and it may not be possible to acquire sufficient data with which to estimate both the detection function and density within each strata. However, note that multiple-covariate distance sampling (MCDS; Marques et al., 2007) can be used to mitigate against limited sample sizes in individual strata.

Ongoing development of spatial distance sampling methods offers an alternative solution that may be used to relax some of the requirements of design-based estimators and provide more detailed ecological inference (Hedley & Buckland, 2004; Miller et al., 2013; Royle et al., 2004; Sillett et al., 2012). The most common implementation uses a two-stage approach where the probability of detection is first estimated, and then a spatial model is used to describe the observed counts using the estimate of detection probability as an offset (Buckland et al., 2015; Hedley & Buckland, 2004). Although this approach is useful in many situations, when transects lie perpendicular to the density gradient, the estimate of the detection function is confounded with the distribution of objects. A full-likelihood approach provides an elegant solution to the problem by simultaneously estimating both detection and occurrence (Buckland et al., 2016; Johnson et al., 2010; Mizel et al., 2018; Oedekoven et al., 2013, 2014; Yuan et al., 2017). By estimating both the detection process and the occurrence processes concurrently, confounding is resolved. Therefore, a full-likelihood spatial modelling approach may provide a reasonable path forward for more efficient and effective designs employing non-random transect placement.

Our exploration of the full-likelihood spatial distance sampling approach as a means to address violations of the design assumption was motivated by our desire to address the relative inefficiency of Dall's sheep (*Ovis dalli*) surveys in Alaska, USA, as currently implemented (Rattenbury et al., 2018; Schmidt & Rattenbury, 2013; Schmidt et al., 2012). In these surveys, transects are systematically placed along elevational contours (i.e. perpendicular to the gradient) throughout the study area, resulting in a design where a sizable subset of transects may be unlikely to contain sheep. Given the cost of aerial surveys, a valid alternative facilitating increased effort in higher density areas would have obvious advantages.

To explore the utility of the full-likelihood spatial distance sampling approach in addressing violations of the design assumption and facilitating survey efficiency, we recast the group-based hierarchical distance sampling model of Schmidt et al. (2012) using a pixel-based framework (Kery & Royle, 2021; Mizel et al., 2018). Doing so allowed the incorporation of habitat covariates and other spatial predictors at a fine spatial grain. We used a simulation study and an applied example from a Dall's sheep survey in Wrangell-St. Elias National Park and Preserve to explore how a spatial full-likelihood model-based approach might be used to relax the design assumption for CDS surveys in cases where transects are placed perpendicular to expected density gradients, along linear features, or other non-random designs. Although our work was motivated by the desire to increase the efficiency of ongoing distance sampling surveys for Dall's sheep, our findings will be of use to practitioners in a variety of situations where the design assumption is not met.

2 | MATERIALS AND METHODS

2.1 | Model structure

We recast the group-based hierarchical distance sampling model of Schmidt et al. (2012) in a pixel-based framework (Kery & Royle, 2016, 2021; Mizel et al., 2018). Under the pixel-based formulation, the study area is overlain with a grid of pixels and instead of modelling the distance to individual groups, one models the distance to individual pixels, a subset of which contains detected groups. The advantage of this formulation is that fine-scale spatial patterns in animal distribution can be incorporated by assigning habitat values to each individual pixel and used to predict group occurrence. We simplified the pixel-based model (Kery & Royle, 2016, 2021; Mizel et al., 2018) to reflect the probability of a group occurring in a given pixel, rather than counts within pixels. This simplification becomes possible when pixel size is small relative to the density of groups on the landscape. The resulting model is a version of a thinned point process model whereby the probability of group occurrence at the pixel level is a function of habitat and observed occurrences are thinned by the detection process (Kery & Royle, 2021, pp. 674–675).

In the pixel-based formulation, the observations, y, in each pixel, i, are modelled as

$$y_i \sim \text{Bernoulli}(\mu_i),$$

where

$$\mu_i = p_i \, \Psi_i.$$

Detection probability, p_i, is represented by the half-normal detection function $p_i = \exp\left(-\frac{x_i^2}{2\sigma_i^2}\right)$, where x_i is the distance from the transect to each group, and σ_i is the scale parameter which is modeled as

$$\log(\sigma_i) = Y_i' \, \boldsymbol{\alpha},$$

where Y_i' is a vector of known covariate values and $\boldsymbol{\alpha}$ are the associated coefficients. Note that other functions, such as the hazard-rate, may be substituted for the half-normal as appropriate. Group occurrence, Ψ_i, is modeled as

$$\Psi_i \sim \text{Bernoulli}(\lambda_i),$$

where

$$\lambda_i = \text{Area}_i \times (1/(1 + \exp(-(X_i' \, \boldsymbol{\beta}))))$$

and Area_i is the proportion of the area of each pixel that was within the surveyed strip, X_i' is a vector of known covariate values and $\boldsymbol{\beta}$ are the associated coefficients. When individuals occur in groups, group size, s_i, can be modeled as

$$\log(\hat{s}_i) = Z_i' \boldsymbol{\delta},$$

where Z_i' is a vector of known covariate values and $\boldsymbol{\delta}$ are the associated coefficients.

The pixel-based model may be cast in a spatial context when spatially indexed covariates are included in X_i'. We fit spatial and non-spatial versions of the pixel-based model to data simulated under several scenarios (see next sub-section) to assess bias given each design. For simplicity, we did not consider models with covariates other than habitat. As our main concern was bias related to the design assumption, we did not investigate other factors expected to primarily impact precision and coverage (e.g. sampling intensity, sample size, strip-width, etc.).

