Adjusting a Finite Population Block Kriging Estimator for Imperfect Detection

Matt Higham[1,2], Jay Ver Hoef[3], Lisa Madsen[4], Andy Aderman[5]

---

[1]Department of Statistics, St. Lawrence University, Canton, NY

[2]Email: mhigham@stlawu.edu

[3]Marine Mammal Laboratory, Alaska Fisheries Science Center,

NOAA Fisheries, Seattle, WA

[4]Department of Statistics, Oregon State University, Corvallis, OR

[5]U.S. Fish and Wildlife Service, Dillingham, AK

August 6, 2020

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.

KEY WORDS: geostatistics, stratified random sampling, sightability, spatial statistics

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 papers 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, 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 (e.g., Gwinn et al., 2016) to species distribution analysis (e.g., Lahoz-Monfort et al., 2014) to estimation of the abundance of rare species (e.g., MacKenzie et al., 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 (Petersen, 1896; Lincoln and others, 1930). Since their inception, mark-recapture methods have increased enormously in use and model complexity (e.g. Gould and Pollock, 2002; McCrea and Morgan, 2014; Otis et al., 1978). The basic mark-recapture can be extended to a hierarchical spatial mark-recapture method for populations with high spatial correlation (Royle and 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

1

26 or a transect line. Like mark-recapture, many studies have used distance sampling. For example,

27 Peters et al. (2014) used distance sampling for moose surveys in Canada. Buckland et al. (2001)

28 and Buckland et al. (2004) provide a more general introduction to distance sampling. Another

29 method used in large-scale abundance estimation surveys is double sampling (Wilm et al., 1944),

30 in which the surveyors sample a subset of the area of interest intensively to provide an adjustment

31 for detection (Eberhardt and Simmons, 1987; Pollock et al., 2002). In estimating total moose

32 abundance in Alaska, the double sampling method is used with stratified sampling on plot-based

33 aerial counts (Gasaway et al., 1986).

34 One major flaw in these approaches is that, though traditional mark-recapture, distance sampling,

35 and double sampling adjust for perception bias (missed animals caused by observer error), they do

36 not easily account for availability bias (missed animals caused by animals not being "available"

37 to be observed). With some studies, a separate "sightability" survey is carried out to address

38 availability bias. Madsen et al. (2020) use separate sightability data collected through radiocollars

39 and parametric bootstrapping to calculate the uncertainty in detection estimates, but assume that

40 the counts are spatially uncorrelated. Boveng et al. (2003) adjust for availability bias in estimating

41 harbor seal counts with radiotelemetry data while Ver Hoef et al. (2014) expand on the model by

42 allowing for spatial autocorrelation among the counts. Other methods only require a single survey,

43 but assume that covariates associated with abundance and detection are readily available (Sólymos

44 et al., 2012).

1.1 Finite Population Block Kriging Background

46 FPBK (Ver Hoef, 2001) was originally developed for moose surveys in Alaska. The Alaska

47 Department of Fish & Game (ADF&G) denotes the FPBK estimator as the Geospatial Population

48 Estimator (GSPE) throughout its literature, which includes an operations manual (Kellie and DeLong,

49 2006) and a software user's guide (DeLong, 2006). The method has been widely used to estimate

50 moose population totals throughout Alaska and western Canada, as 524 moose surveys have been

2

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 et al. (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, pg. 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 et al. (2015), and Vagheggini et al. (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 et al. (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

3

77 more. This is particularly useful if management is interested in small area estimation. For classical

78 sampling, there is no guarantee that adequate sampling will be done in the small area, but, with the

79 possibility of nonrandom sampling, managers have more control over the survey design (Ver Hoef,

80 2002; Kellie and DeLong, 2006).

