Received: 8 September 2019 | Revised: 6 August 2020 | Accepted: 9 August 2020

DOI: 10.1002/env.2654

RESEARCH ARTICLE

WILEY

Adjusting a finite population block kriging estimator for imperfect detection

Matt Higham[1] | **Jay Ver Hoef**[2] | **Lisa Madsen**[3] | **Andy Aderman**[4]

[1]Department of Statistics, St. Lawrence University, Canton, New York, USA

[2]Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, Washington, USA

[3]Department of Statistics, Oregon State University, Corvallis, Oregon,

[4]Togiak National Wildlife Refuge, U.S. Fish and Wildlife Service, Dillingham, Alaska, USA

Correspondence
Matt Higham, Department of Statistics, St. Lawrence University, Canton, NY, USA.
Email: mhigham@stlawu.edu

Abstract

A finite population version of block kriging (FPBK) estimates a total or a mean when there is perfect detection of population units. However, many environmental datasets challenge the assumption of perfect detection. We consider two extensions of FPBK that incorporate imperfect detection. Spatial population estimator with detection: ratio then add (SPEDRA) adjusts observed counts by the estimated detection probability prior to spatial modeling. Spatial population estimator with detection: add then ratio (SPEDAR) uses spatial modeling on observed counts and then adjusts by mean detection probability. Unlike classical sampling approaches such as simple random sampling, SPEDRA and SPEDAR allow for spatial correlation among counts, and, being moment-based, are less computationally intensive than a fully Bayesian model. Both SPEDRA and SPEDAR perform similarly in some simulation settings and give comparable estimates for a moose population total when applied to data from Togiak National Wildlife Refuge (AK). In settings where detection probability varies widely across sites, however, SPEDRA outperforms SPEDAR in reducing root mean square prediction error. We recommend SPEDRA in surveys with imperfect detection because it is more theoretically sound and generally performs better.

KEYWORDS

geostatistics, sightability, spatial statistics, stratified random sampling

1 | INTRODUCTION

Abundance surveys for flora and fauna populations are often used to estimate the total count of a particular species for a variety of purposes, including management and ecological research of population dynamics. However, in a quantitative review of the ecological literature, Kellner and Swihart (2014) find that only 23% of ecological articles involved in estimating species abundance incorporate imperfect detection. Finite population block kriging (FPBK) is a geostatistical approach to predict the total abundance in a particular region from counts that may be spatially autocorrelated; however, it assumes perfect detection (Ver Hoef, 2001; Ver Hoef, 2008). Our overall goal is to extend FPBK to allow for imperfect detection.

Failing to incorporate imperfect detection gives biased results in fields ranging from freshwater biology (Gwinn, Beesley, Close, Gawne, & Davies, 2016) to species distribution analysis (Lahoz-Monfort, Guillera-Arroita, & Wintle, 2014) to estimation of the abundance of rare species (MacKenzie, Nichols, Sutton, Kawanishi, & Bailey, 2005). Gu and

Swihart (2004) argue for the importance of establishing relationships between detection probability and habitat covariates, as even detection that is almost perfect can heavily bias regression coefficients in modeling relationships between covariates and the presence of wildlife. More specifically to the topic of abundance estimation, Kéry and Schmidt (2008) discuss how patterns in the total count of an animal can be confounded with patterns in the detection probability.

A variety of methods have been proposed to estimate population abundance when some animals go unobserved. A basic method of incorporating imperfect detection is the Lincoln–Petersen estimator in mark-recapture studies (Lincoln, 1930; Petersen, 1896). Since their inception, mark-recapture methods have increased enormously in use and model complexity (Gould & Pollock, 2002; McCrea & Morgan, 2014; Otis, Burnham, White, & Anderson, 1978). The basic mark-recapture can be extended to a hierarchical spatial mark-recapture method for populations with high spatial correlation (Royle & Young, 2008). Distance sampling is another class of methods that incorporates imperfect detection by assuming that detection probability is a function of distance from a point or a transect line. Like mark-recapture, many studies have used distance sampling. For example, Peters et al. (2014) used distance sampling for moose surveys in Canada. Buckland, Anderson, Burnham, and Laake (2001) and Buckland, Anderson, Burnham, and Laake (2004) provide a more general introduction to distance sampling. Another method used in large-scale abundance estimation surveys is double sampling (Wilm, Costello, & Klipple, 1944), in which the surveyors sample a subset of the area of interest intensively to provide an adjustment for detection (Eberhardt & Simmons, 1987; Pollock et al., 2002). In estimating total moose abundance in Alaska, the double sampling method is used with stratified sampling on plot-based aerial counts (Gasaway, DuBois, Reed, & Harbo, 1986).

One major flaw in these approaches is that, though traditional mark-recapture, distance sampling, and double sampling adjust for perception bias (missed animals caused by observer error), they do not easily account for availability bias (missed animals caused by animals not being "available" to be observed). With some studies, a separate "sightability" survey is carried out to address availability bias. Madsen, Dalthorp, Huso, and Aderman (2020) use separate sightability data collected through radiocollars and parametric bootstrapping to calculate the uncertainty in detection estimates, but assume that the counts are spatially uncorrelated. Boveng et al. (2003) adjust for availability bias in estimating harbor seal counts with radiotelemetry data while Ver Hoef, Cameron, Boveng, London, and Moreland (2014) expand on the model by allowing for spatial autocorrelation among the counts. Other methods only require a single survey, but assume that covariates associated with abundance and detection are readily available (Sólymos, Lele, & Bayne, 2012).

1.1 | FPBK background

FPBK (Ver Hoef, 2001) was originally developed for moose surveys in Alaska. The Alaska Department of Fish & Game (ADF&G) denotes the FPBK estimator as the geospatial population estimator throughout its literature, which includes an operations manual (Kellie & DeLong, 2006) and a software user's guide (DeLong, 2006). The method has been widely used to estimate moose population totals throughout Alaska and western Canada, as 524 moose surveys have been analyzed using FPBK from 1997 to 2015 covering 303,144 square miles. The ADF&G maintains a database of moose surveys, which, as of 2015, comprises 53,153 records. Estimating moose populations is an important goal for wildlife management throughout many parts of Alaska and Canada. In particular, Boertje, Keech, Young, Kellie, and Seaton (2009) describe the significance of using sustainable yields to regulate the harvesting of moose in the Alaskan interior, and sustainable yields depend on accurate estimates of abundance.

FPBK differs from the usual block kriging method (Cressie, 1993, pp. 106–107) in that the number of sampling units is finite. The differences between the two methods are analogous to using the finite population correction factor in sampling theory (Ver Hoef, 2002). If we only have N distinct sampling units in our area of interest, the standard block kriging estimator will have an inflated variance, particularly if the ratio of sampled units to the total number of units is large.

Classical sampling can also be used to predict a population total with a finite number of sample units under the assumption that a random sample of sites was chosen, which was originally used for moose surveys in Alaska (Gasaway et al., 1986). An advantage of classical sampling compared with FPBK is that classical sampling requires few assumptions about the data because inference comes from the sampling design, which we often have complete control over (Ver Hoef, 2002). Variations of design-based sampling in the spatial setting are investigated in Stevens Jr and Olsen (2003), Fattorini, Corona, Chirici, and Pagliarella (2015), and Vagheggini, Bruno, and Cocchi (2016), all of which assume perfect detection of units.

However, if we make the assumption that the data were generated under a spatial stochastic process, then a model-based approach like FPBK often results in an estimator with lower prediction variance. Chan-Golston, Banerjee, and Handcock (2020) discuss a Bayesian approach to model-based inference for finite populations of a continuous response variable in a spatial setting. For more discussion of design-based versus model-based inference, see Sarndal et al. (1978). Because inference for model-based approaches is based on assumptions about the stochastic process, not the sampling design, FPBK allows for the possibility of nonrandom sampling to lower prediction variance even more. This is particularly useful if management is interested in small area estimation. For classical sampling, there is no guarantee that adequate sampling will be done in the small area, but, with the possibility of nonrandom sampling, managers have more control over the survey design (Kellie & DeLong, 2006; Ver Hoef, 2002).

1.2 | Imperfect detection modeling

There are several sensible estimators when imperfect detection occurs. Observed counts can be divided by estimated detection probabilities site-wise followed by FPBK. This adjusted estimator (which we will call SPEDRA, spatial population estimator with detection: ratio then add) is Horvitz–Thompson like in that it uses detection probabilities in place of inclusion probabilities (Horvitz & Thompson, 1952). A second possibility is to first use FPBK and then divide the estimated total by the mean detection probability (SPEDAR, spatial population estimator with detection: add then ratio). The SPEDAR method is similar to estimators by Manly, McDonald, and Garner (1996) and Borchers, Buckland, Goedhart, Clarke, and Hedley (1998), which both divide averaged counts by a mean detection probability.

For illustration of this problem in a simpler setting, suppose that we have only six plots that are completely independent and that we would like to estimate the total count in the six plots based on a survey of four plots in which the observed counts are $w_i = 3$, 5, 2, and 0. From a separate survey, suppose we estimate that the detection probabilities for these four plots are $\hat{\pi}_i = 0.2$, 0.9, 0.7, and 0.5, respectively. Then, for the first estimator, we would predict the total count $\hat{T}_1$ to be

$$\hat{T}_1 = \sum_{i=1}^{n} \frac{w_i}{\hat{\pi}_i} \cdot \frac{N}{n} = \left(\frac{3}{0.2} + \frac{5}{0.9} + \frac{2}{0.7} + \frac{0}{0.5} \right) \cdot \frac{6}{4} = 35.$$

For the second estimator, we would predict the total count $\hat{T}_2$ to be

$$\hat{T}_2 = \frac{\sum w_i}{\sum \hat{\pi}_i} \cdot N = \left(\frac{3+5+2+0}{0.2+0.9+0.7+0.5} \right) \cdot 6 = 26,$$

where w_i is the observed count on site i, $\hat{\pi}_i$ is the estimated detection probability on site i, N is the total number of sites, and n is the number of sampled sites. The estimators are identical when $\pi_1 = \pi_2 = \ldots = \pi_n$.