2.2 | Simulations

We considered three common survey scenarios that are at risk of violating the design assumption, potentially resulting in bias. For each sampling scenario, we compared the abilities of the spatial and

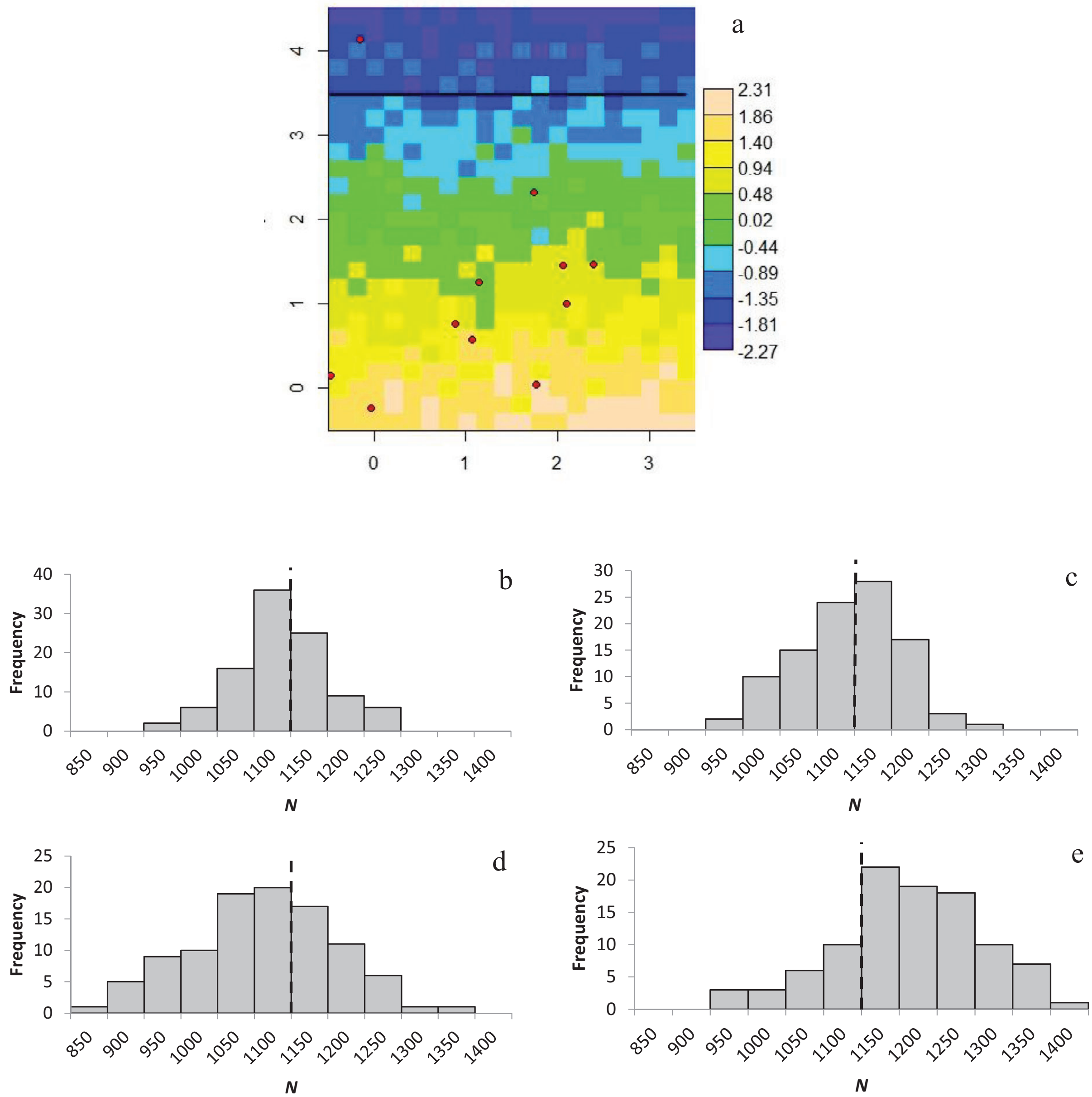

FIGURE 1 Panel a depicts a single iteration of a scenario where transects (black line) are randomly placed perpendicular to the habitat gradient and 11 individuals (red dots) are distributed based on the habitat values of each cell. Each transect for each iteration was generated at a random *y*-value to cover the entire habitat gradient uniformly. The histograms represent the frequency of estimated abundance ($\hat{N}$) based on 100 datasets of 100 transects each analysed under the following conditions: spatial model with the full dataset (b), the spatial model with 50% reduced effort in low-quality habitats (c), the non-spatial model with the full dataset (d) and the non-spatial model with 50% reduced effort in low-quality habitats (e). Vertical dashed lines indicate the true population size

non-spatial approaches to recover the true population size in cases where transects are oriented perpendicular to the habitat gradient. For each replicate under each scenario, we generated 11 individuals that were distributed using occurrence probabilities based on the pixel-level habitat values within a 20 × 25 grid, restricting the distribution to allow only a single individual to occur in each grid cell. We then placed a single transect within the sample unit and simulated a detection process by which some proportion of the 11 individuals was then detected. The generation of a sample unit, individuals within the unit and the sampled line transect was then repeated 100 times to produce a hypothetical dataset consisting of 100 sample units sampled by one transect each, for a true population size of 1100 individuals. We then fit both spatial and non-spatial versions of the pixel-based model to the data to estimate abundance and replicated the whole process 100 times (i.e. 100 datasets) for each scenario.