1.2 Imperfect Detection Modeling

82 There are several sensible estimators when imperfect detection occurs. Observed counts can be

83 divided by estimated detection probabilities site-wise followed by FPBK. This adjusted estimator

84 (which we will call SPEDRA, Spatial Population Estimator with Detection: Ratio then Add) is

85 Horvitz-Thompson-like in that it uses detection probabilities in place of inclusion probabilities

86 (Horvitz and Thompson, 1952). A second possibility is to first use FPBK and then divide the

87 estimated total by the mean detection probability (SPEDAR, Spatial Population Estimator with

88 Detection: Add then Ratio). The SPEDAR method is similar to estimators by Manly et al. (1996)

89 and Borchers et al. (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 $\widehat{\pi}_i$ = 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

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

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

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

90 where w_i is the observed count on site i, $\widehat{\pi}_i$ is the estimated detection probability on site i, N is

91 the total number of sites, and n is the number of sampled sites. The estimators are identical when

92 $\pi_1 = \pi_2 = \ldots = \pi_n$.

4

93   In this very simple example, the estimators are quite different. If we know detection exactly,

94 then SPEDRA is unbiased. However, similar to Horvitz-Thompson estimators (Horvitz and Thompson,

95 1952), the variance of the estimator has the potential to become very large, particularly for small

96 detection probabilities. Therefore, unless the detection probabilities are all equal, we anticipate

97 that SPEDRA will be less biased than SPEDAR, but could have a much higher variance. On the

98 other hand, SPEDAR is obviously biased, but may be more robust than SPEDRA.

1.3 Goals and Organization

100 Our goals are to (1) develop a model-based (geostatistical) approach to predict the total abundance

101 in a region of interest using a version of FPBK that adjusts for imperfect detection, (2) validate

102 and compare the SPEDRA and SPEDAR models through a simulation study, and (3) apply the

103 methods to a moose data set in Togiak National Wildlife Refuge. The adjusted FPBK estimator is

104 useful for population abundance prediction problems where there is spatial autocorrelation in the

105 counts and detection of the animals or plants in a sampling unit is not perfect.

106   In Section 2, we review FPBK assuming perfect detection and then develop the SPEDRA and

107 SPEDAR models for incorporating imperfect detection. Next, in Section 3, we present results from

108 a simulation study before applying the two estimators to real data from a March 2017 Togiak moose

109 survey in Section 4. We conclude in Section 5 with some remarks comparing the two estimators

110 as well as some discussion on possible extensions to the models developed here.

2   Adjusting the FPBK Model for Imperfect Detection

112 We use the following three results frequently in the adjusted FPBK estimators. In particular,

113 the expression for the variance of a product of two random variables (1) will be quite useful for

114 SPEDRA because we will model an observed count at a particular site as the product of the true

115 count and the estimated detection probability at that site.

5

2.1 Basic Equations

117 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.

118 Also, let $\mathbf{x}$ be independent of $\mathbf{y}$. Then, an extension of the variance of the product of two random

119 variables (Goodman, 1960) is the following multivariate version (Ver Hoef et al., 2014):

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

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

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

122 the model,

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

123

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

2.2 Background of FPBK

125 The following is a brief summary of Ver Hoef (2001, 2008), which proposes the FPBK model

126 assuming perfect detection. If D is a spatial lattice indexed on a finite set of points $i = 1, 2, \ldots, N$,

127 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

128 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 $Z(\mathbf{s}_i)$

129 for $i = 1, \ldots, N$ and $\mathbf{b}$ can be a vector where each element is 1 if we want to predict the population

130 total, a vector where each element is $1/N$ if we want to predict the population mean, or a mix of

131 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

132 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

133 data for the sampled locations in D and $\mathbf{a}'$ is a vector of weights.

