

RESEARCH ARTICLE

A Unimodal Model for Double Observer Distance Sampling Surveys

Earl F. Becker*, Aaron M. Christ

Alaska Department of Fish and Game, Anchorage, Alaska, United States of America

* earl.becker@alaska.gov

Abstract

Distance sampling is a widely used method to estimate animal population size. Most distance sampling models utilize a monotonically decreasing detection function such as a half-normal. Recent advances in distance sampling modeling allow for the incorporation of covariates into the distance model, and the elimination of the assumption of perfect detection at some fixed distance (usually the transect line) with the use of double-observer models. The assumption of full observer independence in the double-observer model is problematic, but can be addressed by using the point independence assumption which assumes there is one distance, the apex of the detection function, where the 2 observers are assumed independent. Aerially collected distance sampling data can have a unimodal shape and have been successfully modeled with a gamma detection function. Covariates in gamma detection models cause the apex of detection to shift depending upon covariate levels, making this model incompatible with the point independence assumption when using double-observer data. This paper reports a unimodal detection model based on a two-piece normal distribution that allows covariates, has only one apex, and is consistent with the point independence assumption when double-observer data are utilized. An aerial line-transect survey of black bears in Alaska illustrate how this method can be applied.



OPEN ACCESS

Citation: Becker EF, Christ AM (2015) A Unimodal Model for Double Observer Distance Sampling Surveys. PLoS ONE 10(8): e0136403. doi:10.1371/journal.pone.0136403

Editor: Aaron W. Reed, University of Missouri Kansas City, UNITED STATES

Received: October 22, 2014

Accepted: August 4, 2015

Published: August 28, 2015

Copyright: © 2015 Becker, Christ. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Funding was provided by the Alaska Department of Fish and Game, Federal Aid in Wildlife Restoration Funds.

Competing Interests: The authors have declared that no competing interests exist

Introduction

Animal population size is an important parameter for many wildlife and ecological problems. Distance sampling is a widely used method to estimate animal population size [1]. Conventional distance sampling (CDS) models animal detection as function of the perpendicular distance between the animal and transect to obtain a population estimate. In most applications, there are additional variables that affect an animal's detection probability [2–4] and these variables can be used for modeling animal detection. When appreciable variation in animal detection is present, multiple covariate distance sampling (MCDS) models are superior to CDS models if the appropriate covariates are measured [5]. CDS and MCDS models make the important assumption that animal detection is perfect at the apex of detection, usually on the transect line (distance = 0). If this assumption is true, Burnham et al. [6] show that CDS models under mild assumptions are robust to biases associated with unmodeled heterogeneity, a condition Burnham et al. [7] labeled “pooling robust”. By extension, MCDS models would also be pooling robust if the assumption of perfect detection at the apex were true.

In many cases the perfect detection assumption is unrealistic or suspect. In such cases the collection of double-observer data can be used to remove this assumption. Double-observer data consists of 2 observers trying to detect animals independently of one another. When one observer detects an animal, it is essentially “marked” for the other observer to either detect (recapture) or miss. The incorporation of double-observer data into distance modeling (Mark-Recapture Distance Sampling models, MRDS models) eliminates the assumption of perfect detection at a distance associated with the detection apex and allows this parameter to be estimated [8–11] and even vary due to other variables [11]. The initial MRDS models [8–11] assume independent animal sightings between the observers over all distances, which Borchers et al. [12] call the “full independence” (FI) assumption. While there are several ways to estimate MRDS models under this assumption, a common method is to use a Huggins type [13] mark-recapture model with distance as a covariate [14]. The property of “pooling-robustness” does not hold for MRDS models [6]. Like mark-recapture models, unmodeled heterogeneity in MRDS models is a source of bias [12]. Covariates can reduce the unmodeled heterogeneity but are unlikely to account for all the sources of heterogeneity [15].

Laake [16] analyzed distance data on a known population of 150 wooden stakes, the best CDS estimate was 81% (121) of the true value, while the MRDS model using the FI assumption was 80% (120) of the true value. Additionally, analysis of crabeater seal (*Lobodon carcinophaga*) [12,14] and feral horse (*Equus caballus*) [15] datasets with a MRDS model using the FI assumption resulted in estimates that were considerably lower than those obtained with MCDS models that assume perfect detection on the transect line (the location of maximum detection for these data). These results do not follow the expected pattern of MRDS population estimates being larger than CDS and MCDS estimates. These results can be caused by statistical dependence in the data, even though the 2 observers collected the data independently. The dependence can easily be a function of distance or animal activity. Increased distance to the animal will cause more and more animals to be missed. If at far distances only the most visible animals are detected then they may be more likely to be detected by both observers since they are more visible than the average animal at that distance. This occurs because missing variables explaining detection have not been accounted for; this is unmodeled heterogeneity. In the above case, an animal standing on a ridge top with no ground cover or vegetation in the background (sky-lined), would be very visible at far distances to both observers; the inadequate modeling of the sighting conditions would cause unmodeled heterogeneity. The effect of vegetation obstruction may increase with distance, resulting in the situation described above. Laake et al. [15] document a divergence in detection probabilities over distance obtained from a MCDS model and the mark-recapture probabilities in a feral horse dataset. The higher mark-recapture probabilities at far distances are consistent with unmodeled heterogeneity causing some dependency in the sighting data. Laake et al. [15] state: “it is important to understand that independence is not solely achieved by the observers being unaware of each other.”

There are several ways to deal with this dependence issue. Buckland et al. [17] used a limited independence model that simultaneously modeled the distance data, the mark-recapture data and the dependence between the 2 observers as a function of distance as well as other covariates. Conn et al. [18] modeled observer detections as arising from a probit-Bernoulli model with the correlation between the observers being a function of distance. Laake [16] presented a less restrictive independence assumption that only assumed independence of observer observations at the distance associated with the apex of detection. Borchers et al. [12] denoted these as “point independence” (PI) assumption models. MRDS models that use a point independence assumption fit a MCDS model and use a Mark-recapture model to estimate maximum detection at the MCDS detection apex, this is then used to adjust the height of the MCDS model

[12,16]. Borchers et al. [12] document the mathematical details of a MRDS model under the PI assumption.

Line transect surveys are an excellent tool for obtaining population estimates for large geographic areas. In terrestrial study areas, aircraft are often used to fly transects over large study areas. Assuming perfect detection at some distance from a fast-moving aircraft is unrealistic and requires the use of a MRDS model to estimate animal detection. Due to costs, both observers are usually on the same aircraft which makes it more likely that unmodeled heterogeneity of the sighting conditions will occur. Under these conditions a MRDS model using the PI assumption will be less biased than one that relies on the FI assumption [12, 16]. Large terrestrial study areas often contain mountainous terrain over which it is unsafe to fly straight-line transects; however, these areas can safely be flown using contour transects [19–20]. For contour transects, distance is measured as the closest distance between the transect and the sighted animal using GPS locations [19–21]. Becker and Quang [20] encountered unimodal detection data of brown bears sighted from transects flown with small fixed-wing aircraft. The main reason for this unimodal detection shape is that detection is degraded close to the aircraft due to less time available for searching caused by the plane’s speed and fewer sighting angles around vegetative obstructions, whereas at moderate distances the increased sighting angles and viewing time improve detection despite the negative effect of increased distance. A more detailed explanation is available on pages 209 and 220 of Becker and Quang [20]. They used a unimodal detection function coupled with double-count data modeled under the FI assumption to obtain a population estimate. They also modeled the scale parameter as a function of covariates similar to Marques and Buckland [4]. In the gamma detection model, the addition of covariates results in a different detection apex for each covariate level [20]. The gamma detection function when coupled with double-count data is inconsistent with a MRDS model utilizing the PI assumption, since the covariates cause multiple apexes at various distances and under the PI assumption only 1 distance is assumed to have point independence [12].

The purpose of this paper is to implement a MRDS model for unimodal detection data that does not require independent observations over all distances between the 2 observers. Our solution will involve developing a new detection model that will allow us to implement Borchers et al. [12] MRDS model with the PI assumption. We propose a new detection model that has a unimodal shape with only one detection apex, even when multiple covariates are used to model detection. Additionally, a unimodal model for the double-observer (mark-recapture) data will be presented. A MRDS model using the PI assumption allows for less biased and more realistic population estimates from line transect data which have a unimodal detection shape and avoids the unrealistic assumption of perfect detection [12,16]. If there is no unmodeled heterogeneity then the population estimates are unbiased [12]. As an example, these models will be applied to double-observer, aerial line transect survey of black bears (*Ursus americanus*).

Methods

Borchers et al. [12] partition the likelihood for MRDS data into 4 components: a distance component ($\mathcal{L}_{y|z}(\underline{\theta}_{y|z})$), a mark-recapture component ($\mathcal{L}_{\omega}(\underline{\theta}_{\omega})$), a binomial component ($\mathcal{L}_{n\bullet}(N_c, \underline{\theta}, \underline{\phi}_z)$), and a covariate component ($\mathcal{L}_z(\underline{\phi}_z, \underline{\theta})$) Using their partial likelihood approach, estimates of $\underline{\theta}$, ($\underline{\theta} = \{\underline{\theta}_{y|z}, \underline{\theta}_{\omega}\}$) can be obtained from ($\mathcal{L}_{y|z}(\underline{\theta}_{y|z})$), ($\mathcal{L}_{\omega}(\underline{\theta}_{\omega})$), and a Horvitz-Thompson like estimator is used to obtain the population estimate. Our focus is on aerial surveys which often have blind strip of width (w_b), we use the notation of Borchers et al. [12] with the lower limit denoted by w_b throughout this paper.