The first scenario was directly related to applications which use transects placed perpendicular to a habitat gradient (Figure 1a), as is the case for distance sampling surveys for Dall's sheep (Schmidt & Rattenbury, 2013; Schmidt et al., 2012) and bears (*Ursus* spp.; Becker & Quang, 2009). We simulated a habitat gradient that increased linearly (with some random noise) across each sampled unit so that lower quality habitat occurred at one end and higher quality habitat occurred at the other. We placed each transect randomly in the sample unit but aligned perpendicular to the habitat gradient. We considered two options under this scenario: random distribution of transects, and random distribution of transects with 50% of those transects that occurred in the lowest quality habitats excluded from sampling (i.e. y-values < 3; see Figure 1a). The first option reflects the typical design for surveys employing contour transects (Becker & Quang, 2009; Schmidt et al., 2012) whereby the uniformity assumption is met by sampling all habitats equally across the gradient. In theory, both the spatial and non-spatial estimators should be unbiased given a random distribution of transects relative to the gradient. The reduced effort option reflected a violation of the design assumption by undersampling in low-quality areas where few detections are expected. Theoretically, the non-spatial estimator should be biased when poor-quality habitats are undersampled, but we expected that the full-likelihood spatial estimator would be unbiased under both scenarios. Although stratification and MCDS can be useful in addressing unequal sampling of certain habitat types (Marques et al., 2007), for generality, we assumed stratification was not conducted.

The remaining scenarios we considered explored another common violation of the uniformity assumption that results when linear features such as roads, trails, ridgelines or other non-random linear features are used as transects (Figures 2a and 2b). Sampling along a feature such as a road or trail can be problematic, for example because the road surface itself may form a corridor of low- or high-quality habitat near the transect line where few or no individuals occur. To represent this hypothetical road or trail corridor effect, we generated simulations using a scenario with a random habitat grid and a corridor of low-quality habitat through the middle of the sample unit. For each iteration, the transect was placed along the centre of the 'road corridor', with the expectation the non-spatial estimator would be biased because few detections would occur near the line. In contrast, we expected the spatial estimator would be relatively unbiased after accounting for habitat quality, assuming that the strip half-width extended beyond the road corridor by a reasonable distance. We also considered another more complex scenario that could occur if transects were located along features such as ridgelines where habitat quality changes with elevation (Figure 2b). This type of design creates obvious pathological problems for estimation because distance and occurrence probability are directly confounded, potentially making it very difficult to separate the detection and occurrence processes (Johnson et al., 2010).

We constructed each set of simulations in R 3.6.1 (R Core Team, 2019) using the AHMbook package (Kery et al., 2020). We modified the simDSM function to accommodate designs with multiple linear transects reflecting each simulation scenario described above. Specifically, we modified the habitat distribution for the three scenarios as follows: (1) a habitat gradient [values $= -2$ to $2 + N(0,0.25)$] increasing from $y = 4.5$; (2) random habitat distribution [values $= -3$ to 3] with a band of habitat $= -3$ between $y = 1.5$ and 2.5; and (3) a habitat gradient [values $= -1,1 + N(0,0.15)$] increasing from $y > 2$ and $y < 2$. For simplicity, we assumed that all $\text{Area}_i = 1$. We conducted model fitting using OpenBUGS 3.2.3 (Thomas et al., 2006). For the analysis of each simulated dataset under each scenario, we ran a single chain for 1000 iterations with the first 250 discarded as burn-in, retaining the remainder for inference. Exploratory analysis suggested this was sufficient to reach model convergence (results not shown). We then compared the distribution of the estimates of abundance from each scenario with the true population size to assess bias.

2.3 | Dall's sheep application

We used Dall's sheep survey data collected using a systematic contour transect design in Wrangell-St. Elias National Park and Preserve in 2020 as a worked example. Our sampling design was based on a systematic 7.5-km grid of points generated across the 3303 km^2 study area with each point ($n = 63$) representing a transect centre-point. At each centre-point, we generated a $\leq$15-km transect following the elevational contour (i.e. contour transect) at that location (Figure 3a). The contour transect design resulted in a randomly distributed collection of transects that were oriented perpendicular to the habitat gradient (i.e. elevation). When a full-length transect could not be generated due to a lack of habitat (i.e. mountain top), another transect was generated nearby at the same elevation (outside the study area boundary if necessary). A pilot-observer team surveyed each transect from a fixed-wing aircraft flown at ~90 m above-ground level, recording the perpendicular distance from the transect line to each detected group, as well as the number of individuals in each group. We left-truncated the resulting data at 22 m to account for the partially observable strip beneath the aircraft and right-truncated the data at 685 m, resulting in a 663 m strip-width. Additional details regarding survey design and sampling protocol are presented by Schmidt et al. (2012) and Schmidt and Rattenbury (2013).

Prior to analysis, we generated a grid of 100 × 100 m pixels across the study area and extracted the mean elevation and elevation^2 values for each pixel using a digital elevation model. We chose the pixel size given that a group was defined as a collection of individuals within <100 m (Schmidt et al., 2012), thereby ensuring that only one group could occur within each pixel. We then calculated Area_i and the distance from the transect to each pixel centroid for all pixels that fell within the sampled strip of each transect. We attributed each detected group to the appropriate pixel based on the observed group location. We then fit the pixel-based model to the data under the spatial and non-spatial formulations, replacing the half-normal detection model with the hazard-rate function (Schmidt et al., 2012):

$$p_i = 1 - \exp\left(-\frac{x_i}{\sigma_i}\right)^{-b},$$

which contains the additional shape parameter, b. Given that group sizes were overdispersed, we included a random effect at the pixel level in the group size submodel. We included elevation and elevation^2 as habitat covariates predicting λ_i in the spatial model based on our