In this very simple example, the estimators are quite different. If we know detection exactly, then SPEDRA is unbiased. However, similar to Horvitz–Thompson estimators (Horvitz & Thompson, 1952), the variance of the estimator has the potential to become very large, particularly for small detection probabilities. Therefore, unless the detection probabilities are all equal, we anticipate that SPEDRA will be less biased than SPEDAR, but could have a much higher variance. On the other hand, SPEDAR is obviously biased, but may be more robust than SPEDRA.

1.3 | Goals and organization

Our goals are to (i) develop a model-based (geostatistical) approach to predict the total abundance in a region of interest using a version of FPBK that adjusts for imperfect detection, (ii) validate and compare the SPEDRA and SPEDAR models through a simulation study, and (iii) apply the methods to a moose dataset in Togiak National Wildlife Refuge. The adjusted FPBK estimator is useful for population abundance prediction problems where there is spatial autocorrelation in the counts and detection of the animals or plants in a sampling unit is not perfect.

In Section 2, we review FPBK assuming perfect detection and then develop the SPEDRA and SPEDAR models for incorporating imperfect detection. Next, in Section 3, we present results from a simulation study before applying the two

estimators to real data from a March 2017 Togiak moose survey in Section 4. We conclude in Section 5 with some remarks comparing the two estimators as well as some discussion on possible extensions to the models developed here.

2 | ADJUSTING THE FPBK MODEL FOR IMPERFECT DETECTION

We use the following three results frequently in the adjusted FPBK estimators. In particular, the expression for the variance of a product of two random variables (1) will be quite useful for SPEDRA because we will model an observed count at a particular site as the product of the true count and the estimated detection probability at that site.

2.1 | Basic equations

Let $\mathbf{x}$ and $\mathbf{y}$ be random column vectors with means $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ and variances $\mathbf{V}_x$ and $\mathbf{V}_y$, respectively. In addition, let $\mathbf{x}$ be independent of $\mathbf{y}$. Then, an extension of the variance of the product of two random variables (Goodman, 1960) is the following multivariate version (Ver Hoef et al., 2014):

$$\text{var}(\mathbf{x} \odot \mathbf{y}) = (\boldsymbol{\mu}_x \boldsymbol{\mu}_x') \odot \mathbf{V}_y + (\boldsymbol{\mu}_y \boldsymbol{\mu}_y') \odot \mathbf{V}_x + \mathbf{V}_x \odot \mathbf{V}_y, \tag{1}$$

where $\odot$ denotes element-wise, or Hadamard, product.

We will also make use of the conditional variance and conditional covariance laws in developing the model,

$$\text{var}(Y) = \text{E}[\text{var}(Y|X)] + \text{var}[\text{E}(Y|X)], \tag{2}$$

$$\text{cov}(Y_1, Y_2) = \text{E}[\text{cov}(Y_1, Y_2|X_1, X_2)] + \text{cov}[\text{E}(Y_1|X_1, X_2), \text{E}(Y_2|X_1, X_2)]. \tag{3}$$

2.2 | Background of FPBK

The following is a brief summary of Ver Hoef (2001) and Ver Hoef (2008), which proposes the FPBK model assuming perfect detection. If D is a spatial lattice indexed on a finite set of points $i = 1, 2, \ldots, N$, then let count $Z(\mathbf{s}_i)$ be a random variable at the ith site where $\mathbf{s}_i$ is a vector of the spatial coordinates of the ith site. Our goal is to predict $\mathbf{b}'\tilde{\mathbf{z}}$, where $\tilde{\mathbf{z}}$ is a column vector of the realized values of $Z(\mathbf{s}_i)$ for $i = 1, \ldots, N$ and $\mathbf{b}$ can be a vector where each element is 1 if we want to predict the population total, a vector where each element is $1/N$ if we want to predict the population mean, or a mix of 1's and 0's if we want to predict the total for a subset of D. Then, we want to find $\hat{\tau}(\mathbf{b}'\tilde{\mathbf{z}}) = \mathbf{a}'\tilde{\mathbf{z}}_s$, a linear combination of the observed data in order to predict $\mathbf{b}'\tilde{\mathbf{z}}$, where $\tilde{\mathbf{z}}_s$ is a vector of the observed data for the sampled locations in D and $\mathbf{a}'$ is a vector of weights.

Let $Y(\mathbf{s}_i) = \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} + \epsilon(\mathbf{s}_i)$ be a spatial random field. The error term $\epsilon(\mathbf{s}_i)$ is a spatial error process with 0 mean and a spatially autocorrelated covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ depends on just a few parameters $\boldsymbol{\theta}$. In addition, $\mathbf{x}'(\mathbf{s}_i)$ is a vector of covariates at location $\mathbf{s}_i$, and $\boldsymbol{\beta}$ is a parameter vector. For the remainder of this article, we use the exponential covariance model for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. That is, the i,jth entry for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is

$$\text{cov}(\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_j)|\boldsymbol{\theta}) = \theta_1 \exp(-h_{i,j}/\theta_2) = \boldsymbol{\Sigma}_{i,j}, \tag{4}$$

where $h_{i,j}$ is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. Note, however, that any spatial covariance matrix could be used (Chiles & Delfiner, 1999, pp. 80–93).

We then allow $Z(\mathbf{s}_i)$ to have the following conditional moments:

$$\text{E}[Z(\mathbf{s}_i)|Y(\mathbf{s}_i)] = Y(\mathbf{s}_i),$$

$$\text{var}[Z(\mathbf{s}_i)|Y(\mathbf{s}_i)] = \theta_3.$$

Here, $Z(\mathbf{s}_i)|Y(\mathbf{s}_i)$ is independent of $Z(\mathbf{s}_j)|Y(\mathbf{s}_j)$ when $\mathbf{s}_i \neq \mathbf{s}_j$ and θ_3 is the nugget effect.

A spatial random field with exponential autocovariance function (4) is second-order stationary, meaning that the mean is constant and the covariance is a function only of the separation vector between any two locations, and it is isotropic, meaning it is a function of distance only (Cressie, 1993; Gelfand, Diggle, Fuentes, & Guttorp, 2010). However, the results that follow can be extended to other types of autocovariance models, including models that include one or two extra parameters to model anisotropy (Chiles & Delfiner, 1999, p. 94).

Note that we could also use an autoregressive model, such as a conditional autoregressive model (Besag, 1974) to model the spatial process. In general, such models are natural for finite sets of spatial data (Cressie, 1993, p. 8) and can have a large computational advantage because the inverse covariance matrix is specified (Ver Hoef, Peterson, Hooten, Hanks, & Fortin, 2018). However, when there are missing data, as with incomplete sampling, then obtaining the marginal distribution of the sampled data requires inverting the precision matrix, forming the covariance matrix for the data from the subset of rows and columns, and then inverting again for prediction (Ver Hoef et al., 2018). In addition, the autocorrelation and variances are nonstationary. Therefore, we focus on the geostatistical model.

Using the laws of conditional expectation, conditional variance (2), conditional covariance (3), and the independence of $Z(\mathbf{s}_i)\,|\,Y(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)\,|\,Y(\mathbf{s}_j)$, we obtain $\mathrm{E}[Z(\mathbf{s}_i)] = \boldsymbol{\mu}(\mathbf{s}_i)$, $\mathrm{var}[Z(\mathbf{s}_i)] = \theta_3 + \boldsymbol{\Sigma}_{i,i}$, and $\mathrm{cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] = \boldsymbol{\Sigma}_{i,j}$, where $\boldsymbol{\mu}(\mathbf{s}_i) = \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{i,j}$ is the i,jth element of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

For simpler notation, we can represent $\mathbf{z}$, the vector of the random variables $Z(\mathbf{s}_i)$, using the following linear model, with $\mathbf{z}_s$ denoting the vector of $\{Z(\mathbf{s}_i)\}$ for the sampled locations in D and $\mathbf{z}_u$ denoting the vector of $\{Z(\mathbf{s}_i)\}$ for the unsampled locations in D:

$$\begin{pmatrix} \mathbf{z}_s \\ \mathbf{z}_u \end{pmatrix} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_u \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\delta}_s \\ \boldsymbol{\delta}_u \end{pmatrix}, \tag{5}$$

where $\mathbf{X}_s$ and $\mathbf{X}_u$ are the design matrices for the sampled and unsampled sites, respectively, and $\boldsymbol{\delta}_s$ and $\boldsymbol{\delta}_u$ are zero-mean random errors for the sampled and unsampled sites. Denote $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ as the vector of the $\boldsymbol{\mu}(\mathbf{s}_i)$. If $\boldsymbol{\delta} = [\boldsymbol{\delta}_s \;\; \boldsymbol{\delta}_u]'$, then $E(\boldsymbol{\delta}) = \mathbf{0}$ and

$$\mathrm{var}(\boldsymbol{\delta}) \equiv \mathbf{D} = \mathrm{diag}(\theta_3) + \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{diag}(\theta_3) + \boldsymbol{\Sigma}_{ss} & \boldsymbol{\Sigma}_{su} \\ \boldsymbol{\Sigma}'_{su} & \mathrm{diag}(\theta_3) + \boldsymbol{\Sigma}_{uu} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{D}_{ss} & \mathbf{D}_{su} \\ \mathbf{D}'_{su} & \mathbf{D}_{uu} \end{pmatrix}, \tag{6}$$

where $\mathrm{diag}(\theta_3)$ is the diagonal matrix with diagonal elements θ_3, $\boldsymbol{\Sigma}_{ss}$ is the submatrix of $\mathrm{var}(\epsilon) \equiv \boldsymbol{\Sigma}(\boldsymbol{\theta})$ for the sampled locations, $\boldsymbol{\Sigma}_{su}$ is the covariance of the sampled sites with the unsampled sites, and $\boldsymbol{\Sigma}_{uu}$ is the submatrix of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ for the unsampled locations.