Distance Data

Assume random transect placement and truncation of the distance data at w . Let y represent the closest distance between the transect and the sighted animal, \underline{z} represent a vector of covariates that affect the detection probability, and i index the sighted animal ($i = 1, 2, \dots, n$). Define $p_{\bullet}(y_i, \underline{z}_i)$ as the probability that an animal at distance y_i with covariate vector \underline{z}_i is detected by at least one observer in the (w_b, w) strip. Denote the pdf of y as $\pi(y) = \frac{1}{(w-w_b)}$.

Borchers et al. [12] give the general form of the distance likelihood as:

$$\mathcal{L}_{y|z}(\theta_{y|z}) = \prod_{i=1}^n \frac{p_{\bullet}(y_i, \underline{z}_i)\pi(y)}{p_{\bullet}(\underline{z}_i)}, \tag{1}$$

where

$$p_{\bullet}(\underline{z}_i) = \int_{w_b}^w p_{\bullet}(y_i, \underline{z}_i)\pi(y)dy, \tag{2}$$

is the probability that at least one observer detects animal i .

Our interest is to develop a MRDS model that is consistent with the PI assumption of independent sighting data between observers only at some distance $y = y^*$. This requires a unimodal detection function that also allows the use of detection covariates. The two-piece normal distribution [22], also known as a split-normal distribution, is a simple distribution that meets this requirement. The two-piece normal distribution is essentially 2 half-normal distributions that share a common mode (μ) but have different variances. Using the terminology of Marques and Buckland [4] the key function is:

$$p_{\bullet}(y_i, \underline{z}_i) = \begin{cases} \exp\left[-\left(\frac{y_i - \mu}{\sqrt{2}\sigma_{1i}}\right)^2\right] & \text{for } w_b \leq y_i < \mu \\ \exp\left[-\left(\frac{y_i - \mu}{\sqrt{2}\sigma_{2i}}\right)^2\right] & \text{for } y_i \geq \mu, \end{cases} \tag{3}$$

where μ denotes the mode of the detection distribution and $\mu = y^*$ under the PI assumption.

Assuming the covariates, \underline{z}_i , affect detection via the scale parameters $(\sigma_{1i}, \sigma_{2i})$ [4],

$$\sigma_{1i} = \exp\left(\beta_0 + \sum_{k_1} \beta_{k_1} z_{k_1i} + \sum_{k_2} I(w_b \leq y_i < \mu) \beta_{k_2} z_{k_2i}\right), \text{ and} \tag{4}$$

$$\sigma_{2i} = \exp\left(\beta_0 + \sum_{k_1} \beta_{k_1} z_{k_1i} + \sum_{k_3} I(y_i \geq \mu) \beta_{k_3} z_{k_3i}\right). \tag{5}$$

The covariates that affect the scale of both half-normal curves are modeled by the $\sum_{k_1} \beta_{k_1} z_{k_1i}$ terms, the covariates that affect only σ_{1i} are modeled by the $\sum_{k_2} \beta_{k_2} z_{k_2i}$ terms, and the covariates that affect only σ_{2i} are modeled by the $\sum_{k_3} \beta_{k_3} z_{k_3i}$ terms. The total number of covariates in the model is k , ($k = k_1 + k_2 + k_3$). The $I(\)$ notation is a standard indicator function that is 1 when the condition in the parentheses is true and 0 otherwise, the purpose of this notation is to make explicit which scale parameters are being modelled.

MCDS parameter estimates are obtained by maximizing the log-likelihood equation:

$$\begin{aligned} \ell_{y|z}(\hat{\theta}_{y|z}) &= \sum_{i=1}^n [\ln(p_{\bullet}(y_i, z_i)) + \ln(\pi(y)) - \ln(p_{\bullet}(z_i))] \\ &= \sum_{i=1}^n \left[- \left(\frac{y_i - \hat{\mu}}{\sqrt{2}[I(w_b \leq y_i < \hat{\mu})\hat{\sigma}_{1i} + I(y_i \geq \hat{\mu})\hat{\sigma}_{2i}]} \right)^2 \right] \\ &\quad - \sum_{i=1}^n \ln \left(\int_{y=w_b}^{\hat{\mu}} \exp \left[- \left(\frac{y_i - \hat{\mu}}{\sqrt{2}\hat{\sigma}_{1i}} \right)^2 \right] dy \right) - \sum_{i=1}^n \ln \left(\int_{y=\hat{\mu}}^w \exp \left[- \left(\frac{y_i - \hat{\mu}}{\sqrt{2}\hat{\sigma}_{2i}} \right)^2 \right] dy \right). \end{aligned} \tag{6}$$

Numerical methods for maximizing Eq (6) above require initial starting values for the parameters. We follow Ramsey et al. [23] to obtain initial starting values for the $\hat{\beta}_k$'s, and use the apex of a kernel estimator [24] for the initial starting value of $\hat{\mu}$.

Double-Observer Data

Our focus is the implementation of a point independence MRDS model, which will utilize the results of a two-piece normal distance model in order to obtain a population estimate. Following Borchers et al. [12] the mark-recapture likelihood is:

$$\mathcal{L}_{\omega}(\theta_{\omega}) = \prod_{i=1}^n \frac{\Pr(\omega_i | y_i, z_i)}{p_{\bullet}(y_i, z_i)}. \tag{7}$$

Here ω_i denotes the encounter history of the i^{th} animal ($\omega_i \in \{(1,0), (0,1), (1,1)\}$); the first number in the history indexes whether the animal was seen by the 1st observer (0 = no, 1 = yes) and the second number indexes whether the animal was seen by the 2nd observer.

Borchers et al. [12] denote the degree of dependence between the 2 observers as $\delta(y, z)$, σ_{12} as the covariance between the 2 observers' detection probabilities that is induced by excluding covariates u from the model, and $p_j(y, z) = E_u[p_j(y, z, u)]$ for observer $j, j = 1, 2$;

$$\delta(y, z) = 1 + \left(\frac{\sigma_{12}(y, z)}{p_1(y, z)p_2(y, z)} \right), \tag{8}$$

$$\sigma_{12}(y, z) = Cov_u[p_1(y, z, u), p_2(y, z, u)]. \tag{9}$$

The calculation of $p_{\bullet}(y_i, z_i)$ requires knowledge of $\delta(y_i, z_i)$ which cannot be obtained from the capture history [12],

$$p_{\bullet}(y_i, z_i) = p_1(y_i, z_i) + p_2(y_i, z_i) - p_1(y_i, z_i)p_2(y_i, z_i)\delta(y_i, z_i). \tag{10}$$

They show that using conditional probabilities under the PI assumption allows for the estimation of $p_{\bullet}(y_i, z_i)$. Without making any independence assumptions:

$$p_{\bullet}^c(y_i, z_i) = p_{1|2}(y_i, z_i) + p_{2|1}(y_i, z_i) - p_{1|2}(y_i, z_i)p_{2|1}(y_i, z_i), \tag{11}$$

$$p_{\bullet}^c(y_i, z_i) = p_{\bullet}(y_i, z_i)\delta(y_i, z_i). \tag{12}$$

Assuming point independence (PI) at y^* , we assume $\delta(y^*, z_i) = 1$, so

$$p_{\bullet}^c(y^*, z_i) = p_{\bullet}(y^*, z_i)\delta(y^*, z_i) = p_{\bullet}(y^*, z_i). \tag{13}$$

Their approach to fitting a point independence MRDS model requires a logistic model that contains different observers (1 and 2), and other possible covariates to obtain estimates of $\hat{p}_{1|2}(y_i, z_i)$, and $\hat{p}_{2|1}(y_i, z_i)$. A linear form of distance (y) was a potential explanatory variable in their logistic models. In addition to the linear form of distance as a potential covariate, we also wanted to consider other forms that could mimic the general shape of the two-piece normal detection model. We used a polynomial spline [25] on the distance variable as a possible covariate to accomplish this. We placed no restrictions on the apex location of the spline fit. Depending upon the nature of the unmodeled heterogeneity and other variables being considered in the model, distance may not be an important covariate [15] or have a unimodal shape. Our MRDS model uses a two-piece normal detection function to model the distance data, so we assume point independence at the detection apex $y^* = \hat{\mu}$ and obtain $\hat{p}_{1|2}(\hat{\mu}, z_i)$, and $\hat{p}_{2|1}(\hat{\mu}, z_i)$ from our logistic model to calculate $\hat{p}_\bullet(\hat{\mu}, z_i)$,

$$\hat{p}_\bullet(\hat{\mu}, z_i) = \hat{p}_{1|2}(\hat{\mu}, z_i) + \hat{p}_{2|1}(\hat{\mu}, z_i) - \hat{p}_{1|2}(\hat{\mu}, z_i)\hat{p}_{2|1}(\hat{\mu}, z_i). \tag{14}$$

Population Estimation

Following Borchers et al. [12], let N_c be the number of animals in the covered region, which is estimated by using a Horvitz-Thompson like estimator:

$$\hat{N}_c = \sum_{i=1}^n \frac{s_i}{\hat{\pi}_i}, \tag{15}$$

where s_i denotes the size of the i^{th} animal group, and $\hat{\pi}_i$ denotes the estimated inclusion probability of the i^{th} group and is estimated by:

$$\hat{\pi}_i = \hat{p}_\bullet(\hat{\mu}, z_i)\hat{p}_\bullet(z_i) \tag{16}$$

[12]. The $\hat{p}_\bullet(\hat{\mu}, z_i)$ term is calculated from the mark-recapture data (Eq (14)) and the $\hat{p}_\bullet(z_i)$ term is calculated from the distance data (Eq (2)). Transects were randomly placed so inferences about animal density and population size can be made for the entire study area.