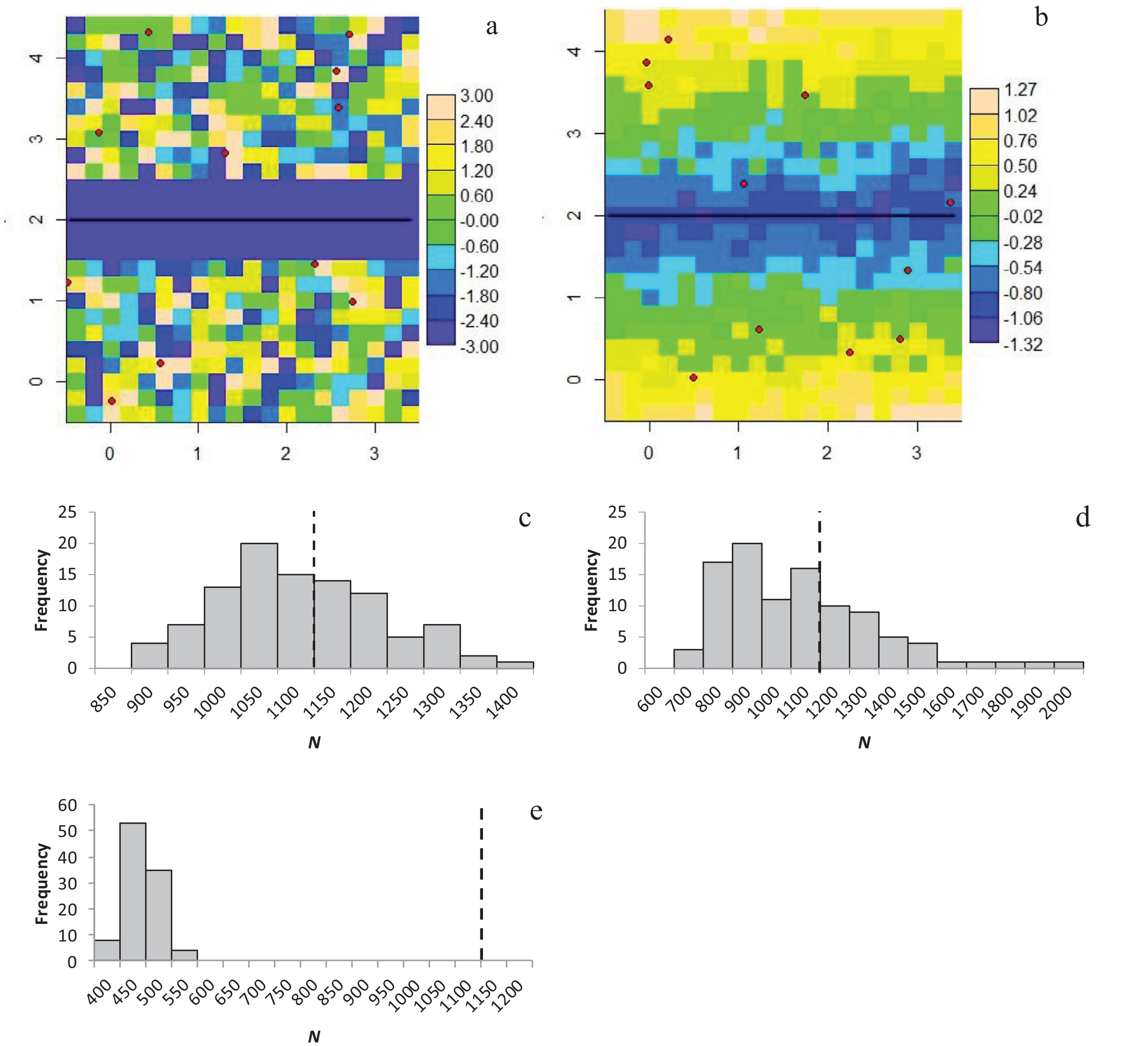

FIGURE 2 The first two panels represent a single iteration for each of two scenarios: transects placed along a linear feature with a corridor of poor-quality habitat at close distances (e.g. a road corridor; a), and transects placed along a linear feature aligned perpendicular to a habitat gradient (e.g. a ridgeline; b). All transects were generated at $y = 2$ to represent sampling along the linear feature. Each example depicts one iteration containing a single transect (black line) and 11 individuals (red dots) that are distributed based on the habitat values of each cell. The histograms represent the frequency of estimated abundance ($\hat{N}$) based on 100 datasets of 100 transects each analysed under the following conditions: transects placed along a road corridor (spatial model; c), transects placed along a linear feature where habitat quality increases with distance (spatial model; d) and transects placed along a road corridor (non-spatial model; e). Vertical dashed lines indicate the true population size

observation that sheep tend to occur less frequently at either end of the elevational distribution. We did not consider any additional covariates in any of the submodels. For completeness, we also fit the equivalent non-spatial group-based model of Schmidt et al. (2012) to the data to demonstrate the equivalence between the group-based and pixel-based model structures.

In addition to the above analyses, we also considered the 'reduced effort' case where we omitted data from 50% of the transects occurring at high elevations (i.e. >1 SD above mean elevation) where sheep were least likely to occur. Doing so represented an ~11% reduction in total survey effort and is comparable to the first simulation scenario with reduced effort. The reduced effort case addresses the question of whether under-sampling in low-quality habitats can be used to reduce survey cost and increase logistical efficiency in future surveys. We refit both the spatial and non-spatial versions of the pixel-based model for comparison with the results from the full dataset. Due to a relatively low number of group detections ($n = 73$, full dataset; $n = 72$, reduced dataset), we used informed priors for the shape and scale parameters of the detection function: $\sigma \sim N(0.7, 0.044)$ and $b \sim N(2, 0.11)$, based on prior work (Rattenbury et al., 2018; Schmidt & Rattenbury, 2013).