Then, the best linear unbiased predictor (BLUP) of $\mathbf{b}'\mathbf{z}$ is

$$\hat{\tau}(\mathbf{b}'\mathbf{z}) = \mathbf{b}'_s\mathbf{z}_s + \mathbf{b}'_u\hat{\mathbf{z}}_u, \tag{7}$$

where $\mathbf{b}_s$ and $\mathbf{b}_u$ are subvectors of $\mathbf{b}$ corresponding to the sampled and unsampled locations in D, respectively, and $\hat{\mathbf{z}}_u$ is the BLUP of $\mathbf{z}_u$,

$$\hat{\mathbf{z}}_u = \mathbf{D}'_{su}(\mathbf{D}_{ss})^{-1}(\mathbf{z}_s - \hat{\boldsymbol{\mu}}_s) + \hat{\boldsymbol{\mu}}_u, \tag{8}$$

with $\hat{\boldsymbol{\mu}}_s = \mathbf{X}_s\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, $\hat{\boldsymbol{\mu}}_u = \mathbf{X}_u\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$, and $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\mathbf{X}'_s\mathbf{D}_{ss}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}'_s(\mathbf{D}_{ss})^{-1}\mathbf{z}_s$, the generalized least squares estimator of $\boldsymbol{\beta}$.

The prediction variance of the FPBK estimator is

$$\mathbf{b}'\mathbf{D}\mathbf{b} - \mathbf{G}'(\mathbf{D}_{ss})^{-1}\mathbf{G} + \mathbf{H}'\mathbf{E}\mathbf{H}, \tag{9}$$

with

$$\begin{aligned} \mathbf{G} &= (\mathbf{D}_{ss})\mathbf{b}_s + \mathbf{D}_{su}\mathbf{b}_u, \\ \mathbf{H} &= \mathbf{X}'\mathbf{b} - \mathbf{X}'_s(\mathbf{D}_{ss})^{-1}\mathbf{G}, \\ \mathbf{E} &= (\mathbf{X}'_s(\mathbf{D}_{ss})^{-1}\mathbf{X}_s)^{-1}. \end{aligned}$$

The derivation of Equations (7)–(9) can be found in Ver Hoef (2008). Note that the above model assumes that all animals or plants are detected so that there is no difference in the number of observed animals or plants and the true

number of animals or plants at a particular site. In the following two models, we introduce the possibility that not all units at a particular site are observed.

2.3 | Spatial population estimator with detection: Ratio then add

The first proposed model adjusts the observed counts for each sample unit by their estimated detection probabilities prior to spatial modeling. We first motivate our model by considering $W(\mathbf{s}_i)|\{Z(\mathbf{s}_i), P(\mathbf{s}_i)\} \sim \text{Binomial}(Z(\mathbf{s}_i), P(\mathbf{s}_i))$, where $W(\mathbf{s}_i)$ is the observed count, $P(\mathbf{s}_i)$ is the estimated probability of detection of a single animal or plant, and $Z(\mathbf{s}_i)$ is the random variable for the true count at location $\mathbf{s}_i$, defined in (5), which assumes perfect detection. Site s_i may be either a sampled or unsampled site.

Now we suppose s_i denotes a sampled site. Let $\mathbf{p}_s$ denote the vector of $P(\mathbf{s}_i)$, the estimated probabilities of detection on the sampled sites, $\boldsymbol{\pi}_s$ denote the vector of expected values of $\mathbf{p}_s$, and $\mathbf{w}_s$ denote the vector of observed counts on the sampled sites. Let $\boldsymbol{\mu}_s$ denote $\mathbf{X}_s\boldsymbol{\beta}$. If we assume that $P(\mathbf{s}_i)$ is independent of $Z(\mathbf{s}_j)$ for all $\mathbf{s}_i$ and $\mathbf{s}_j$, then $\text{E}(\mathbf{w}_s) = \boldsymbol{\pi}_s \odot \boldsymbol{\mu}_s$ and, using Equations (1)–(3),

$$\begin{aligned}\text{var}(\mathbf{w}_s) \equiv \mathbf{C} = \text{diag}(\boldsymbol{\mu}_s \odot \boldsymbol{\pi}_s \odot (\mathbf{1} - \boldsymbol{\pi}_s)) + \boldsymbol{\pi}_s\boldsymbol{\pi}_s' \odot \mathbf{D}_{ss} \\ + \boldsymbol{\mu}_s\boldsymbol{\mu}_s' \odot \mathbf{V}_{ss} + \mathbf{D}_{ss} \odot \mathbf{V}_{ss},\end{aligned} \tag{10}$$

where $\mathbf{V}_{ss}$ is the covariance matrix of the $\mathbf{p}_s$ and $\mathbf{1}$ denotes a column vector of 1's. While we could model $\mathbf{p}_s$ with spatial random effects in much the same way as we did for abundance, it may often be the case that detection is an observation process that is under our control with well-known covariates that mitigate spatial autocorrelation. On the other hand, the abundance of natural resources, like animals or plants, are often inherently aggregated on the landscape with autocorrelated residual error. In addition, in our application, we have very few nondetections for modeling spatial autocorrelation. Hence, in our simulations, and application, we assume detection variables are spatially independent.

We can also write the covariance between the observed sample counts $\mathbf{w}_s$ and the true counts on all of the sites $\mathbf{z}$ as $\text{cov}(\mathbf{w}_s, \mathbf{z}') \equiv \mathbf{R} = \boldsymbol{\pi}_s \odot \begin{pmatrix}\mathbf{D}_{ss} & \mathbf{D}_{su}\end{pmatrix}$.

We want to find $\hat{\tau}(\mathbf{b}'\mathbf{z}) = \mathbf{a}'\mathbf{w}_s$ to predict $\mathbf{b}'\tilde{\mathbf{z}}$. The goal is to find weights $\mathbf{a}$, but now these weights are applied to the observed counts $\mathbf{w}_s$, not the true counts $\mathbf{z}_s$ because the $\mathbf{z}_s$ are unknown. Let $M_{\mathbf{a}} = \text{E}[(a'\mathbf{w}_s - b'\mathbf{z})^2]$ be the mean square prediction error (MSPE) for any particular $\mathbf{a}$. Our goal is to find the BLUP; similar to Ver Hoef (2008), we want the "best" weights λ such that

1 $\text{E}(\lambda'\mathbf{w}_s) = \text{E}(\mathbf{b}'\mathbf{z})$ and
2 $M_{\mathbf{a}} - M_{\lambda}$ is nonnegative for all $\mathbf{a} \neq \lambda$.

First, we can restrict ourselves to the class of all unbiased estimators such that $\text{E}(a'\mathbf{w}_s) = \text{E}(b'\mathbf{z})$ for all $\boldsymbol{\beta}$ in the marginal linear model for $\mathbf{z}$ in Equation (5). This implies that $\mathbf{a}'(\boldsymbol{\pi}_s \odot \mathbf{X}_s)\boldsymbol{\beta} = \mathbf{b}'\mathbf{X}\boldsymbol{\beta}$ for all $\boldsymbol{\beta}$, or, equivalently, $\mathbf{a}'\mathbf{X}_s^* = \mathbf{b}'\mathbf{X}$ with $\mathbf{X}_s^* = (\boldsymbol{\pi}_s \odot \mathbf{X}_s)$.

Now we need to find the λ that makes $M_{\mathbf{a}} - M_{\lambda}$ nonnegative such that the unbiasedness constraint holds. If we minimize the MSPE, then we obtain the prediction equations

$$\begin{pmatrix}\mathbf{C} & \mathbf{X}_s^* \\ \mathbf{X}'^*_s & \mathbf{0}\end{pmatrix}\begin{pmatrix}\lambda \\ \mathbf{L}\end{pmatrix} = \begin{pmatrix}\mathbf{Rb} \\ \mathbf{X}'\mathbf{b}\end{pmatrix},$$

where $\mathbf{L}$ is the LaGrange multiplier in the system of equations. Note the similarity in form between these prediction equations and the prediction equations in standard block kriging and also in FPBK assuming perfect detection in Ver Hoef (2008). Solving for λ,

$$\lambda' = \mathbf{b}'\mathbf{R}'\mathbf{C}^{-1} + (\mathbf{b}'\mathbf{X} - \mathbf{b}'\mathbf{R}'\mathbf{C}^{-1}\mathbf{X}_s^*)(\mathbf{X}_s^{*'}\mathbf{C}^{-1}\mathbf{X}_s^*)^{-1}\mathbf{X}_s^{*'}\mathbf{C}^{-1}, \tag{11}$$

we obtain our predictor as $\hat{\tau}(\mathbf{b}'\mathbf{z}) = \lambda'\mathbf{w}_s$, with a prediction variance of

$$\text{var}[\hat{\tau}(\mathbf{b}'\mathbf{z})] = \lambda'\mathbf{C}\lambda - \mathbf{b}'\mathbf{R}'\lambda - \lambda'\mathbf{Rb} + \mathbf{b}'\mathbf{Db}. \tag{12}$$

The above model could also be cast as a hierarchical model, $[\mathbf{z}_s, \mathbf{y}_s, \mathbf{p}_s, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \theta_3 | \mathbf{w}_s, \mathbf{K}, \mathbf{X}]$, where $\mathbf{K}$ is a vector of 1's and 0's resulting from the sightability trials. A natural approach would be to adopt a Bayesian model with prior distributions and MCMC sampling. However, such an approach can be time consuming and relies on a great amount of statistical expertise. Our goal is to have a fairly automatic way for a nonstatistician user to obtain results quickly and reliably. In several applications that use FPBK so far, dozens of surveys are completed annually, and each one cannot be a carefully tuned analysis. The present approach is fast and robust, relying only on moment assumptions. However, for the interested reader, we have also analyzed these data with a Bayesian hierarchical model (Higham, 2019).