Laake and Borchers [14] have shown that averaging the $\hat{p}_\bullet(\hat{\mu}, z_i)$'s over the observed covariates is also valid and can be used to calculate the estimated inclusion probability. We denote this average as $\hat{p}_\bullet(\hat{\mu})$, and followed equation 6.50 of Laake and Borchers [14]:

$$\hat{p}_\bullet(\hat{\mu}) = \frac{\sum_{i=1}^n \hat{p}_\bullet(\hat{\mu}, z_i) / \hat{E}(\hat{p}_\bullet(z_i))}{\sum_{i=1}^n 1 / \hat{E}(\hat{p}_\bullet(z_i))}, \tag{17}$$

where:

$$\hat{E}(\hat{p}_\bullet(z_i)) = \int_{w_b}^w \hat{p}_\bullet(y, z_i)\pi(y)dy. \tag{18}$$

The estimated inclusion probabilities are now calculated by:

$$\hat{\pi}_i = \hat{p}_\bullet(\hat{\mu})\hat{p}_\bullet(z_i). \tag{19}$$

The variance can be obtained following Borchers et al. [12] or by using a transect-based bootstrap method. We used the lowest AIC [1] to determine the best MCDS model and the best logistic model. Marques and Buckland [4] report that estimators based on estimated Horvitz-Thompson inclusion probabilities can have substantial bias if many of the estimated

inclusion probabilities are small. Marques and Buckland [5] recommend restricting model selection to models whose estimated inclusion probabilities are all greater than 0.1 and less than 5% of them are below 0.2; we excluded models that did not meet these criteria from our AIC selection process. Assessing the fit of the MCDS model includes a Goodness-of-Fit test [1], q-q plots, and a Kolmogorov-Smirnov test [6]. Model fit for the logistic model and the MRDS model follow Borchers et al. [12].

Application to a South-Central Alaska Black Bear Population

Data collection involved aerial observation of black bears (*Ursus americanus*), for which no agency field permit or approval was required. All Federal Aviation Administration regulations were followed in collecting the data. No lands were accessed in the data collection, nor were any endangered or protected species sampled. No animals were handled during the survey, so no IACUC permit (a permit that is required to ensure the humane treatment and handling of captured animals) was required.

In the spring of 2007, an aerial line transect survey for black bears was conducted in the 21,398 km² Skwetna study area in South-Central Alaska. This area is bordered on the south by the Chakachamna Lake, Chakachamna River and Cook Inlet; on the east by the Susitna and Yentna rivers; and on the north and west by the spine of the Alaska Range. Elevation ranged from 0 to 6,196 meters. Prominent habitats consisted of: birch and mixed birch-spruce forests, and muskegs at lower elevations; mixtures of grasslands and alder on the lower mountain slopes; alpine habitat at intermediate elevations; and rocky terrain with variable snow cover and glaciers at higher elevations. The survey was conducted between 10–25 May, a period after black bear den emergence and prior to full leaf-out of the birch forest.

A total of 470 30-km long transects were flown by 5 small tandem-seat aircraft that contained 2 observers: the pilot and a backseat observer. No black bears were encountered above 1,067 meters and terrain above this elevation was not considered black bear habitat. The 14,895 km² of black bear habitat were sampled by 441 of the 470 flown transects. Roughly half of the bear habitat was too mountainous for straight-line transects and required contour transects [19–20]. Straight-line transects were hinged in the middle with a random angle between 30 and 330 degrees, and this facilitated efficient sampling of irregular shaped subareas that were suitable for “straight-line” transects (Fig 1). Bear searches were restricted to one-side of the transect to facilitate the collection of double-count data. Searches along contour transects were conducted on the uphill side, while a random side was used for straight-line transects. Other than using hinged straight-line transects, survey methodology follows the details given in Becker and Quang [20]. Possible covariates included: bear activity (bedded, sitting, standing, moving), group size, percent cover within 10 m of the bear, percent snow cover within 10 m of the bear, plane speed, search distance, and transect type (straight, contour). Percent cover and percent snow covariates were divided by 20, and air speed was divided by 100 to increase numerical efficiency of the optimizer used to obtain maximum likelihood estimates. Search distance represented how far out the observers were searching at the time the bear was detected and was measured by overflying that location and obtaining a GPS location, from which a distance to the transect was calculated. The covariate $\text{Bin}.\ln\text{SD}350$ represents the search distance covariate which was truncated at w , transformed to the natural log scale, and binned to into 9 classes. The truncation and binning was done to minimize the correlation between this covariate and distance.

A total of 260 black bear groups were detected from these transects. The distance data was truncated at 350 m ($w = 350$), in addition there was a 22 m blind spot underneath the plane, so $\pi(y) = \frac{1}{(350-22)}$. Within this 22–350 m strip, a total of 235 bear groups were detected, so 9.6% of the data were trimmed off. The distance data for these detections exhibited a unimodal

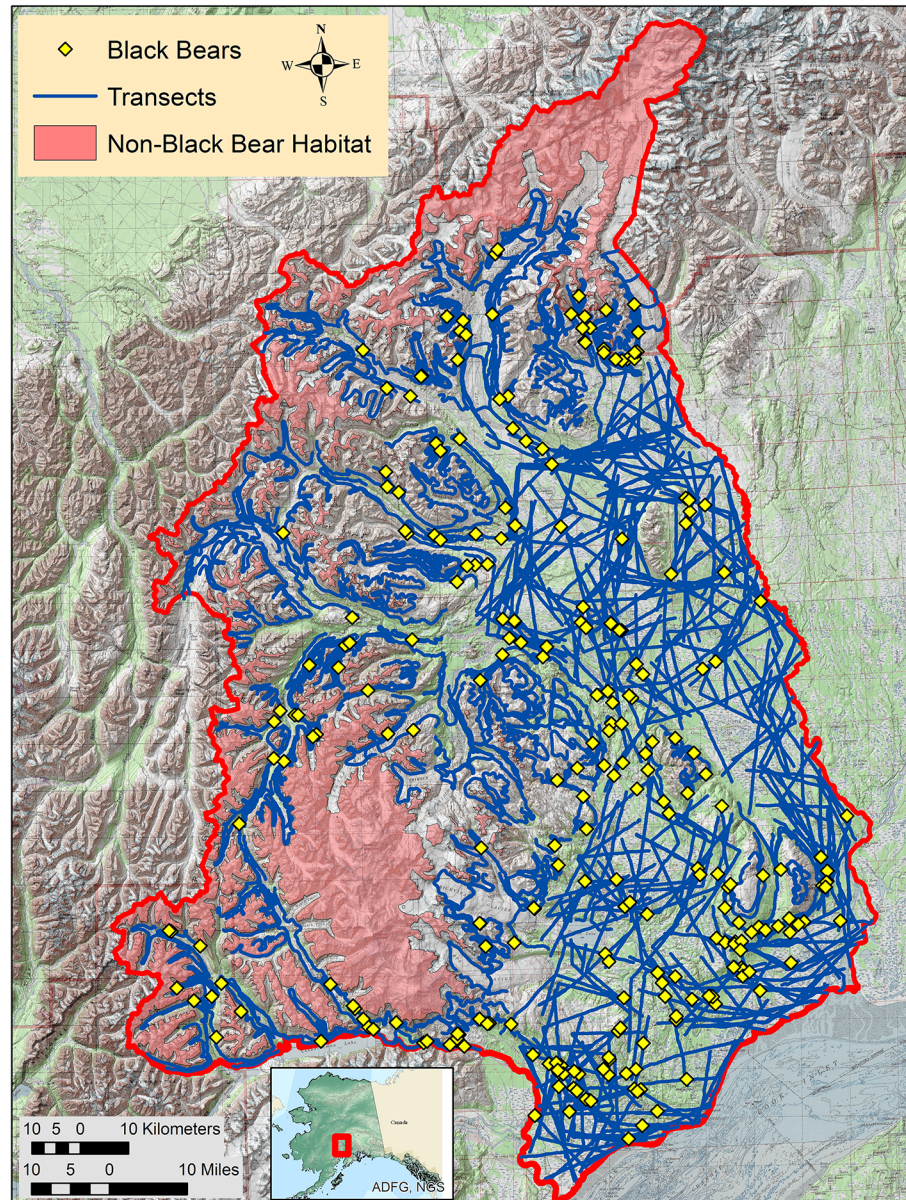


Fig 1. The Skwetna study area of South-Central Alaska that was used in a 2007 aerial distance sampling survey of black bears. The study area is enclosed by the red border and Cook Inlet to the southeast. Red-shaded areas within the study area denote locations that were above 1,067 m in elevation that were classified as non-black bear habitat. Yellow triangles denote the location of detected black bear groups. The blue lines denote the actual flown transects, with contour transects flown in mountainous terrain and straight transects that contain a midpoint hinge being flown in the lower-elevation, flat terrain.