For each model fit to the Dall's sheep data, we ran two independent chains of 10,000 iterations each in OpenBUGS, discarding the initial

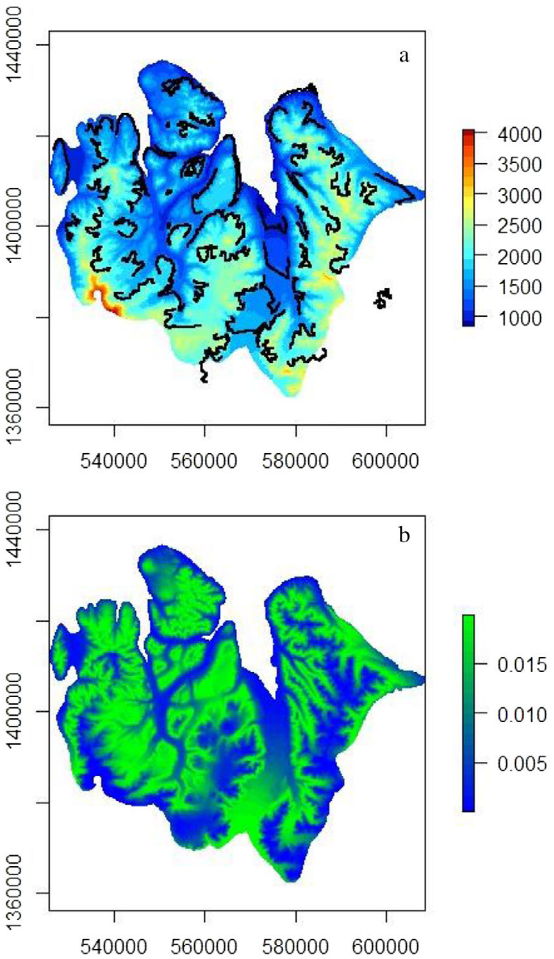

FIGURE 3 Panel a shows the contour transects surveyed (black lines) for Dall's sheep in Wrangell-St. Elias National Park and Preserve, Alaska, USA in 2020 in relation to the distribution of elevation (m). Portions of some surveyed transects were located outside the study area to facillitate uniform sampling across the distribution of elevations. Panel b depicts predicted probability of Dall's sheep group occurrence in 2020 in a portion of Wrangell-St. Elias National Park and Preserve based on the full-likelihood spatial model including a curvilinear relationship between elevation and group occurrence. The *x* and *y* axes represent spatial coordinates in meters

5000 iterations as burn-in and retaining the remainder for inference. For the non-spatial models, estimates of abundance within the sampled area were extrapolated directly to the entire study area. For the spatial models, we predicted pixel-level abundance using the posteriors of the predictors and then summed the predictions across all pixels to estimate total abundance and 95% credible intervals.

3 | RESULTS

3.1 | Simulations

The spatial model was relatively unbiased under the habitat gradient scenario under both the full and reduced sampling options (Figures 1b and 1c), indicating the spatial approach adequately addressed the violation of the design assumption caused by targeted undersampling of some habitats. As expected, the non-spatial estimator showed little evidence of bias when transects were distributed randomly with respect to the habitat gradient (Figure 1d). However, when the lowest quality habitats were undersampled by 50%, abundance estimates tended to be biased high (Figure 1e), following our expectation of the consequences of violating the design assumption. The variability in abundance estimates from the spatial model was also lower as compared to the equivalent non-spatial results (Figure 1).

When transects were placed along a hypothetical road corridor, results based on the spatial model were relatively unbiased, suggesting that the model was largely able to compensate for a substantial violation of the design assumption by modelling the spatial distribution of habitat (Figure 2c). Somewhat surprisingly, results were also generally unbiased even when habitat quality and distance from the line were directly confounded (Figure 2d), although the distribution of estimates suggested some instability. In contrast, the non-spatial model performed poorly in the road-corridor scenario (Figure 2e). This result was predictable given that few detections were expected near the transect line. Overall, the simulation results confirm that the non-spatial estimator was unbiased when the design assumption was met but was biased when the design assumption was violated. In contrast, the full-likelihood spatial modelling approach was generally unbiased under either scenario, suggesting that the incorporation of habitat as a predictor of group occurrence largely mediated violations of the design assumption.

3.2 | Dall's sheep application

The results from the group-based model were comparable to those from the non-spatial pixel-based model, reflecting the equivalence of the two formulations (Table 1). The pixel-based formulation was ~19% more precise, presumably due to the finer spatial grain of the analysis (i.e. 100 × 100 m pixels vs. total area surveyed per transect), and the spatial approach incorporating a curvilinear elevation effect

TABLE 1 Comparisons of estimated Dall's sheep abundance ($\hat{N}$) in the Wrangell-St. Elias National Park and Preserve study area under the group-based and pixel-based model formulations. We also include the 95% Bayesian credible intervals (95%CrI) and the coefficient of variation (CV) for each case. Results with and without the curvilinear effect of elevation on group occurrence, as well as the impact of undersampling by 50% at elevations >1 SD above the mean (reduced), are also shown

Model	$\hat{N}$	95%CrI	CV
Group-based (full, without elevation)	3762	2716–5057	16%
Pixel-based (full, without elevation)	3776	2882–4876	13%
Pixel-based (full, with elevation)	3715	2809–4720	16%
Pixel-based (reduced, without elevation)	4096	3111–5307	13%
Pixel-based (reduced, with elevation)	3744	2804–4795	16%

produced comparable estimates. As was the case for the simulations, the reduction of sampling in low-quality habitats (i.e. high elevation) had little impact on the spatial estimator when elevation was included as a covariate, but there was positive apparent bias in the non-spatial estimator (Table 1). The estimated relationship between elevation and Dall's sheep occurrence based on the results of the spatial model was curvilinear, indicating that sheep tended to occur more frequently at mid-elevations (Figure 3b), in agreement with our expectation.