2.4 | The logistic regression model

In order to use the above model, we must have some way to estimate the detection probabilities for the sites in the population of interest. In some animal surveys, there are radiocollared animals and separate sightability trials to model the imperfect detection of animals. Suppose that we have n radiocollared animals to be used for the sightability trials, indexed $i = 1, \ldots, n$. For each i, $K_i \sim \text{Bernoulli}(\pi_i)$, where K_i is the random variable for whether or not a radiocollared animal is detected. Also suppose that we have a design matrix $\mathbf{U}$ where each row contains covariates like habitat conditions or time spent surveying for predicting the detection probability in a particular site. Denote $\mathbf{U}_r$ as the design matrix for the sites with the radiocollared animals and $\boldsymbol{\pi}_r$ as the vector of probabilities that a radiocollared animal is sighted. Then, we assume that we have the regression model

$$\text{logit}(\boldsymbol{\pi}_r) = \mathbf{U}_r \boldsymbol{\gamma}, \tag{13}$$

where $\boldsymbol{\gamma}$ is the parameter vector and logit$(\cdot)$ takes the logit of each element of $(\cdot)$.

2.5 | Estimation

We next discuss estimation of the quantities in (11) and (12). If we denote $\mathbf{U}_s$ as the rows of $\mathbf{U}$ corresponding to the sampled sites in the animal count survey, then $\text{logit}(\hat{\boldsymbol{\pi}}_s) = \mathbf{U}_s \hat{\boldsymbol{\gamma}}$, where $\hat{\boldsymbol{\gamma}}$ is the estimator for $\boldsymbol{\gamma}$ in standard logistic regression. The estimated detection probabilities $\hat{\boldsymbol{\pi}}_s$ are then obtained through the element-wise inverse logit transformation on $\mathbf{U}_s \hat{\boldsymbol{\gamma}}$. The $\hat{\boldsymbol{\pi}}_s$ take the place of $\mathbf{p}_s$ in the binomial model for $\mathbf{w}_s$, though, in a more general setting, $\mathbf{p}_s$ could be obtained from a method other than logistic regression.

We investigated using either nonparametric bootstrapping (Efron (1992)) or the delta method (Dorfman (1938), Ver Hoef (2012)) to estimate $\mathbf{V}_{ss}$, the covariance matrix of $\hat{\boldsymbol{\pi}}_s$, ultimately finding that bootstrapping produced more accurate results in simulations. Letting $\mathbf{U}_{all}$ denote the matrix combining the design matrix $\mathbf{U}_r$ and the sightability response vector $(K_1, K_2, \ldots, K_n)'$, we perform 1,400 nonparametric bootstraps of the rows of $\mathbf{U}_{all}$, estimate $\hat{\boldsymbol{\gamma}}_{\text{boot}}$ for each bootstrap, calculate $\hat{\boldsymbol{\pi}}_{s,\text{boot}}$, and find the empirical covariance of the 1,400 $\hat{\boldsymbol{\pi}}_{s,\text{boot}}$ vectors to obtain an estimator of $\mathbf{V}_{ss}$.

We use maximum likelihood to obtain estimates for $\boldsymbol{\mu}_s$, denoted $\hat{\boldsymbol{\mu}}_s$, and the covariance parameters in $\mathbf{D}$, denoted $\hat{\mathbf{D}}$, assuming the exponential covariance structure defined in (4). We use a normal likelihood model for the $\mathbf{w}_s$ with the covariance matrix in Equation (10), plugging in $\hat{\boldsymbol{\pi}}_s$ and $\hat{\mathbf{V}}_{ss}$. Note that, for ease of estimation, we are using a misspecified normal likelihood for the correlated binomial counts. Our estimators for the mean and covariance parameters from misspecified maximum likelihood are consistent estimators for the $\boldsymbol{\mu}_s$ and $\boldsymbol{\theta}$ parameters that minimize the Kullback–Leibler distance between the misspecified distribution and the true distribution (Kullback & Leibler, 1951; White, 1982). Huber (1967) and White (1982) give additional examples and properties of maximum likelihood estimators under distribution misspecification. Then, we have $\hat{\mathbf{D}} = \text{diag}(\hat{\theta}_3) + \hat{\boldsymbol{\Sigma}}$, $\hat{\mathbf{X}}_s^* = (\hat{\boldsymbol{\pi}}_s \odot \mathbf{X}_s)$, and, partitioning $\hat{\mathbf{D}}$ as in Equation (6), $\hat{\mathbf{R}} = \hat{\boldsymbol{\pi}}_s \odot \begin{pmatrix} \hat{\mathbf{D}}_{ss} & \hat{\mathbf{D}}_{su} \end{pmatrix}$. The estimated variance of the observed counts is

$$\widehat{\text{var}}(\mathbf{w}_s) \equiv \hat{\mathbf{C}} = \text{diag}(\hat{\boldsymbol{\mu}}_s \odot \hat{\boldsymbol{\pi}}_s \odot (\mathbf{1} - \hat{\boldsymbol{\pi}}_s)) + \hat{\boldsymbol{\pi}}_s \hat{\boldsymbol{\pi}}_s' \odot \hat{\mathbf{D}}_{ss} + \hat{\boldsymbol{\mu}}_s \hat{\boldsymbol{\mu}}_s' \odot \hat{\mathbf{V}}_{ss} + \hat{\mathbf{D}}_{ss} \odot \hat{\mathbf{V}}_{ss}. \tag{14}$$

Finally, we obtain the following plug-in estimators for λ, $\tau(\mathbf{b}'\mathbf{z})$, and $\text{var}[\tau(\mathbf{b}'\mathbf{z})]$:

$$\hat{\lambda}' = \mathbf{b}'\hat{\mathbf{R}}'\hat{\mathbf{C}}^{-1} + (\mathbf{b}'\mathbf{X} - \mathbf{b}'\hat{\mathbf{R}}'\hat{\mathbf{C}}^{-1}\hat{\mathbf{X}}_s^*)(\hat{\mathbf{X}}_s^{*'}\hat{\mathbf{C}}^{-1}\hat{\mathbf{X}}_s^*)^{-1}\hat{\mathbf{X}}_s^{*'}\hat{\mathbf{C}}^{-1},$$

$$\hat{\tau}(\mathbf{b}'\mathbf{z}) = \hat{\lambda}'\mathbf{w}_s,$$

$$\widehat{\text{var}}[\hat{\tau}(\mathbf{b}'\mathbf{z})] = \hat{\lambda}'\hat{\mathbf{C}}\hat{\lambda} - \mathbf{b}'\hat{\mathbf{R}}'\hat{\lambda} - \hat{\lambda}'\hat{\mathbf{R}}b + \mathbf{b}'\hat{\mathbf{D}}b.$$

2.6 | Spatial population estimator with detection: Add then ratio

The second proposed model uses the FPBK estimator defined in (7) on the observed counts and then divides by the estimated mean detection probability. Here, we use the notation **obs** as a reminder that kriging uses the observed counts. Using this notation, we can use the standard FPBK in (Ver Hoef, 2008) on the observed counts to obtain

$$\hat{\tau}_{\mathbf{obs}}(\mathbf{b}'\mathbf{w}) = \mathbf{b}'_s\mathbf{w}_s + \mathbf{b}'_u\hat{\mathbf{w}}_u. \tag{15}$$

Denote this predictor for the observed counts as $\hat{\tau}_{\mathbf{obs}}$. Also define $\hat{\bar{\pi}} = \sum_{j=1}^{m} \frac{\hat{\pi}_j}{m}$, the estimated mean detection probability where m is the number of sites in the count survey and j indexes the detection probabilities for the sites in the count survey. Then, we adjust estimator (15) by the mean detection probability across the sampled sites to obtain the SPEDAR estimator

$$\hat{\tau}_2(\mathbf{b}'\mathbf{w}) = \hat{\tau}_{\mathbf{obs}}\hat{\bar{\pi}}^{-1}. \tag{16}$$

We obtain the prediction variance using a plug-in estimator in the scalar form of Equation (1),

$$\text{var}(\hat{\tau}_{\mathbf{obs}}\hat{\bar{\pi}}^{-1}) = (\text{E}(\hat{\tau}_{\mathbf{obs}}))^2\text{var}(\hat{\bar{\pi}}^{-1}) + (\text{E}(\hat{\bar{\pi}}^{-1}))^2\text{var}(\hat{\tau}_{\mathbf{obs}}) + \text{var}(\hat{\tau}_{\mathbf{obs}})\text{var}(\hat{\bar{\pi}}^{-1}). \tag{17}$$

We use bootstrapping to obtain an approximation for $\text{E}(\hat{\bar{\pi}}^{-1})$ and $\text{var}(\hat{\bar{\pi}}^{-1})$, resampling the rows of $\mathbf{U}_{all}$ as in the bootstrapping for SPEDRA estimator. $\hat{\text{E}}(\hat{\bar{\pi}}^{-1})$ and $\widehat{\text{var}}(\hat{\bar{\pi}}^{-1})$ are then the mean and variance of the 1,400 bootstrapped quantities of $\hat{\bar{\pi}}^{-1}$. Putting the bootstrap estimates back into Equation (17), as well as using the observed counts in Equation (7) as an estimate for $\text{E}(\hat{\tau}_{\mathbf{obs}})$ and using the observed counts in Equation (9) to estimate $\text{var}(\hat{\tau}_{\mathbf{obs}})$, gives an approximation to the prediction variance for the SPEDAR estimator.