doi:10.1371/journal.pone.0136403.g001

distribution and were modeled using a two-piece normal detection function. The R-code [26] used to obtain the parameter estimates, standard errors, confidence intervals, graphs, and fit diagnostics are listed in appendix 1. Standard errors and confidence intervals were based on Borchers et al. [12] variance formula. Based on AIC, the best MCDS model (Fig 2) included covariates for transect type (Fig 3) and percent cover (Fig 4) which affect both scale parameters (σ_1, σ_2). The Bin.lnSD350 covariate was included in modeling the σ_{2i} scale parameter (Fig 5). The mode ($\hat{\mu}$) was estimated to be 102.61 meters (se = 1.013). This analysis ignored observer

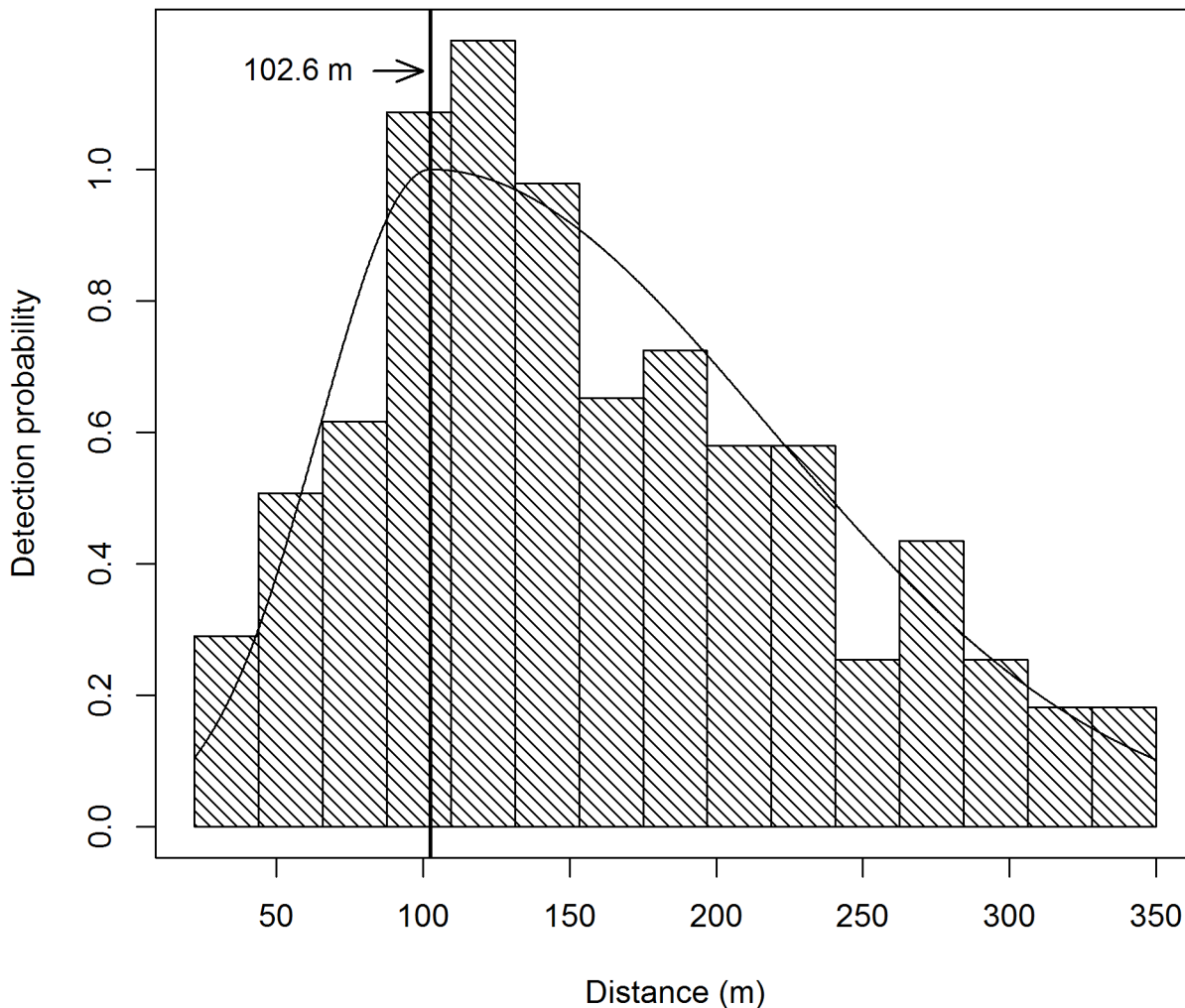


Fig 2. Two-piece normal detection model for black bears in Skwetna study area of South-Central Alaska. The detection curve is for averaged covariate values of binned search distance, percent cover, and transect type. The vertical line at 102.6 m depicts $\hat{\mu}$. The histogram is of the observed black bear group detection distances.

doi:10.1371/journal.pone.0136403.g002

type (pilot or backseat observer) which resulted in a detection model for the survey plane. Parameter estimates and standard errors are given in [Table 1](#). A Goodness of Fit test for the MCDS model using 15 bins resulted in a Chi-sq. statistic of 7.742, $df = 8$, $p = 0.459$. A Kolmogorov-Smirnov test ($p = 0.843$) and a q-q plot ([Fig 6](#)) revealed no abnormalities. Collectively, these results indicate that the two-piece normal detection model provided an adequate fit to the data.

The Double-Observer model was fit using the R-package *mrds* [26, 27]. In addition to the fixed observer effect, the covariates that were considered in the MCDS model for this data were also considered here. Based on AIC [28] (AIC = 3064.3), the best Double-Observer model included a 4th order polynomial spline of distance, percent cover, air speed, bear activity of bedded/sitting, and observer. Parameter estimates are given in [Table 2](#) and a graph of the average covariate values are given in [Figs 7](#) and [8](#) for the pilot and backseat observer models respectively. A Goodness of Fit test resulted in a Chi-sq. value of 14.776, $df = 11$, $p = 0.193$. At 102.61 m ($\hat{\mu}$), the model resulted in an estimate of 0.926 (se = 0.038) for $\hat{p}_\bullet(\hat{\mu})$.

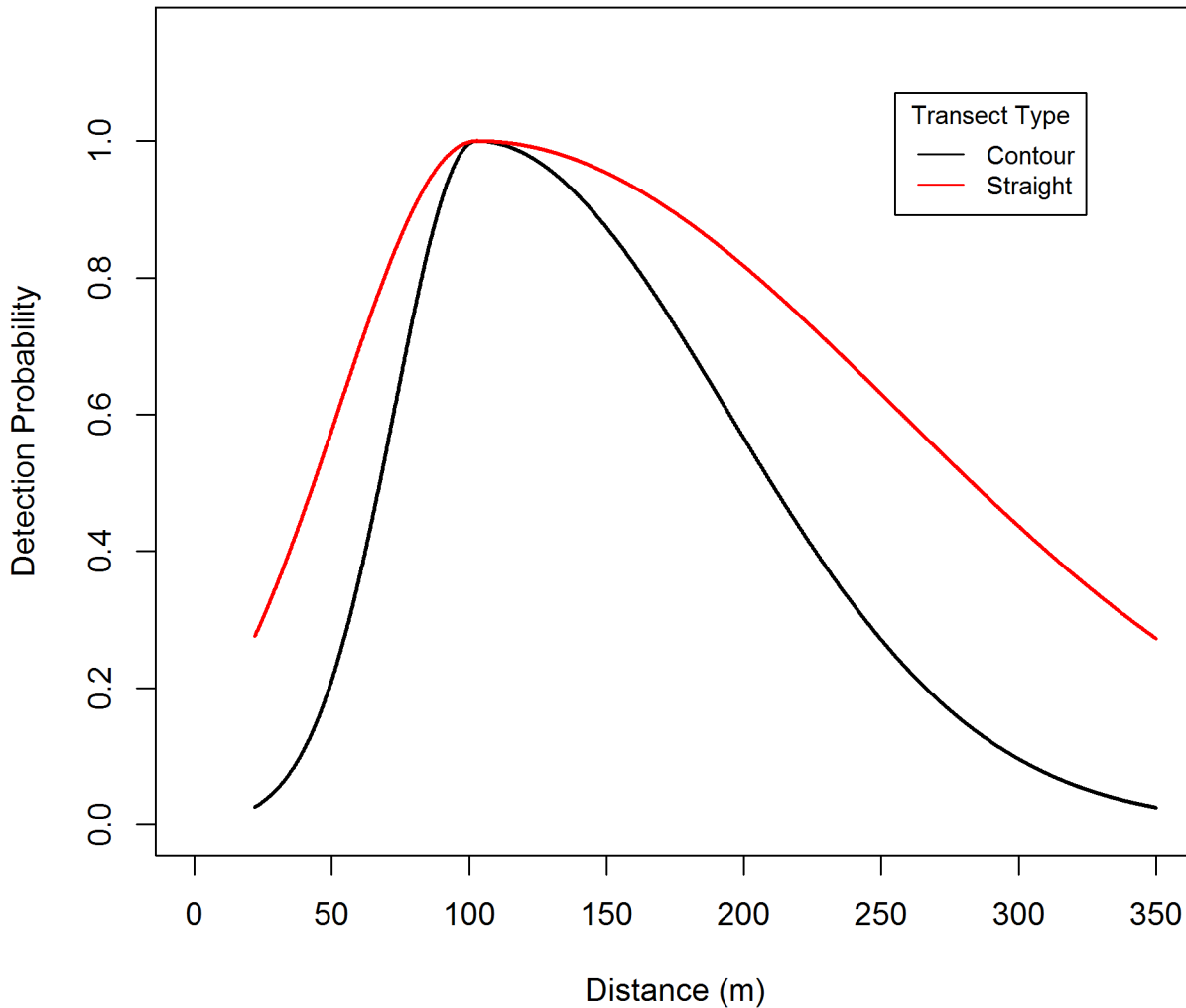


Fig 3. The effect of transect type on average covariate detection of black bears in the Skwetna study area of South-Central Alaska.