4 | DISCUSSION

Here, we have demonstrated how a spatial distance sampling framework can be used to mitigate against common violations of the design assumption in distance sampling applications and lead to practical modifications of inefficient survey designs. The idea that a spatial approach could be used to compensate for non-random and opportunistic designs is not new (Johnson et al., 2010); however, a direct assessment of some of the more common applications that risk violating the design assumption will help practitioners understand how these approaches may be employed in commonly encountered situations. Many applications use linear features as transects or must place transects perpendicular to the habitat gradient, and our work will serve as a valuable resource for researchers designing projects that do not adhere to the design assumption. Overall, we expect that our work will prove useful for practitioners desiring to apply model-based approaches to distance sampling in complex sampling scenarios.

The design assumption is a fundamental component of most CDS applications (Buckland et al., 2001) because estimator bias is directly proportional to the degree to which the sample is unrepresentative of the area of interest. Herein lies the difficulty with using non-random designs (i.e. convenience sampling, roads as transects), without knowledge of the spatial distribution of individuals, the degree to which the estimator will be biased is typically unknowable. Spatial modelling approaches can account for non-representative sampling by explicitly modeling the distribution of groups relative to spatial processes (Buckland et al., 2015; Johnson et al., 2010), but inference is conditioned on the sample. This is a critical point to consider. If certain portions of the population of interest are excluded from sampling (e.g. certain habitat types), model predictions are likely to be inaccurate. As an extreme example, consider a survey that only samples forested habitats within a study area. Under such a design, predicting abundance in open meadow habitats would likely be nonsensical, regardless of the framework employed. The issue is similar for scenarios using road corridors or ridgelines as transects. If the strip-width is narrow relative to the width of the habitat gradient along the linear feature, the model-based estimator will be biased because some habitat types will be excluded from sampling. The design considerations necessary to ensure unbiased inference will vary depending on the particular survey situation, but regardless, the practitioner must ensure that the sample is representative of the population of interest.

Spatial distance sampling approaches have been shown to be useful in a variety of applications, including those with non-random survey designs (e.g. Hedley & Buckland, 2004; Johnson et al., 2010; Miller et al., 2013; Mizel et al., 2018). Here, we have expanded on this body of work to further investigate the utility of spatial approaches, specifically in cases where transects are oriented perpendicular to the habitat gradient, in direct contrast to recommended practice (Buckland et al., 2001, p. 239). Given our goal of broadly determining the role of spatial approaches in mitigating bias for a particular suite of non-random designs, we limited our exploration to a broad assessment of bias for a few plausible scenarios. Our results were consistent with past work indicating that spatial modelling approaches can indeed mitigate against bias, even in the somewhat extreme scenarios we considered. However, we acknowledge that our exploration was limited and that there are many factors that would determine the effectiveness of a given design (e.g. effective strip-width, habitat gradient, object density, sample sizes). We recommend that researchers treat our findings as a starting point and encourage practitioners to carefully consider the details of a proposed application when assessing whether a spatial approach is indeed likely to avoid bias for a particular survey design. Indeed, the simDSM function (Kery et al., 2020) provides a convenient framework for practitioners to assess whether a spatial approach is appropriate for their specific sampling problem.

Our reframing of the group-based distance sampling model of Schmidt et al. (2012) in the pixel-based framework of Kery and Royle (2021) provides a natural link between existing applications and ongoing development of full-likelihood spatial modelling approaches. Although spatial structure could be incorporated in the group-based model of Schmidt et al. (2012), doing so requires the assumption that within-transect distribution of individuals is uniform, as was done in a point transect context (Royle et al., 2004). The coarseness of the spatial grain (i.e. ~1000-fold larger in the Dall's sheep application) of the spatial group-based model may not be particularly useful in many situations. Recasting the group-based model in the pixel-based framework allowed the direct incorporation of spatial information at a more relevant (i.e. finer) spatial grain. The Dall's sheep application indicated that inference was effectively identical under the non-spatial version of both formulations, although the non-spatial pixel-based version was more precise, presumably due to reduced variation among pixels as compared to variation among transects in the group-based model. Indeed, increased estimator precision is one of the expected benefits of the spatial modelling approach (Mizel et al., 2018).

The work we presented focused on the use of habitat variables as predictors of pixel-level occurrence to create a density surface across the study area. Absent a strong relationship between density and habitat, or in cases where important habitat covariates are not available, the spatial structure in the data can be modelled directly using a variety of strategies (e.g. Banerjee et al., 2004; Ver Hoef et al., 2018), although generalized additive models (Wood, 2006) tend to be most commonly employed in spatial distance sampling applications (e.g. Hedley & Buckland, 2004; Johnson et al., 2010; Kery & Royle, 2021; Sigourney et al., 2020). The modelling of spatial structure directly can also be used in combination with covariate modelling to help account for residual structure that is not explained by the available habitat variables. Modelling the spatial structure directly is a useful alternative in situations

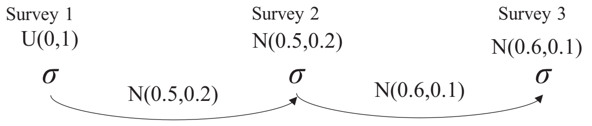

FIGURE 4 Diagrammatic example of how information on the detection process can be incorporated through time or among surveys using informed priors on the scale parameter, σ. The prior distribution for each survey is specified under each survey heading and the resulting posterior distribution (i.e. prior + data) placed above the connecting arrows. In this example, the posterior from survey 1 is used as the prior for survey 2 and so on

where sufficient habitat covariate values are not available or when inference to ecological relationships is not of interest.