It is unclear whether SPEDAR is better than SPEDRA in terms of unbiasedness and prediction variance, but results from distance sampling (Borchers et al., 1998; Manly et al., 1996) and weighted semivariograms (Zimmerman & Ver Hoef, 2017) indicate that there could be substantial reduction in variance at the expense of a little more bias. Therefore, we include both the SPEDAR and the SPEDRA models here in an effort to compare the two in a simulation study in the FPBK setting.

Note that we make slightly different assumptions about the nature of the spatial correlation in the counts in the SPEDRA and SPEDAR models. For SPEDRA, we assume that the *total* counts of animals per site are spatially autocorrelated according to some model while, for SPEDAR, we assume that the *observed* counts of animals per site are spatially autocorrelated according to some model. These assumptions are slightly different because, for SPEDRA, we incorporate detection before performing kriging while, for SPEDAR, we perform kriging on the observed counts. The implications of these assumptions for which model might be better to use are given in Section 5.

3 | SIMULATION STUDY

We ran a Latin Square simulation experiment to cover a wide range of possible detection probabilities. A 3×3 Latin Square with three sightability sample sizes $n_{\text{det}} \in \{60, 100, 200\}$, three possible levels of mean detection $\gamma \in \{0.25, 0.5, 0.75\}$, and three possible range parameters $\theta_2 \in \{0.1, 5, 30\}$ was replicated for each combination of two population means $\mu \in \{4, 8\}$ and two detection variances $\gamma_1 \in \{1, 4\}$ for low and high detection variances, respectively (Appendix A). $\gamma_{0\text{cat}}$ refers to the level of mean detection (Low has a mean detection of 0.25, Medium has a mean detection of 0.5, and High has a mean detection of 0.75). We simulated counts on a regular 1×1 unit grid. The total number of sites was 400, a 20×20 grid. For each simulation setting, we ran 1,400 simulations with the partial sill fixed at 2, the nugget fixed at 0.02, and the count survey sample size fixed at 100.

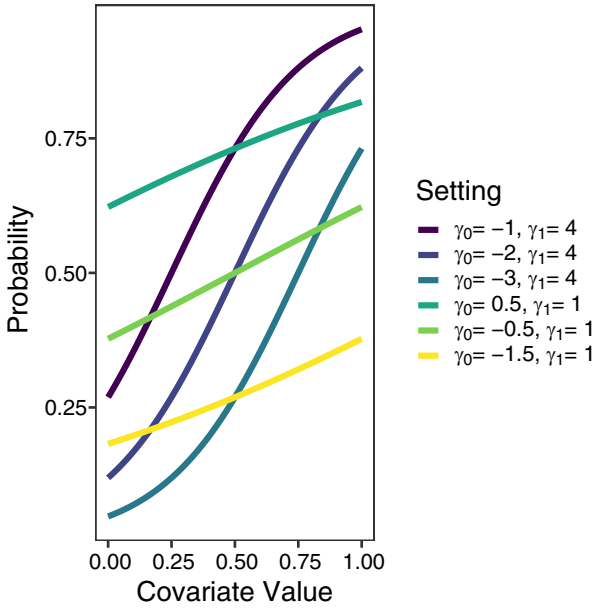

FIGURE 1 Logistic regression curves for the settings with high detection probability variance ($\gamma_1 = 4$) and low detection probability variability ($\gamma_1 = 1$) for three different mean detection levels

We simulated spatially correlated normal random variables with the specified mean and covariance parameters assuming the exponential spatial autocorrelation model (rounding to the nearest nonnegative integer) as the true counts. We then simulated the true detection probabilities for each site using the logistic regression parameters in Figure 1 with covariates simulated as Unif(0, 1) random variables. The observed counts were then simulated as binomial random variables with sizes equal to the true counts and probabilities equal to the true detection probabilities on a random sample of sites corresponding to the number of sites in the moose count survey. The sightability data were formed by generating n_{det} Bernoulli random variables using the true detection probabilities.

The full results of the simulations are given in Appendix A. The coverage across all settings for SPEDRA is 0.882 with a minimum of 0.841 while the coverage across all settings for SPEDAR is 0.911 with a minimum of 0.882. We also included a simple random sample (SRS) estimator in the simulations to compare with SPEDRA and SPEDAR. SRS was calculated as $\frac{N}{m}\sum_{i=1}^{m}\frac{w_i}{\hat{\pi}_i}$, where m denotes the number of sites in the count survey and N denotes the total number of sites in the study region. Results from three of the scenarios are shown in Figure 2. As expected, SRS performs similarly to the two spatial estimators when the range parameter is small, as seen in Setting C. On the other hand, in Setting A, where there is a large amount of spatial correlation, SRS performs much more poorly than the spatial estimators. In all settings, SRS had a comparable or worse root mean square prediction error (rMSPE) than the SPEDRA and SPEDAR estimators so we only consider SPEDRA and SPEDAR henceforth.

The results from the simulations are summarized as linear models, with the experimental units as the individual simulation settings, in Tables 1,2, and 3. Table 1 uses SPEDRA rMSPE as the response with the simulation factors as predictors. All of the factors considered are significant at a significance level of 0.05 except for the level of the range parameter, θ_2. Table 2 uses SPEDAR rMSPE as the response with the simulation factors as predictors. The results are similar to those in Table 1, but now the detection variance factor ($\gamma_1 = 4$) is only weakly significant, indicating that, for SPEDAR, the variability in detection is a less important influence on rMSPE than it is for SPEDRA rMSPE. The signs of the point estimates in Tables 1 and 2 are identical, showing that rMSPE tends to increase for: (i) larger means, (ii) larger ranges, (iii) lower mean detection, (iv) lower detection variability, and (v) lower detection sample size.

Table 3 uses the efficiency, rMSPE of SPEDRA divided by rMSPE of SPEDAR, as the response with the simulation factors as predictors. An efficiency greater than 1 indicates that SPEDAR is better than SPEDRA while an efficiency less than 1 indicates that SPEDRA is better than SPEDAR. SPEDRA does better than SPEDAR for higher levels of detection variability and higher levels of mean detection. The form of the detection function therefore drives the change in efficiency between SPEDRA and SPEDAR. As detection gets smaller, the efficiency increases due to the instabilities in dividing by small individual detections.

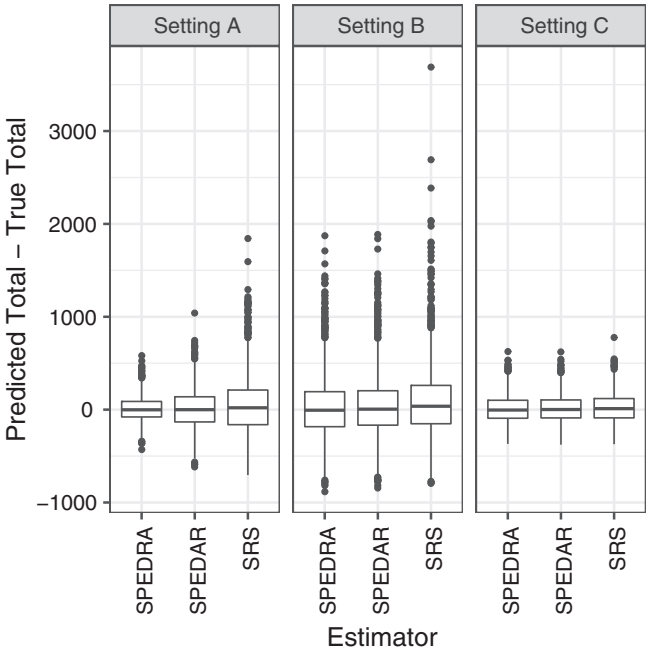

FIGURE 2 Comparison of SPEDRA, SPEDAR, and simple random sampling (SRS) for three simulations scenarios. Setting A corresponds to a setting where SPEDRA outperforms SPEDAR and SRS with $\mu = 8$, $\theta_2 = 5$, $n_{\text{det}} = 100$, $\gamma_{0\text{cat}}$ is High, and $\gamma_1 = 4$. Setting B corresponds to a setting where SPEDRA and SPEDAR outperform SRS with $\mu = 4$, $\theta_2 = 30$, $n_{\text{det}} = 100$, $\gamma_{0\text{cat}}$ is Low, and $\gamma_1 = 1$. Setting C corresponds to a setting where all three estimators perform similarly with $\mu = 4$, $\theta_2 = 0.1$, $n_{\text{det}} = 200$, $\gamma_{0\text{cat}}$ is Medium, and $\gamma_1 = 1$. SPEDAR, spatial population estimator with detection: add then ratio; SPEDRA, spatial population estimator with detection: ratio then add

TABLE 1 Linear model results with SPEDRA rMSPE as the response and the simulation settings as experimental units

Parameter	Estimate	Std. error	*p*-Value
Intercept	176.61	34.55	<.0001
$\mu = 8$	146.28	23.03	<.0001
$\theta_2 = 30$	11.58	28.21	.6846
$\theta_2 = 5$	7.83	28.21	.7834
$\gamma_{0\text{cat}} = \text{Low}$	261.83	28.21	<.0001
$\gamma_{0\text{cat}} = \text{Medium}$	89.33	28.21	.0038
$\gamma_1 = 4$	−90.72	23.03	.0005
$n_{\text{det}} = 100$	−72.50	28.21	.0160
$n_{\text{det}} = 200$	−138.33	28.21	<.0001

Note: $\gamma_{0\text{cat}}$ indicates the level of mean detection (Low = 0.25, Medium = 0.5, and High = 0.75). The reference group has the following simulation parameters: true mean count $\mu = 4$, range $\theta_2 = 0.1$, $\gamma_{0\text{cat}} = \text{High}$, logistic regression slope $\gamma_1 = 1$, and detection sample size $n_{\text{det}} = 60$.
Abbreviations: rMSPE, root mean square prediction error; SPEDRA, spatial population estimator with detection: ratio then add.