doi:10.1371/journal.pone.0136403.g003

The estimated inclusion probabilities from our best MRDS model ranged from 0.137 to 0.790, with a median value of 0.555 (Fig 9). Only 2.98% of the estimated probabilities were below 0.2, and all were within the limits suggested by Marques and Buckland [5]. Our MRDS model resulted in a population estimate of 2,377.0 black bears (se = 185.29, CV = 7.80%) and a 95% confidence interval of (1990.9, 2838.1) in the study area. Converting to density (bears/1000 km²) results in a point estimate of 159.58 (se = 12.44) with a 95% confidence interval of (133.7, 190.5).

Discussion

In our best MRDS model of the black bear dataset, distance was an important covariate and was fit using a 4th degree polynomial spline. Our model did not restrict the apex locations of the MCDS and MRDS models to be the same. The apexes, 102.6 and 110.8 m respectively, differed by 8.2 meters. The main reasons the apexes were not restricted to be the same were: 1) distance may not be included in the best MRDS model, 2) computational complexity, and 3) if distance were in the MRDS model, unmodeled heterogeneity in the mark-resight data may

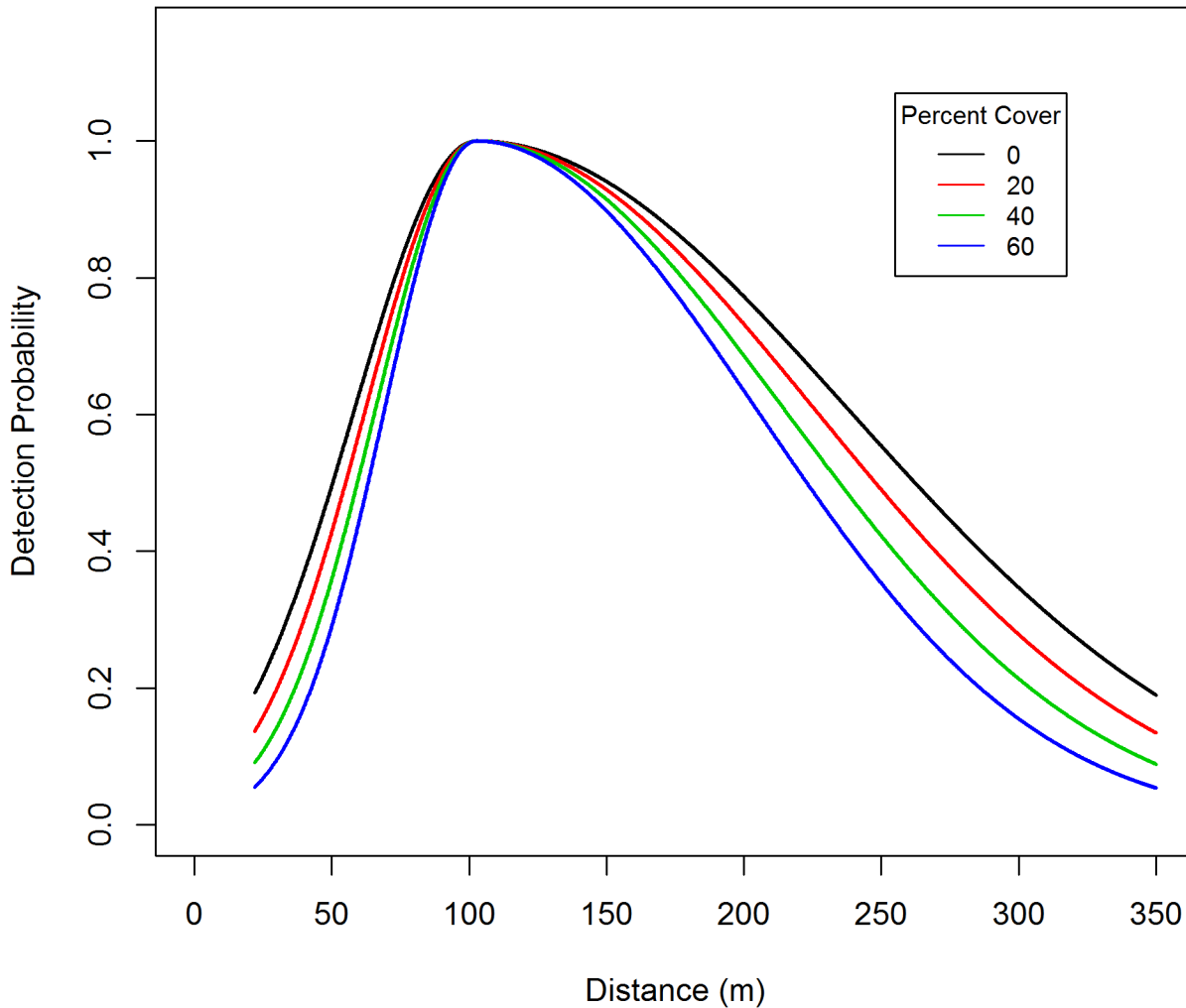


Fig 4. The effect of vegetative cover on average covariate detection of black bears in the Skwetna study area of South-Central Alaska.

doi:10.1371/journal.pone.0136403.g004

cause the 2 apex locations to differ. In our black bear example, the difference in the estimates of $\hat{p}_\bullet(\hat{\mu})$ is negligible between these apex locations (0.926 and 0.928 respectively).

Besides goodness of fit tests to assess model fit, we also required the estimated inclusion probabilities obtained from the model to meet the recommendations of Marques and Buckland [5]. This additional criterion ensures that the population estimate is not being driven by a few observations with very low estimated inclusion probabilities. An analogous situation exists for traditional Horvitz-Thompson estimators [29] (where the inclusion probabilities for the observations are known, not estimated). The inclusion probabilities are contained in the denominator for both the point estimate and the variance. If very small inclusion probabilities are encountered then the variance becomes very large and the point estimate is driven by these few probabilities. In traditional sampling, care is taken to avoid small inclusion probabilities, and optimal sample designs have the inclusion probabilities proportional to size [29]. The recommendation of Marques and Buckland [5] formalizes the avoidance of this problem with their criteria of no estimated inclusion probabilities lower than 0.1, and less than 5% falling in the 0.1 to 0.2 range.