Although we considered several related survey design scenarios, the motivation for our work was to improve the accuracy and cost-effectiveness of the ongoing Dall's sheep monitoring program in Alaska. We confirmed that a systematic sampling design is indeed unbiased when analysed in a non-spatial framework. However, our findings suggest that surveys could be made more cost-effective by prioritizing sampling in areas where sheep are more likely to occur, thereby increasing the number of detections for a given level of survey effort and providing more data with which to estimate complex habitat associations. Full coverage of the elevational gradient would still be required, but equivalent allocation of effort across the available habitat gradient is not necessary. For long-term monitoring programs, a logical path forward would be to consider sharing information over time to better estimate habitat associations in a particular area, as has been done for the parameters of the detection process (Schmidt & Rattenbury, 2013; Figure 4). Doing so would leverage data from multiple years to inform better spatial allocation of sampling effort and improve estimation of habitat associations. When long-term data are not available, it may be possible to leverage information collected from other areas or for similar species, either through joint analysis (e.g. Schmidt et al., 2014) or through the creation of informed priors (McCarthy & Masters, 2005).

The use of spatial distance sampling for group-dwelling species such as Dall's sheep can be complex, particularly for monitoring projects conducted through time (e.g. Schmidt & Rattenbury, 2018) or in a spatial context because variation in both group occurrence and group size must be considered. A natural extension to the model we used would be the inclusion of spatial covariates related to both group occurrence and group size (Hedley & Buckland, 2004), reflecting different spatial processes governing where groups occur versus the size of each group. For example the occurrence of groups in general may be related to one particular habitat feature (e.g. elevation), but the size of individual groups may be related to another (e.g. vegetation quality). The interplay between group occurrence and the aggregation of individuals within groups has potentially important ecological implications that may be lost if the spatial process is only considered at one level. Similarly, more complex structures may be useful in addressing changes in spatial distribution in both space and time (e.g. Barnett et al., 2020; Camp et al., 2020).

Spatial distance sampling approaches are powerful tools that, when used appropriately, can leverage information collected under a variety of sample designs. Although spatial approaches cannot completely ignore sampling design, the added flexibility for field implementation can be very useful. We anticipate that ongoing development of spatial approaches will ultimately enable researchers to address increasingly interesting and complex ecological questions in a variety of settings.

ACKNOWLEDGEMENTS

We thank J. A. Putera for logistical support and D. T. Schertz for technical assistance throughout the project. J. Cummings safely piloted the aircraft for the survey in Wrangell St. Elias National Park and Preserve. S. T. Buckland, M. C. MacCluskie, J. D. Mizel, E. J. Wald and an anonymous reviewer provided helpful comments on an earlier draft of this manuscript. Funding for this work was provided by the U.S. National Park Service through the Arctic and Central Alaska Inventory and Monitoring Networks and Wrangell St. Elias National Park and Preserve. Any use of trade names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

JS and WD jointly conceived the idea. JS conducted the analyses and wrote the manuscript. JS and WD edited the manuscript. WD collected the data in the applied example.

DATA AVAILABILITY STATEMENT

The Dall's sheep data used in the applied example are available at https://irma.nps.gov/DataStore/Reference/Profile/2284639 (Schmidt & Deacy, 2021).

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/2688-8319.12091

ORCID

Joshua H. Schmidt https://orcid.org/0000-0002-5803-6431

REFERENCES

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC Press.

Barnett, L. A. K., Ward, E. J., & Anderson, S. C. (2020). Improving estimates of species distribution change by incorporating local trends. *Ecography*, *44*(3), 427–439. https://doi.org/10.1111/ecog.05176

Becker, E. F., & Quang, P. X. (2009). A gamma-shaped detection function for line-transect surveys with mark-recapture and covariate data. *Journal of Agricultural, Biological, and Environmental Statistics*, *14*, 207–223. https://doi.org/10.1198/jabes.2009.0013

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling*. Oxford University Press.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2004). *Advanced distance sampling*. Oxford University Press.

Buckland, S. T., Oedekoven, C. S., & Borchers, D. L. (2016). Model-based distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, *21*, 58–75. https://doi.org/10.1007/s13253-015-0220-7

Buckland, S. T., Rexstad, E. A., Marques, T. A., & Oedekoven, C. S. (2015). *Distance sampling: Methods and applications*. Springer International Publishing.

Camp, R. J., Miller, D. L., Thomas, L., Buckland, S. T., & Kendall, S. J. (2020). Using density surface models to estimate spatio-temporal changes in population densities and trend. *Ecography*, *43*, 1079–1089. https://doi.org/10.1111/ecog.04859

Hedley, S. L., & Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, *9*, 181–199. https://doi.org/10.1198/1085711043578

Johnson, D. S., Laake, J. L., & Ver Hoef, J. M. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics*, *66*, 310–318. https://doi.org/10.1111/j.1541-0420.2009.01265.x

Kery, M., & Royle, J. A. (2016). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance, and species richness in R and BUGS* (Vol. 1). Academic Press.

Kery, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance, and species richness in R and BUGS* (Vol. 2). Academic Press.