4 | APPLICATION TO TOGIAK MARCH 2017 SURVEY

4.1 | Study area description

The Togiak Wildlife Refuge, an area of about 21,000 km^2 of land in southwestern Alaska, has seen an increase in moose since the early 1980s, when wildlife biologists believed there to be fewer than 35 moose in the area (Benson et al., 2015). Togiak has historically been snow-covered beginning in November, the time of year traditionally used for moose surveys. On snow-covered ground, moose were easily sighted during surveys, resulting in a detection probability very close to 1 and assumed to be exactly 1 in subsequent statistical analyses. However, warming trends have resulted in inadequate snowfall for good moose detection in the past 10 years and future changes are not expected to improve the situation (Park, Yabuki, & Ohata, 2012).

TABLE 2 Linear model results with SPEDAR rMSPE as the response and the simulation settings as experimental units

Parameter	Estimate	Std. error	*p*-value
Intercept	176.92	33.13	<.0001
$\mu = 8$	169.94	22.08	<.0001
$\theta_2 = 30$	17.17	27.05	.5310
$\theta_2 = 5$	13.50	27.05	.6217
$\gamma_{0\text{cat}} = \text{Low}$	256.83	27.05	.0000
$\gamma_{0\text{cat}} = \text{Medium}$	89.58	27.05	.0026
$\gamma_1 = 4$	−42.50	22.08	.0649
$n_{\text{det}} = 100$	−82.50	27.05	.0051
$n_{\text{det}} = 200$	−162.08	27.05	<.0001

Note: $\gamma_{0\text{cat}}$ indicates the level of mean detection (Low = 0.25, Medium = 0.5, and High = 0.75). The reference group has the following simulation parameters: true mean count $\mu = 4$, range $\theta_2 = 0.1$, $\gamma_{0\text{cat}} = \text{High}$, logistic regression slope $\gamma_1 = 1$, and detection sample size $n_{\text{det}} = 60$.
Abbreviations: rMSPE, root mean square prediction error; SPEDAR, spatial population estimator with detection: add then ratio.

TABLE 3 Linear model results with efficiency (rMSPE of SPEDRA divided by rMSPE of SPEDAR) as the response and the simulation settings as experimental units

Parameter	Estimate	Std. error	*p*-value
Intercept	0.961	0.032	<.0001
$\mu = 8$	−0.030	0.021	.1721
$\theta_2 = 30$	−0.038	0.026	.1601
$\theta_2 = 5$	−0.027	0.026	.3097
$\gamma_{0\text{cat}} = \text{Low}$	0.106	0.026	.0004
$\gamma_{0\text{cat}} = \text{Medium}$	0.052	0.026	.0573
$\gamma_1 = 4$	−0.195	0.021	<.0001
$n_{\text{det}} = 100$	0.001	0.026	.9735
$n_{\text{det}} = 200$	0.035	0.026	.1996

Note: $\gamma_{0\text{cat}}$ indicates the level of mean detection (Low = 0.25, Medium = 0.5, and High = 0.75). The reference group has the following simulation parameters: true mean count $\mu = 4$, range $\theta_2 = 0.1$, $\gamma_{0\text{cat}} = \text{High}$, logistic regression slope $\gamma_1 = 1$, and detection sample size $n_{\text{det}} = 60$.
Abbreviations: rMSPE, root mean square prediction error; SPEDAR, spatial population estimator with detection: add then ratio; SPEDRA, spatial population estimator with detection: ratio then add.

Togiak is divided into sites that are approximately 17.6 km^2 (Figure 3). The sampling frame for moose surveys excludes sites with airports, communities, coastal, and mountainous sites with no moose habitats. Togiak wildlife biologists stratify sample units into a high expected moose count (H) stratum and a low expected moose count (L) stratum based on previous knowledge of the moose distribution. Within each stratum, most sites are randomly selected to be part of a particular survey. The remaining sites that are sampled are nonrandomly selected so that there are no large gaps or "holes" in the study region. Because the survey time typically only spans a few days, surveyors assume that there is no migration across sites during the time when moose are counted. Biologists then use FPBK on the observed counts from each stratum, and subsequently sum the stratum estimates to obtain a total abundance estimate. However, because of changing snow conditions and more difficult moose detectability, the number of observed moose is now thought to be substantially less than the true number of moose.

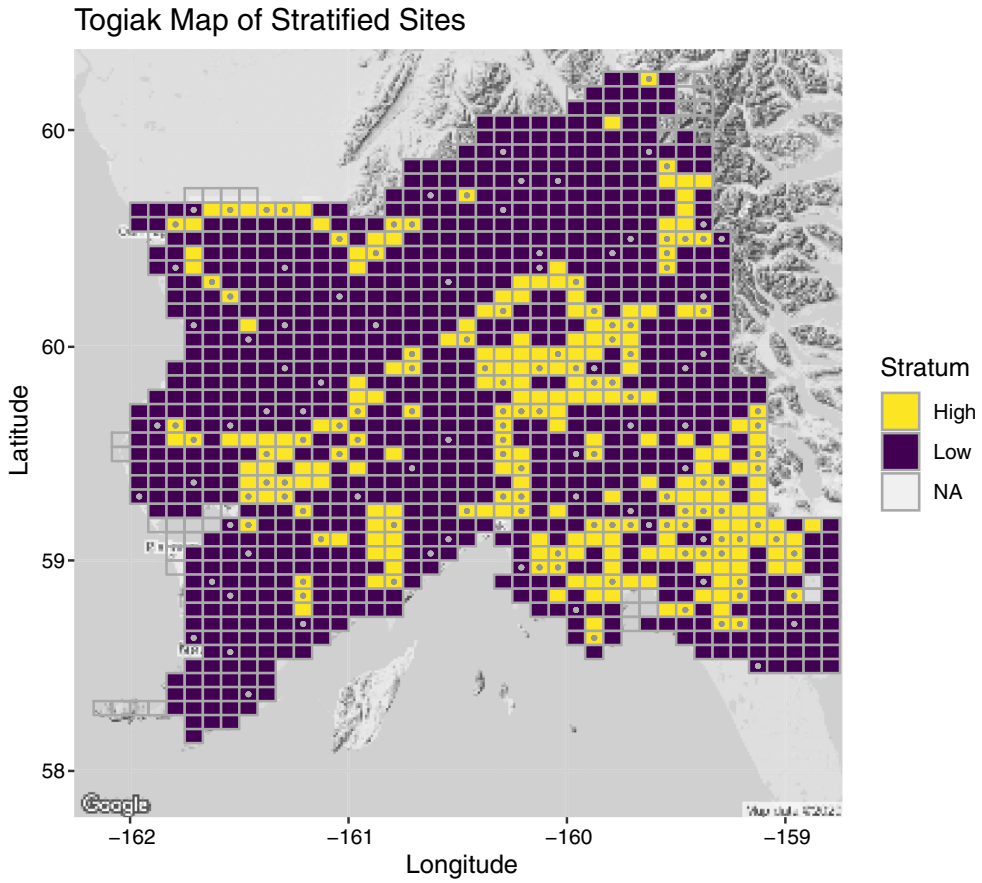

FIGURE 3 A map of the stratification and sampling of the sites in Togiak. A mark in the middle of a site indicates that the site was sampled for the March 2017 count survey. Sites marked as NA were excluded from the sampling frame

4.2 | Stratification

We applied both estimators to both the October 2016 Togiak moose survey data and the March 2017 Togiak moose survey data. In the Togiak surveys, stratification is used to increase the precision of the estimate of the total because wildlife biologists familiar with the study region typically have some knowledge about where to expect high counts and where to expect low counts of animals. In addition, in geostatistical approaches, we typically assume stationarity across the study area. Stratifying can help with the assumption of a constant mean, particularly in models without any covariates. Instead of assuming a constant mean across all sites, we now assume a constant mean within each stratum. When we incorporate stratification in the FPBK model with imperfect detection, we assume there is no crosscorrelation between strata. However, correlation between the estimators for the two strata still arises because the estimators for the totals in the two strata depend on the same sightability model. In other words, the $\hat{\boldsymbol{\pi}}_s$ vectors for each stratum will typically be positively correlated. In addition, the sites in Togiak are not all exactly equal in area because of each site's varying latitude and longitude. Therefore, we actually perform SPEDRA and SPEDAR on the densities, not the counts. The details of transforming from densities back to counts and of working stratification into the SPEDRA and SPEDAR estimators are given in Appendices B and C, respectively.