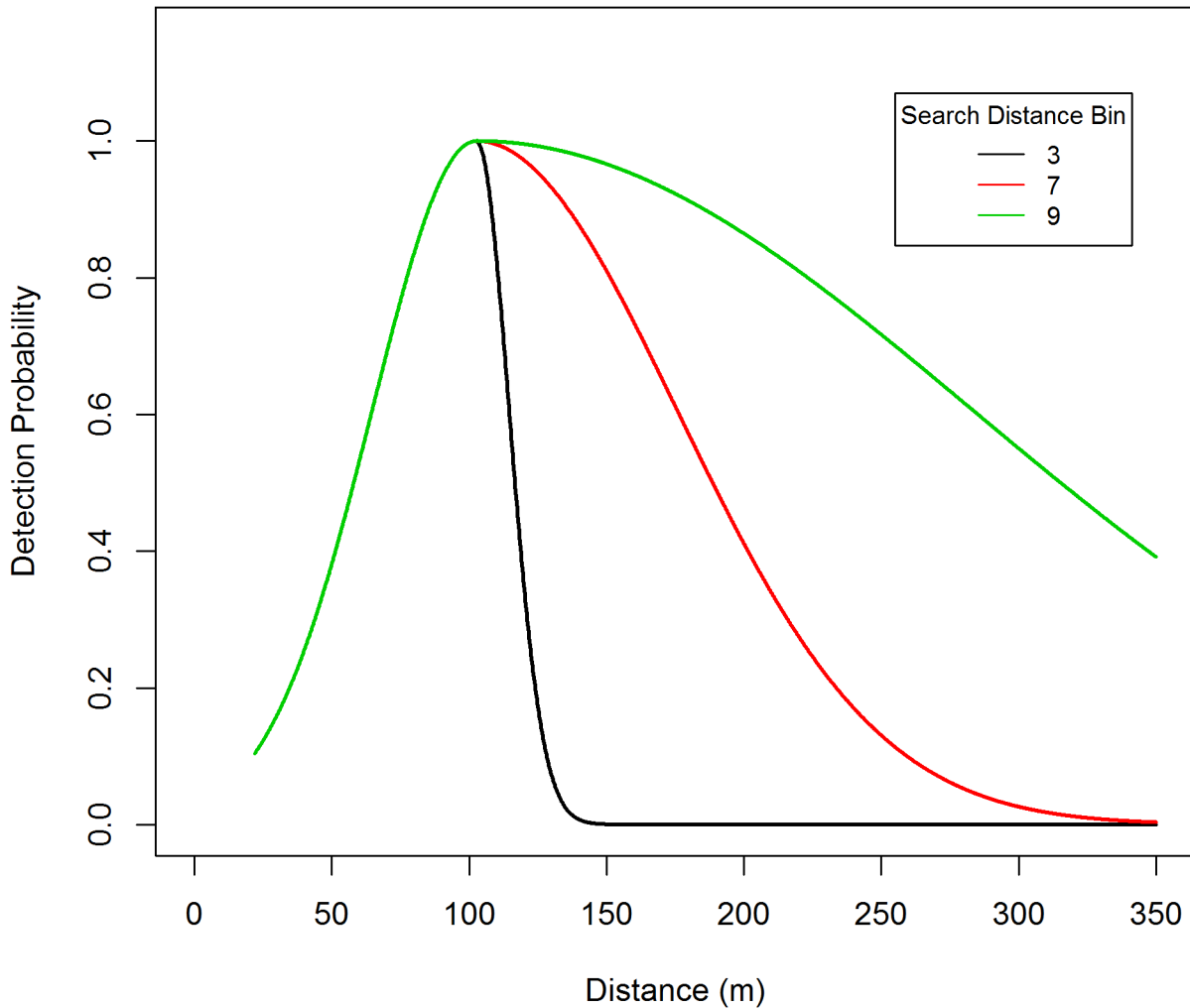


Fig 5. The effect of binned search distance on average covariate detection of black bears in the Skwetna study area of South-Central Alaska.

doi:10.1371/journal.pone.0136403.g005

Our black bear example illustrates the utility of distance sampling to make inferences over large study areas. Areas of management interest are often so large that aerial distance sampling is one of the few logistically feasible options. Double-count modeling is needed since the assumption of perfect detection at the detection apex would not be tenable without supporting

Table 1. Coefficients and standard errors for the best fitting multiple covariate distance sampling model using a two-piece normal detection model. The parameter notation Bin.InSD denotes the 9 binned values of the natural log of truncated search distance (SD), for which $\ln(350)$ m was substituted for $\ln(\text{SD})$ if the SD value was beyond 350 m (w).

| Parameter | Coefficient | SE |
|--------------------------|-------------|--------|
| Intercept | 3.5549 | 0.0644 |
| $I(y > \mu)$ | -2.5125 | 0.6103 |
| $I(y > \mu)$ Bin.InSD350 | 0.4526 | 0.0060 |
| Transect (Flat) | 0.5199 | 0.0358 |
| Percent Cover | -0.0944 | 0.0044 |
| $\ln(\mu)$ | 4.6309 | 0.0128 |

doi:10.1371/journal.pone.0136403.t001

QQ Plot

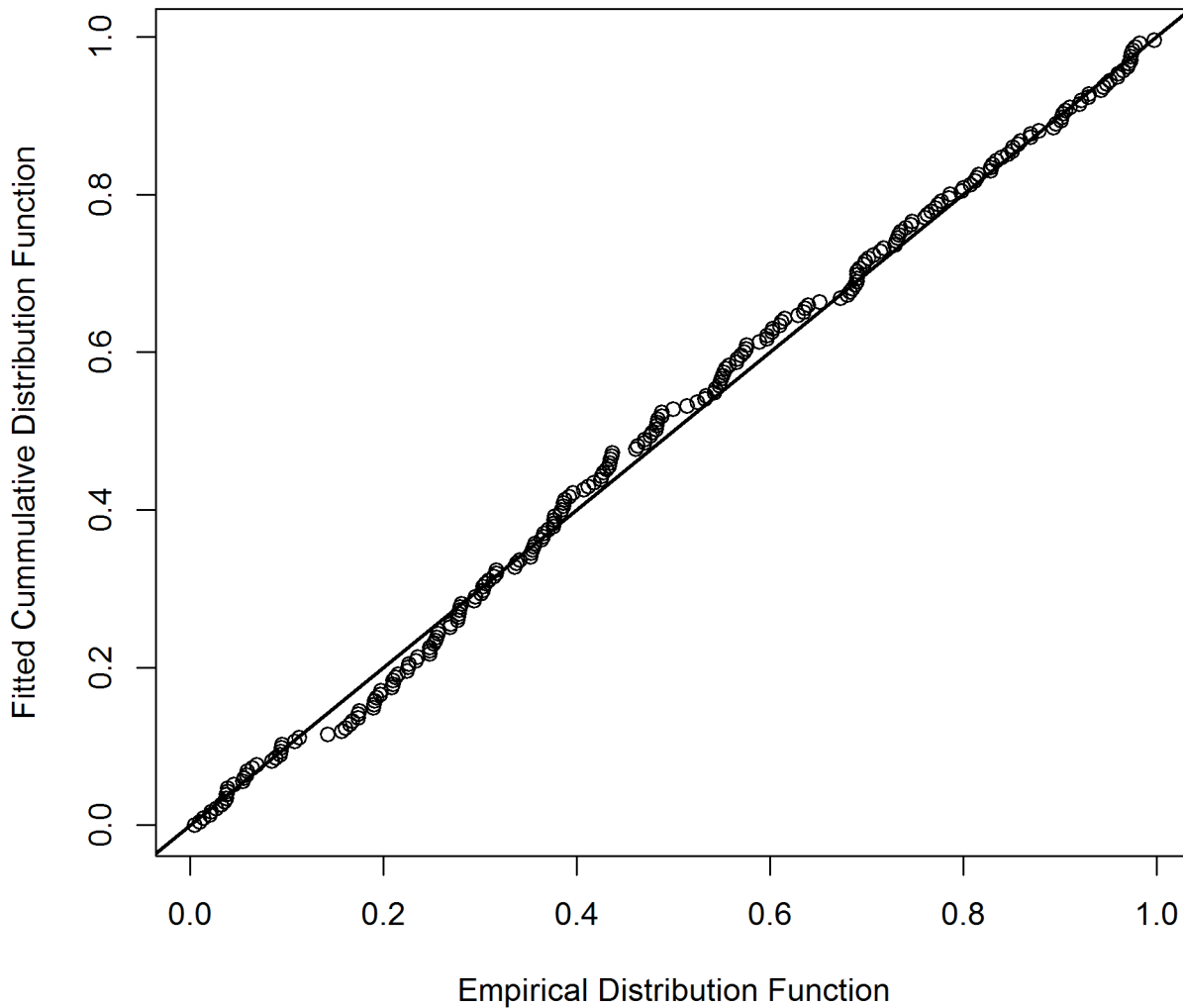


Fig 6. Q-q plot of the multiple covariate distance model fit.

doi:10.1371/journal.pone.0136403.g006

Table 2. Coefficients and standard errors for the best fitting double observer model which included a quadratic spline over distance. The 4-bs(distance, degree = 4) parameters denote the 4 parameters of the 4th order polynomial spline of distance.

| Parameter | Coefficient | SE |
|---------------------------|-------------|--------|
| Intercept | 2.2843 | 1.6794 |
| bs(distance, degree = 4)1 | 9.1280 | 3.2812 |
| bs(distance, degree = 4)2 | -2.4780 | 2.7692 |
| bs(distance, degree = 4)3 | 5.7301 | 3.5575 |
| bs(distance, degree = 4)4 | 0.8770 | 1.6418 |
| Observer (backseat) | -1.1690 | 0.2024 |
| Percent Cover | -0.3819 | 0.1356 |
| Air Speed | -2.8225 | 1.5882 |
| Bedded/Sitting | -0.5289 | 0.2960 |