Kery, M., Royle, J. A., & Meredith, M. (2020). *AHMbook*: F*unctions and data for the book 'Applied Hierarchical Modeling in Ecology' Vols 1 and 2. R package version 0.2.2*. https://CRAN.R-project.org/package=AHMbook

Marques, T. A., Buckland, S. T., Bispo, R., & Howland, B. (2013). Accounting for animal density gradients using independent information in distance sampling surveys. *Statistical Methods and Applications*, *22*, 67–80. https://doi.org/10.1007/s10260-012-0223-2

Marques, T. A., Thomas, L., Fancy, S. G., & Buckland, S. T. (2007). Improving estimates of bird density using multiple-covariate distance sampling. *Auk*, *124*, 1229–1243. https://doi.org/10.1093/auk/124.4.1229

McCarthy, M. A., & Masters, P. (2005). Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology*, *42*, 1012–1019. https://doi.org/10.1111/j.1365-2664.2005.01101.x

Miller, D. L., Burt, M. L., Rexstad, E. A., & Thomas, L. (2013). Spatial models for distance sampling data: Recent developments and future directions. *Methods in Ecology and Evolution*, *4*, 1001–1010. https://doi.org/10.1111/2041-210X.12105

Mizel, J. D., Schmidt, J. H., & Lindberg, M. S. (2018). Accommodating temporary emigration in spatial distance sampling models. *Journal of Applied Ecology*, *55*, 1456–1464. https://doi.org/10.1111/1365-2664.13053

Obbard, M. E., Stapleton, S., Middel, K. R., Thibault, I., Brodeur, V., & Jutras, C. (2015). Estimating the abundance of the Southern Hudson Bay polar bear subpopulation with aerial surveys. *Polar Biology*, *38*, 1713–1725. https://doi.org/10.1007/s00300-015-1737-5

Oedekoven, C. S., Buckland, S. T., Mackenzie, M. L., Evans, K. O., & Burger, L. W., Jr. (2013). Improving distance sampling: Accounting for covariates and non-independency between sampled sites. *Journal of Applied Ecology*, *50*, 786–793. https://doi.org/10.1111/1365-2664.12065

Oedekoven, C. S., Buckland, S. T., Mackenzie, M. L., King, R., Evans, K. O., & Burger, L. W., Jr. (2014). Bayesian methods for hierarchical distance sampling methods. *Journal of Agricultural, Biological, and Environmental Statistics*, *19*, 219–239. https://doi.org/10.1007/s13253-014-0167-0

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rattenbury, K. L., Schmidt, J. H., Swanson, D. K., Borg, B. L., Mangipane, B. A., & Sousanes, P. J. (2018). Delayed spring onset drives declines in abundance and recruitment in a mountain ungulate. *Ecosphere*, *9*, e02513. https://doi.org/10.1002/ecs2.2513

Royle, J. A., Dawson, D. K., & Bates, S. (2004). Modeling abundance effects in distance sampling. *Ecology*, *85*, 1591–1597. https://doi.org/10.1890/03-3127

Schmidt, J. H., & Deacy, W. W. (2021). *Data from: Using spatial distance sampling models to optimize survey effort and address violations of the design assumption*. IRMA Data Store. https://irma.nps.gov/DataStore/Reference/Profile/2284639

Schmidt, J. H., Flamme, M. J., &and, & Walker, J. (2014). Habitat use and population status of yellow-billed and Pacific loons in western Alaska, USA. *Condor: Ornithological Applications*, *116*, 483–492. https://doi.org/10.1650/CONDOR-14-28.1

Schmidt, J. H., & Rattenbury, K. L. (2013). Reducing effort while improving inference: Estimating Dall's sheep abundance and composition in small areas. *Journal of Wildlife Management*, *77*, 1048–1058. https://doi.org/10.1002/jwmg.557

Schmidt, J. H., & Rattenbury, K. L. (2018). An open-population distance sampling framework for assessing population dynamics in group-dwelling species. *Methods in Ecology and Evolution*, *9*, 936–945. https://doi.org/10.1111/2041-210X.12932

Schmidt, J. H., Rattenbury, K. L., Lawler, J. P., & MacCluskie, M. C. (2012). Using distance sampling and hierarchical models to improve estimates of Dall's sheep abundance. *Journal of Wildlife Management*, *76*, 317–327. https://doi.org/10.1002/jwmg.216

Sigourney, D. B., Cahvez-Rosales, S., Conn, P. B., Garrison, L., Josephson, E., & Palka, D. (2020). Developing and assessing a density surface model in a Bayesian hierarchical framework with a focus on uncertainty: Insights from simulations and an application to fin whales (*Balaenoptera physalus*). *PeerJ*, *8*, e8226. https://doi.org/10.7717/peerj.8226

Sillett, T. S., Chandler, R. B., Royle, J. A., Kery, M., & Morrison, S. A. (2012). Hierarchical distance-sampling models to estimate population size and habitat-specific abundance of an island endemic. *Ecological Applications*, *22*, 1997–2006. https://doi.org/10.1890/11-1400.1

Stapleton, S., Peacock, E., & Garshelis, D. (2016). Aerial surveys suggest long-term stability of the seasonally ice-free Foxe Basin (Nunavut) polar bear population. *Marine Mammal Science*, *32*, 181–201. https://doi.org/10.1111/mms.12251

Thomas, A., O'Hara, B., Ligges, U., & Sturtz, S. (2006). Making BUGS open. *R News*, *6*, 12–17.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R. B., Marques, T. A., & Burnham, K. P. (2010). Distance software: Design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, *47*, 5–14. https://doi.org/10.1111/j.1365-2664.2009.01737.x

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., & Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs*, *88*, 36–59. https://doi.org/10.1002/ecm.1283

Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations*. Academic Press.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Chapman and Hall/CRC.

Yuan, B. Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., Rue, H., & Gerrodette, T. (2017). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, *11*, 2270–2297.

How to cite this article: Schmidt, J. H., & Deacy, W. W. (2021). Using spatial distance sampling models to optimize survey effort and address violations of the design assumption. *Ecological Solutions and Evidence*, *2*, e12091. https://doi.org/10.1002/2688-8319.12091