4.3 | Application results

We present the results for the March 2017 survey in detail here. Survey time and the proportions of water, birch, dwarf shrub, alder, and willow in a particular site served as possible covariates for the detection model. For the March sightability trials, 42 of the 50 moose were detected and, through an all-subsets selection procedure based on the Akaiki information criterion (Akaike, 1974), none of the possible covariates were informative in predicting the detection probability. There is little spatial autocorrelation among the moose counts in the low stratum, but there is a moderate amount of spatial autocorrelation in the high stratum (Figure 4). We also checked the assumption of stationarity in the high stratum by

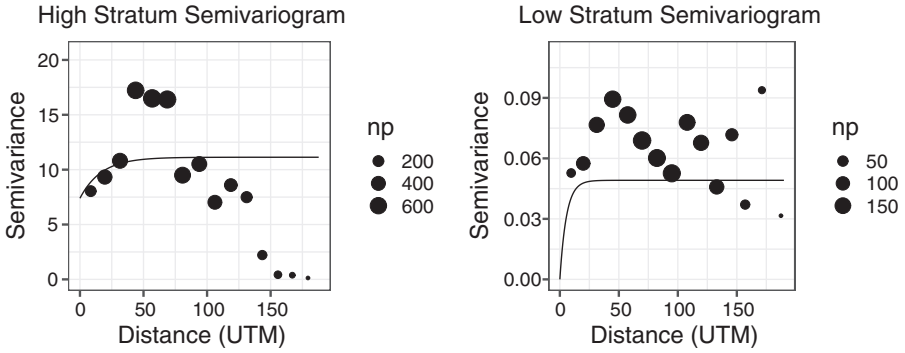

FIGURE 4 Semivariograms for each stratum for the March 2017 Togiak Wildlife Refuge survey using the estimated total moose in the surveyed sites for SPEDRA. The observed counts divided by the site-wise detection probabilities are used as the data for the empirical semivariogram while the fitted semivariogram model uses maximum likelihood to estimate the exponential covariance parameters. The sizes of the points in the semivariograms depend on the number of pairs used in that particular bin. The fitted lines are exponential semivariogram models that used maximum likelihood to estimate the nugget, partial sill, and range. Note the change in the semivariance scale between the high and low strata. SPEDRA, spatial population estimator with detection: ratio then add

comparing variograms for different quadrants of the study area. In this case, there was no strong violation of stationarity in the high stratum while the low stratum had too few data points to construct multiple semivariograms for different regions. Note that each semivariogram uses the naive $\hat{z}_i = w_i/\hat{\pi}_i$ as observations to compute the empirical variogram. Therefore, particularly in the low stratum where there is a large abundance of zeroes, the variogram appears to fit the "observations" poorly. The zeroes are not inflated at all in our naive "observations," which drives up the variance in the empirical variogram while our fitted variogram model using maximum likelihood estimation accounts for some inflation of the zero counts.

Figure 5 shows a few sites with very high predicted counts in the high stratum while most other sites are predicted to have few or no moose. The SPEDRA estimator yielded a total estimate of 3,658 moose while SPEDAR estimator yielded a total estimate of 3,645. Approximate 90% normal-based prediction intervals for the total number of moose at the Togiak Wildlife Refuge were (2,839, 4,478) moose using SPEDRA and (2,785, 4,505) moose using SPEDAR. In this case, both the estimates and the standard errors were approximately equal for the two estimators.

4.4 | Application simulation

In addition to the full simulation study in Section 3, we also ran 2,000 simulations for a stratified scenario with counts and detection similar to the March 2017 Togiak survey, where we wanted to incorporate simulated data that is zero-inflated as well as overdispersed for the animal counts. Therefore, we simulated the true counts on each site as spatially correlated negative binomial random variables. The procedure follows that of Madsen and Birkes (2013) except that, in order to save simulation time, accounting for ties in the discrete negative binomial data is ignored so that the spatial correlation between the counts only approximately follows the specified correlation. SPEDRA and SPEDAR had similar biases (31.6 and 29.3, respectively), rMSPEs (524 and 513), coverages (0.906 and 0.897), and median confidence interval lengths (1,664 and 1,626) for the March 2017 Togiak stratified simulation setting. Because detection probabilities were estimated to be identical across all sites, SPEDRA and SPEDAR should be similar (though not identical) so the comparable results of the application simulation are not surprising.

5 | DISCUSSION

We have developed a spatial model that allows incorporation of imperfect detection in population abundance prediction. The model that incorporates detection probability site-wise (SPEDRA) outperforms the model that divides the predicted observed count total by the mean detection (SPEDAR) in simulation settings. Using detection data from radiocollared

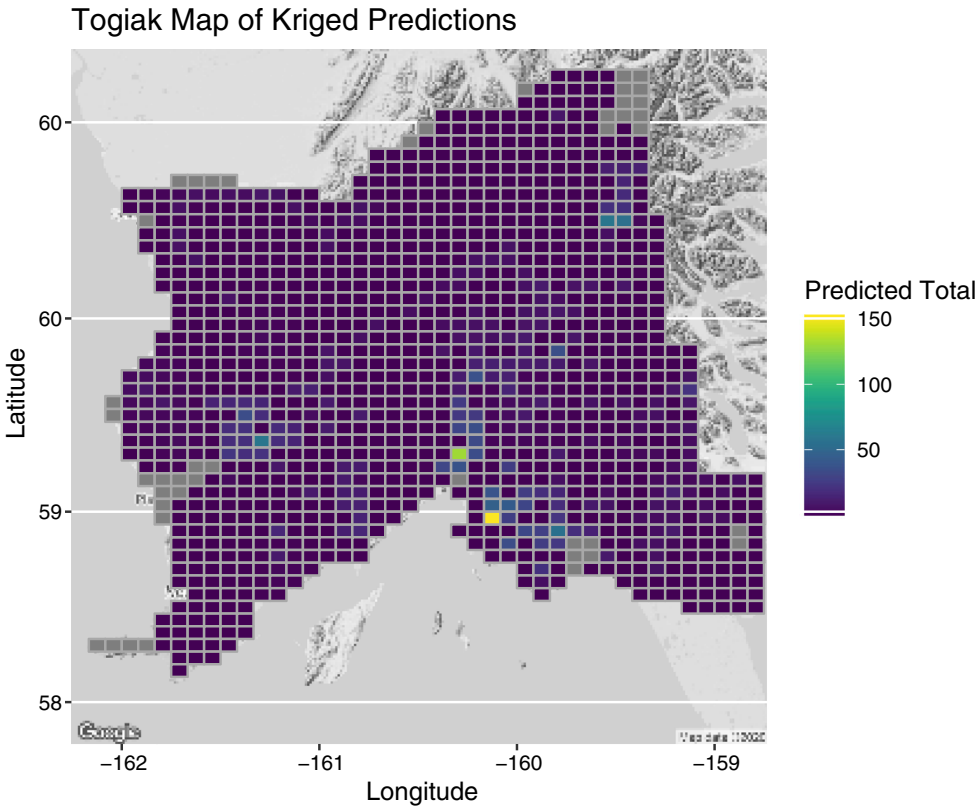

FIGURE 5 Map of Togiak displaying the predicted totals for each site in the March 2017 survey. Note the very high predicted counts in two of the sites in the south central region of Togiak, which creates a large amount of overdispersion in the high stratum

animals, we have applied the model to predict moose abundance in the Togiak National Wildlife Refuge. Both SPEDRA and SPEDAR yield similar predictions and standard errors for the Togiak moose data.

Note that SPEDRA and SPEDAR have lower rMSPE (131 and 215 units, respectively) than the SRS estimator (320 units) (Figure 2, Setting A), unless counts were simulated with little spatial correlation (Figure 2, Setting C). The SRS estimator shows no advantage in our simulations in terms of bias or rMSPE, even in settings with little spatial autocorrelation. Therefore, we only recommend it for its simplicity when there is little spatial autocorrelation and sites are chosen randomly. We estimated positive spatial correlation in one of the strata in our application (Figure 4), and unsampled sites near sites with large counts had large predicted counts (Figure 5). When no spatial autocorrelation occurs, SPEDRA/SPEDAR predictions at all unsampled sites in each stratum will be the sample mean for that stratum (either before or after adjusting for detection), and we would get the same result if using an extension of SRS to predict at unsampled sites.

SPEDRA outperformed SPEDAR for rMSPE in almost all simulation scenarios (Appendix A). In addition, a major difference in the two estimators is that SPEDRA assumes stationarity in the true counts while SPEDAR assumes stationarity in the observed counts. Checking this assumption is difficult because we never actually observe the true counts. However, we believe that it is more reasonable to assume stationarity in the true counts, not the observed counts. For example, suppose there are no covariates for the counts so that the true counts have a constant mean throughout some study area. Let the northern half of the area have low sightability, both true and estimated, and let the southern half have high sightability, both true and estimated. Then the different detection rates would induce a trend in the observed counts that would show up as autocorrelation in the spatial model of SPEDAR, but SPEDRA would correct for that trend. We do note, however, that SPEDRA has an advantage in our simulation study in that we always generate observed counts using the SPEDRA model.

In making a map of site predictions, SPEDRA predictions for sites with observed zeros are slightly positive, taking into account the possibility that a unit was missed, while SPEDAR predictions for sites with observed zeros are exactly zero, which is not realistic unless detection is perfect. In choosing between the two estimators, we recommend SPEDRA as opposed to SPEDAR because SPEDRA is more theoretically sound and it performed better in simulations, with the main exception occurring when detection probabilities were very low (Table 3).

Further research is needed on other spatial models for finite populations that incorporate detection. Fully model-based Bayesian hierarchical models are becoming faster and easier to use, but still require substantial time and expertise. Peters

et al. (2014) used a distance sampling method to estimate a moose population total in Alberta, Canada, and it would be interesting to compare SPEDRA to a distance sampling survey of roughly equal cost. Sightability trials for SPEDRA and SPEDAR may be conducted for each survey, or a static detection model could be developed from historical sightability data for a region. Christ (2011) notes the high cost of obtaining sightability data, favoring a static model, but that conditions could vary in the future, favoring fresh data for each survey. More research and practice with SPEDRA and SPEDAR will help create future improvements.