doi:10.1371/journal.pone.0136403.t002

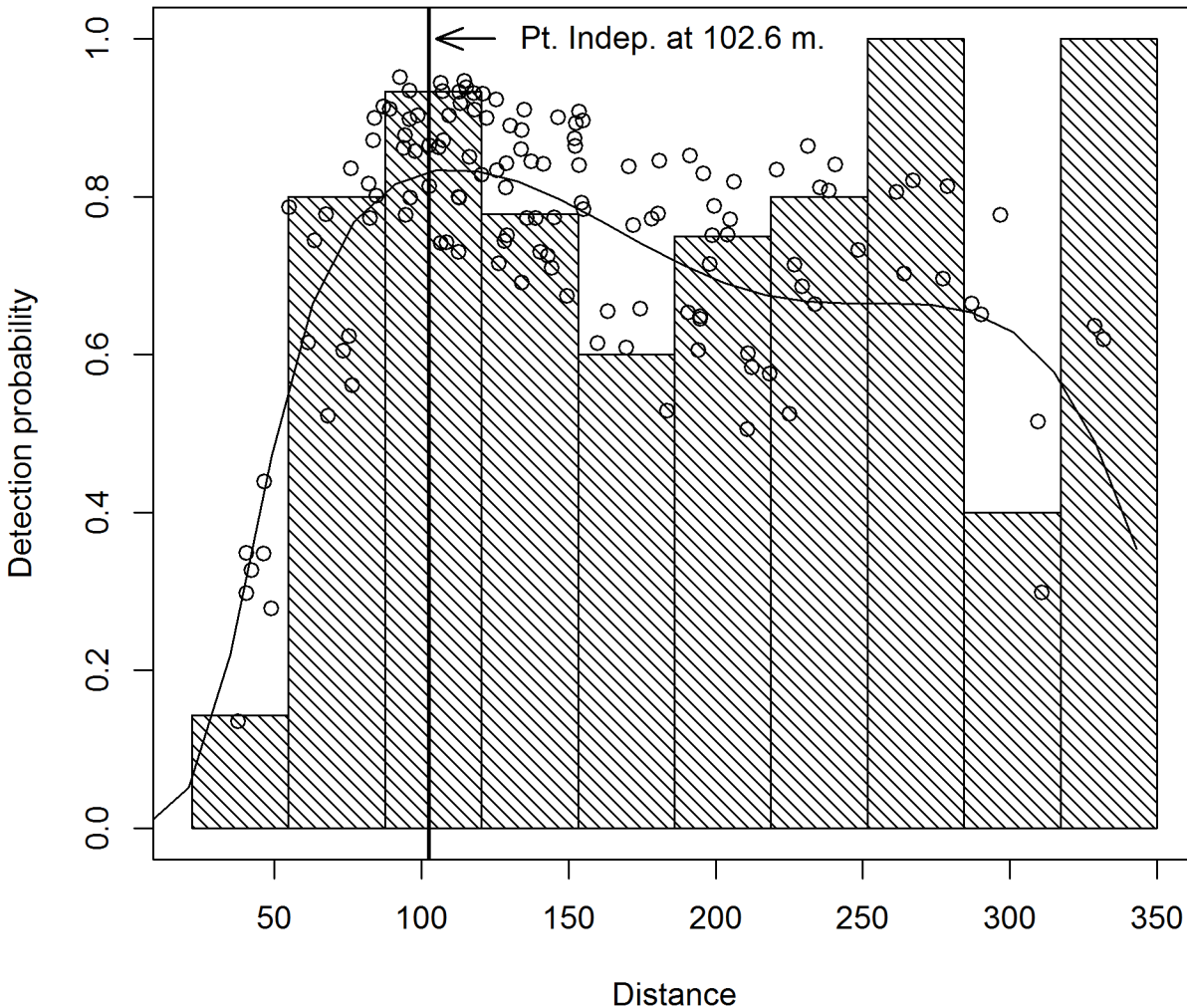


Fig 7. Logistic model of detection probabilities of the pilot given the backseat observer detected a black bear. The solid line denotes the conditional detection probability for the average values of the distance, percent cover, observer, air speed, and bedded/sitting bear covariates. The circles represent the model estimated conditional probabilities for each bear, based on their observed covariates.

doi:10.1371/journal.pone.0136403.g007

data. Our double count model indicates that perfect detection at the apex may be reasonable for some black bear sightings but not for all sightings. This is consistent with our expectations that black bears are not perfectly detectable at some apex distance from fast-moving aircraft over various covariate conditions. A MCDS estimate that assumes perfect detection at the apex would generate a population estimate of 2,220.6 bears while a MRDS model assuming full independence would estimate 1,716.6 bears. These estimates are 7.4% and 27.8% lower than our reported estimate of 2,377.0 black bears.

Very few options exist for modeling distance sampling data that exhibit a unimodal form. The gamma detection model developed by Becker and Quang [20] allows for covariates and has proven to be a useful model with aerielly collected distance sampling data if the assumption of perfect detection at the apex is reasonable. To determine if that assumption is reasonable, double-observer data must be collected [14]. If the analysis of the double-count data indicates a MRDS model is required, the use of the gamma detection function to model the distance data could be problematic. Using a gamma detection function in a MRDS model with PI would require a single apex location for the distance model, so no covariates could be used; thus a

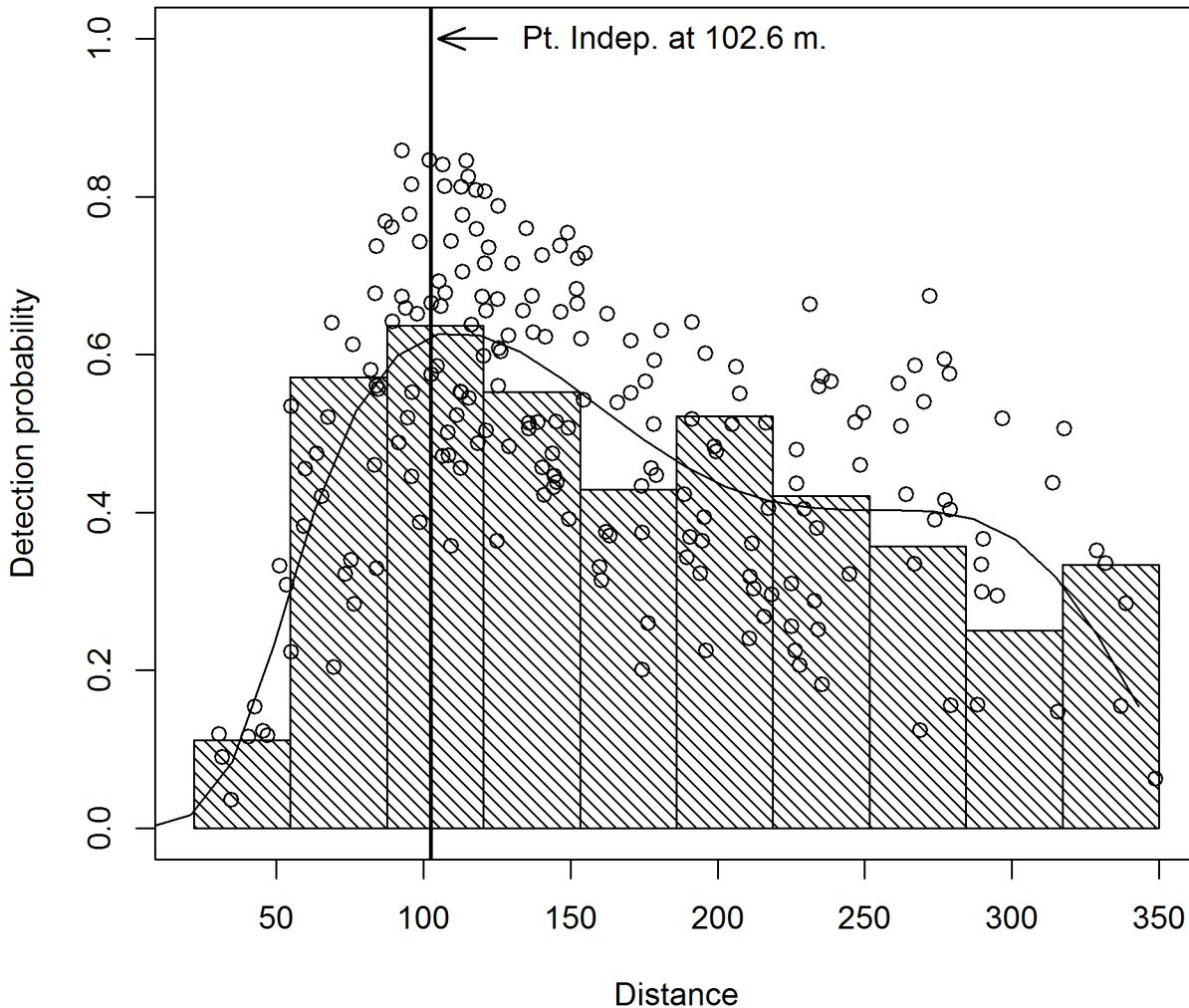


Fig 8. Logistic model of detection probabilities of the backseat observer given the pilot detected a black bear. The solid line denotes the conditional detection probability for the average values of the distance, percent cover, observer, air speed, and bedded/sitting bear covariates. The circles represent the model estimated conditional probabilities for each bear, based on their observed covariates.

doi:10.1371/journal.pone.0136403.g008

CDS model using a gamma detection function would model the distance data. When appreciable variation in animal detection exists, as in the black bear example, the inability of the CDS model to incorporate covariates would result in an inferior estimate (AIC = 3111.5 versus 3064.3 for our model) [5]. Neilson et al. [30] used a non-monotonic, non-parametric Gaussian kernel estimator to model aerially collected distance data on golden eagles (*Aquila chrysaetos*) with a MRDS model that used the PI assumption. Their nonparametric detection model is more flexible than a unimodal detection model, but does not allow for covariates in the distance portion of the MRDS model.