Additional information and supporting material for this article is available online at the journal's website. The data used for the March 2017 Togiak analysis is also provided.

ACKNOWLEDGEMENTS

This work is supported by grant F16AC01127 to Oregon State University from the USDI Fish and Wildlife Service. We would like to thank the Togiak National Wildlife Refuge biologists and Fish and Wildlife Service biometricians for providing the data used in the application of the proposed model as well as clear descriptions about the study area and how the data were gathered. We also thank the editors and reviewer for their constructive comments and suggestions. The findings and conclusions in the article of the author do not necessarily represent the views of the reviewers nor the National Marine Fisheries Service, NOAA, or the U.S. Fish and Wildlife Service. Any use of trade, product, or firm names does not imply an endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Zenodo at https://zenodo.org/record/3402478#.XXUbC5NKgWo, reference number [3402478] (Aderman, Higham, Ver Hoef, & Madsen, 2019). The datasets are in the "data" folder with titles MarchMoose for the count survey and MarchSight for the sightability trials.

REFERENCES

Aderman, A., Higham, M., Ver Hoef, J., & Madsen, L. (2019). *Togiak national wildlife refuge data*. Zenodo. https://464zenodo.org/record/3402478#.XXUbC5NKgWo.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Benson, A.-M., Frye, G., Maier, H., Cobb, M., Walsh, P., & Aderman, A. (2015). *Estimation of moose abundance using the geospatial population estimator combined with a sightability model on Togiak national wildlife refuge: Site-specific research protocol*. Western Alaska Landscape Conservation Cooperative: U.S. Fish and Wildlife Service.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*, 192–225.

Boertje, R. D., Keech, M. A., Young, D. D., Kellie, K. A., & Seaton, C. T. (2009). Managing for elevated yield of moose in interior Alaska. *Journal of Wildlife Management*, *73*, 314–327.

Borchers, D. L., Buckland, S. T., Goedhart, P. W., Clarke, E. D., & Hedley, S. L. (1998). Horvitz-Thompson estimators for double-platform line transect surveys. *Biometrics*, *54*(4), 1221–1237.

Boveng, P. L., Bengtson, J. L., Withrow, D. E., Cesarone, J. C., Simpkins, M. A., Frost, K. J., & Burns, J. J. (2003). The abundance of harbor seals in the Gulf of Alaska. *Marine Mammal Science*, *19*, 111–127.

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

Buckland, S. T., Anderson, D. R., Burnham, K. P., & Laake, J. (2004). *Advanced distance sampling: Estimating abundance of biological populations*. Oxford, UK: Oxford University Press.

Chan-Golston, A. M., Banerjee, S., & Handcock, M. S. (2020). Bayesian inference for finite populations under spatial process settings. *Environmetrics*, *31*, e2606.

Chiles, J.-P., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty*. New York, NY: John Wiley & Sons.

Christ, A. (2011). *Sightability correction for moose population surveys*. Juneau, Alaska: Alaska Department of Fish and Game.

Cressie, N. (1993). *Statistics for Spatial Data, Wiley series in probability and mathematical statistics: Applied probability and statistics*. New York, NY: John Wiley.

DeLong, R. A. (2006). *Geospatial population estimator software user's guide. Technical report*. Juneau, Alaska: Department of Fish and Game.

Dorfman, R. (1938). A note on the delta-method for finding variance formulae. *The Biometric Bulletin*, *1*, 92.

Eberhardt, L., & Simmons, M. (1987). Calibrating population indices by double sampling. *The Journal of Wildlife Management*, *51*(3), 665–675.

Efron, B. (1992). *Bootstrap methods: Another look at the jackknife*. In *Breakthroughs in Statistics* (pp. 569–593). New York, NY: Springer.

Fattorini, L., Corona, P., Chirici, G., & Pagliarella, M. C. (2015). Design-based strategies for sampling spatial units from regular grids with applications to forest surveys, land use, and land cover estimation. *Environmetrics*, *26*, 216–228.

Gasaway, W. C., DuBois, S. D., Reed, D. J., & Harbo, S. J. (1986). *Estimating moose population parameters from aerial surveys. Technical report*. Juneau, Alaska: University of Alaska. Institute of Arctic Biology.

Gelfand, A. E., Diggle, P. J., Fuentes, M., & Guttorp, P. (2010). *Handbook of spatial statistics*. Boca Raton, FL: CRC Press.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, *55*, 708–713.

Gould, W. R., & Pollock, K. H. (2002). *CaptureRecapture methods*. In *Encyclopedia of environmetrics* (pp. 243–251). Hoboken, NJ: John Wiley & Sons Ltd.

Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation*, *116*, 195–203.

Gwinn, D. C., Beesley, L. S., Close, P., Gawne, B., & Davies, P. M. (2016). Imperfect detection and the determination of environmental flows for fish: Challenges, implications and solutions. *Freshwater Biology*, *61*, 172–180.

Higham, N. (2019). *Spatial prediction for finite populations with ecological applications* (Ph.D. thesis, Oregon State University).

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*, 663–685.

Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. (Vol. 1, pp. 221–233). Cambridge, CA: University of California Press.

Kellie, K. A., & DeLong, R. A. (2006). *Geospatial survey operations manual. Technical report*. Juneau, Alaska: Alaska Department of Fish and Game.

Kellner, K. F., & Swihart, R. K. (2014). Accounting for imperfect detection in ecology: A quantitative review. *PLoS One*, *9*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0111436.

Kéry, M., & Schmidt, B. (2008). Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, *9*, 207–216.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*, 79–86.

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, *23*, 504–515.

Lincoln, F. C. (1930). *Calculating waterfowl abundance on the basis of banding returns, no. 118*. Washington D.C: US Department of Agriculture.

MacKenzie, D. I., Nichols, J., Sutton, N., Kawanishi, K., & Bailey, L. L. (2005). Improving inferences in population studies of rare species that are detected imperfectly. *Ecology*, *86*, 1101–1113.

Madsen, L., & Birkes, D. (2013). Simulating dependent discrete data. *Journal of Statistical Computation and Simulation*, *83*, 677–691.

Madsen, L., Dalthorp, D., Huso, M. M. P., & Aderman, A. (2020). Estimating population size with imperfect detection using a parametric bootstrap. *Environmetrics*, *31*, e2603.

Manly, B. F., McDonald, L. L., & Garner, G. W. (1996). Maximum likelihood estimation for the double-count method with independent observers. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*(2), 170–189.

McCrea, R. S., & Morgan, B. J. (2014). *Analysis of capture-recapture data*. Boca Raton, FL: CRC Press.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, *1*(62), 3–135.

Park, H., Yabuki, H., & Ohata, T. (2012). Analysis of satellite and model datasets for variability and trends in Arctic snow extent and depth, 1948-2006. *Polar Science*, *6*, 23–37.

Peters, W., Hebblewhite, M., Smith, K. G., Webb, S. M., Webb, N., Russell, M., … Anderson, R. B. (2014). Contrasting aerial moose population estimation methods and evaluating sightability in west-central Alberta, Canada. *Wildlife Society Bulletin*, *38*, 639–649.

Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Limfjord from the German Sea. *Report of the Danish Biological Station for 1895*, *6*, 1–48.

Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., & Sauer, J. R. (2002). Large scale wildlife monitoring studies: Statistical methods for design and analysis. *Environmetrics: The Official Journal of the International Environmetrics Society*, *13*, 105–119.

Royle, J. A., & Young, K. V. (2008). A hierarchical model for spatial capture-recapture data. *Ecology*, *89*, 2281–2289.

Sarndal, C.-E., Thomsen, I., Hoem, J. M., Lindley, D., Barndorff-Nielsen, O., & Dalenius, T. (1978). Design-based and model-based inference in survey sampling [with discussion and reply]. *Scandinavian Journal of Statistics*, *5*(1), 27–52.

Sólymos, P., Lele, S., & Bayne, E. (2012). Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. *Environmetrics*, *23*, 197–205.

Stevens, D. L., Jr., & Olsen, A. R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, *14*, 593–610.

Vagheggini, A., Bruno, F., & Cocchi, D. (2016). A competitive design-based spatial predictor. *Environmetrics*, *27*, 454–465.

Ver Hoef, J. (2002). Sampling and geostatistics for spatial data. *Ecoscience*, *9*, 152–161.

Ver Hoef, J. M. (2001). *Predicting finite populations from spatially correlated data*. Proceedings of the 2000 ASA Section on Statistics and the Environment (pp. 93–98). American Statistical Association.

Ver Hoef, J. M. (2008). Spatial methods for plot-based sampling of wildlife populations. *Environmental and Ecological Statistics*, *15*, 3–13.

Ver Hoef, J. M. (2012). Who invented the delta method? *The American Statistician*, *66*, 124–127.

Ver Hoef, J. M., Cameron, M. F., Boveng, P. L., London, J. M., & Moreland, E. E. (2014). A spatial hierarchical model for abundance of three ice-associated seal species in the eastern Bering Sea. *Statistical Methodology*, *17*, 46–66.

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.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, *50*(1), 1–25.

Wilm, H. G., Costello, D. F., & Klipple, G. (1944). Estimating forage yield by the double-sampling method. *Agronomy Journal*, *36*, 194–203.

Zimmerman, D. L., & Ver Hoef, J. M. (2017). The Torgegram for fluvial variography: Characterizing spatial dependence on stream networks. *Journal of Computational and Graphical Statistics*, *26*, 253–264.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Higham M, Ver Hoef J, Madsen L, Aderman A. Adjusting a finite population block kriging estimator for imperfect detection. *Environmetrics*. 2020;e2654. https://doi.org/10.1002/env.2654