The purpose of this paper is to model unimodal distance datasets with a MRDS model that addresses unmodeled heterogeneity in the mark-resight data and allows for the use of covariates in modeling animal detection. There are several approaches that could be modified to provide a solution. The limited independence model [17] is one possible solution; the logistic detection function may have to be modified to adequately model the unimodal aerial distance data. The probit-Bernoulli model [18] is another possible solution; it would need to be modified to model observer dependence as a function of distance from the detection apex. This

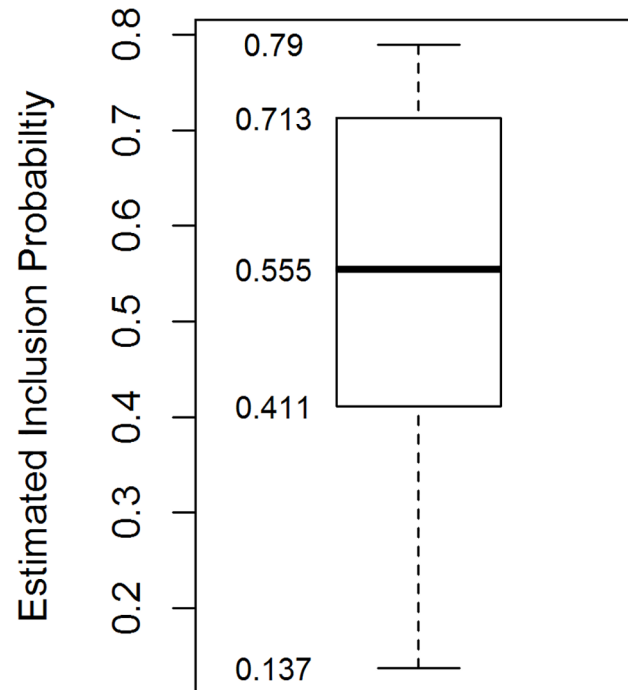


Fig 9. Boxplot of the estimated Horvitz-Thompson inclusion probabilities.

doi:10.1371/journal.pone.0136403.g009

model may allow for multiple detection apices. We choose to take the approach of Borchers et al. [12] to solve this problem. We used the two-piece normal detection function with covariates in a MRDS model with the PI assumption obtain a population estimate from our aerially collected bear data. This approach provided a good fit to our unimodal detection data, avoids the unrealistic assumption of perfect detection at the apex, avoids the FI assumption that has caused biases in many MRDS datasets, and allows for the use of covariates in the MCDS model to obtain a potentially superior model [5]. Researchers with unimodal detection data should consider the two-piece normal detection model for MRDS models and MCDS models.

Supporting Information

S1 File. BlackBear.MRDS.data.csv. A csv file containing the black bear distance data that was analyzed in the manuscript; this file is required by PlosOneMRDSAnalysis.txt.
(CSV)

S2 File. SkwetnaTranData350.csv. A csv file containing the transect area information; this file is required by SkwetnaBlackBearAnalysis.txt.
(CSV)

S1 Text. PlosOneMRDSAnalysis.txt. A R script file that will replicate the results found in this manuscript.
(TXT)

S2 Text. Two.Piece.Normal.txt. A R file containing all the R-functions written to fit a MRDS model that uses a two-piece normal detection model. These functions are required by the SkwetnaBlackBearAnalysis.txt file.
(TXT)

Acknowledgments

Rebecca Strauch, Alaska Department of Fish and Game, provided invaluable GIS support for this project. David Miller, University of St. Andrews, made the *mrds* R-code [24–25] available for us to obtain variance estimates. Steve Buckland, University of St. Andrews, provided invaluable encouragement and insights on modeling dependence in MRDS models. Jeff Laake, National Marine Mammal Laboratory, patiently answered numerous questions, and provided numerous insights into MRDS models. Louise Burt and David Miller, both at the University of St. Andrews, provided invaluable assistance in using and understanding the *mrds* R-package. We thank the pilots: J. DeCreft, D. Filkill, A. Greenblat, M. Litzen, and H. McMahan, for their expert and safe survey flying. We also thank the backseat observers: L. Gustavsen, T. Peltier, and H. Reynolds for their untiring efforts in collecting the data. This work was supported by the Alaska Department of Fish and Game and the Federal Aid in Wildlife Restoration. We thank Paul Conn and another anonymous referee for their constructive reviews that has greatly improved this manuscript.

The R-code [25] used in this analysis is included in the appendix. The two-piece normal detection model is planned to be incorporated into the *mrds* R-package [26–27] in the future.

Author Contributions

Conceived and designed the experiments: EFB. Performed the experiments: EFB AMC. Analyzed the data: EFB. Contributed reagents/materials/analysis tools: EFB. Wrote the paper: EFB AMC.

References

1. Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L. Introduction to distance sampling. New York: Oxford University Press; 2001.
2. Drummer TD, McDonald LL. Size bias in line-transect sampling. *Biometrics*. 1987 Mar; 43(1):13–21.
3. Quang PX, Becker EF. Line transect sampling under varying conditions with applications to aerial surveys. *Ecology*. 1996 Jun; 77(4):1297–1302.
4. Marques FFC, Buckland ST. Incorporating covariates into standard line transect analyses. *Biometrics*. 2003 Dec; 59(4):924–935. PMID: [14969471](#)
5. Marques FCC, Buckland ST. Covariate models for the detection function. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L, editors. *Advanced Distance Sampling*. Oxford: Oxford University Press; 2004. pp. 31–47.
6. Burnham KP, Buckland ST, Laake JL, Borchers DL, Marques TA, Bishop JRN, et al. Further topics in distance sampling. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L, editors. *Advanced Distance Sampling*. Oxford: Oxford University Press; 2004. pp. 307–392.
7. Burnham KP, Anderson DR, Laake JL. Estimation of density from line transect sampling of biological populations. *Wildlife Monographs*. 1980; 72:1–202.
8. Alpizar-Jara R, Pollock KH. A combination of line transect and capture-recapture sampling model for multiple observers in aerial surveys. *Environ Ecol Stat*. 1998 Dec; 3(4):311–327.
9. Manly BFJ, McDonald LL, Garner GW. Maximum likelihood estimation for the double-count method with independent observers. *J Agric Biol Environ Stat*. 1996 Jun; 1(2):170–189.
10. Quang PX, Becker EF. Combining line transect and double count sampling techniques for aerial surveys. *J Agric Biol Environ Stat*. 1997 Jun; 2(2):230–242.
11. Borchers DL, Zucchini W, Fewster RM. Mark-recapture models for line transect surveys. *Biometrics*. 1998 Dec; 54(4):1207–1220.
12. Borchers DL, Laake JL, Southwell C, Paxton CGM. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics*. 2006 Jun; 62(2):372–378. PMID: [16918901](#)
13. Huggins RM. On the statistical analysis of capture experiments. *Biometrika*. 1989 Mar; 76(1):133–140.
14. Laake JL, Borchers DL. Methods for incomplete detection at distance zero. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L, editors. *Advanced Distance Sampling*. Oxford: Oxford University Press; 2004. pp. 108–189.

15. Laake J, Dawson MJ, Hone J, Visibility bias in aerial survey: mark-recapture, line-transect or both? *Wildlife Research*. 2008 Jan; 35(4):299–309.
16. Laake JL. Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. In: Garner GW, Amstrup SC, Laake JL, Manly BFJ, McDonald LL, Robertson DG, editors. *Marine mammal survey and assessment methods*. Rotterdam: Balkema Press; 1999. pp. 137–148.
17. Buckland ST, Laake JL, Borchers DL. Double-observer line transect methods: levels of independence. *Biometrics*. 2010 Mar; 66(1):169–177. doi: [10.1111/j.1541-0420.2009.01239.x](https://doi.org/10.1111/j.1541-0420.2009.01239.x) PMID: [19432793](https://pubmed.ncbi.nlm.nih.gov/19432793/)
18. Conn PB, Laake JL, Johnson DS. A hierarchical modeling framework for multiple observer transect surveys. *PLoS One*. 2012 Aug 8; 7(8):e42294. doi: [10.1371/journal.pone.0042294](https://doi.org/10.1371/journal.pone.0042294) PMID: [22905121](https://pubmed.ncbi.nlm.nih.gov/22905121/)
19. Quang PX, Becker EF. Aerial line transect sampling of contour transects using double-count and covariate data. In: Garner GW, Amstrup SC, Laake JL, Manly BFJ, McDonald LL, Robertson DG, editors. *Marine mammal survey and assessment methods*. Rotterdam: Balkema Press; 1999. pp. 87–97.
20. Becker EF, Quang PX. A gamma-shaped detection function for line-transect surveys with mark-recapture and covariate data. *J Agric Biol Environ Stat*. 2009 Jun; 14(2):207–223.
21. Hiby L, Krishna MB. Line transect sampling from a curved path. *Biometrics*. 2001 Sep; 57(3):727–731. PMID: [11550921](https://pubmed.ncbi.nlm.nih.gov/11550921/)
22. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions, Volume 1*. New York: John Wiley and Sons; 1994.
23. Ramsey FL, Wildman V, Engbring J. Covariate adjustments to effective area in variable-area wildlife surveys. *Biometrics*. 1987 Mar; 43(1):1–11.
24. Venables WN, Ripley BD. *Modern applied statistics with S-PLUS*. New York: Springer-Verlag; 1997.
25. Hastie TJ. Generalized additive models. In: Chambers JM, Hastie TJ, editors. *Statistical Models*. New York: Chapman and Hall; 1997. pp. 249–307.
26. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing; 2014.
27. Laake JL, Borchers DL, Thomas LJ, Miller DL, Bishop J. Mark-recapture distance sampling (mrds); 2015 Jul. Software: mrds [Internet]. Available: <https://cran.r-project.org/web/packages/mrds/index.html>.
28. Akaike H. Prediction and entropy. In: Atkinson AC, Finberg SE, editors. *A celebration of statistics*. Berlin: Springer-Verlag; 1985. pp. 1–24.
29. Thompson SK. *Sampling*, 2nd ed. New York: John Wiley and sons; 2002.
30. Nielson RM, Mcmanus L, Rintz T, McDonald LL, Murphy RK, Howe WH, et al. Monitoring abundance of Golden Eagles in the Western United States. *J Wildl Manage*. 2014 May; 78(4):721–730.