134 Let $Y(\mathbf{s}_i) = \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta} + \boldsymbol{\epsilon}(\mathbf{s}_i)$ be a spatial random field. The error term $\boldsymbol{\epsilon}(\mathbf{s}_i)$ is a spatial error

135 process with $\mathbf{0}$ mean and a spatially autocorrelated covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ depends

136 on just a few parameters $\boldsymbol{\theta}$. Also, $\mathbf{x}'(\mathbf{s}_i)$ is a vector of covariates at location $\mathbf{s}_i$, and $\boldsymbol{\beta}$ is a parameter

6

137 vector. For the remainder of this paper, we use the exponential covariance model for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. That

138 is, the i, j^{th} entry for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is

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

139 where $h_{i,j}$ is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. Note, however, that any spatial covariance matrix

140 could be used (Chiles and Delfiner, 1999, pg. 80 - 93).

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

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

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

141 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.

142 A spatial random field with exponential autocovariance function (4) is second-order stationary,

143 meaning that the mean is constant and the covariance is a function only of the separation vector

144 between any two locations, and it is isotropic, meaning it is a function of distance only (Cressie,

145 1993; Gelfand et al., 2010). However, the results that follow can be extended to other types

146 of autocovariance models, including models that include one or two extra parameters to model

147 anisotropy (Chiles and Delfiner, 1999, pg. 94).

148 Note that we could also use an autoregressive model, such as a conditional autoregressive

149 (CAR) model (Besag, 1974) to model the spatial process. In general, such models are natural for

150 finite sets of spatial data (Cressie, 1993, pg. 8) and can have a large computational advantage

151 because the inverse covariance matrix is specified (e.g., Ver Hoef et al., 2018). However, when

152 there are missing data, as with incomplete sampling, then obtaining the marginal distribution of

153 the sampled data requires inverting the precision matrix, forming the covariance matrix for the

154 data from the subset of rows and columns, and then inverting again for prediction (Ver Hoef et al.,

155 2018). Additionally, the autocorrelation and variances are nonstationary. Therefore, we focus on

156 the geostatistical model.

7

157 Using the laws of conditional expectation, conditional variance (2), conditional covariance (3),

158 and the independence of $Z(\mathbf{s}_i)|Y(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)|Y(\mathbf{s}_j)$, we obtain $\mathrm{E}[Z(\mathbf{s}_i)] = \boldsymbol{\mu}(\mathbf{s}_i)$, $\mathrm{var}[Z(\mathbf{s}_i)] =$

159 $\theta_3 + \Sigma_{i,i}$, and $\mathrm{cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] = \Sigma_{i,j}$, where $\boldsymbol{\mu}(\mathbf{s}_i) = \mathbf{x}'(\mathbf{s}_i)\boldsymbol{\beta}$ and $\Sigma_{i,j}$ is the i, j^{th} element of

160 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

161 For simpler notation, we can represent $\mathbf{z}$, the vector of the random variables $Z(\mathbf{s}_i)$, using the

162 following linear model, with $\mathbf{z}_s$ denoting the vector of $\{Z(\mathbf{s}_i)\}$ for the sampled locations in D and

163 $\mathbf{z}_u$ denoting the vector of $\{Z(\mathbf{s}_i)\}$ for the unsampled locations in D:

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

164 where $\mathbf{X}_s$ and $\mathbf{X}_u$ are the design matrices for the sampled and unsampled sites, respectively, and

165 $\boldsymbol{\delta}_s$ and $\boldsymbol{\delta}_u$ are zero-mean random errors for the sampled and unsampled sites. Denote $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ as

166 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

$$
\mathrm{var}(\boldsymbol{\delta}) \equiv \mathbf{D} = \mathrm{diag}(\boldsymbol{\theta}_3) + \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \mathrm{diag}(\boldsymbol{\theta}_3) + \boldsymbol{\Sigma}_{ss} & \boldsymbol{\Sigma}_{su} \\ \boldsymbol{\Sigma}'_{su} & \mathrm{diag}(\boldsymbol{\theta}_3) + \boldsymbol{\Sigma}_{uu} \end{pmatrix} \equiv \begin{pmatrix} \mathbf{D}_{ss} & \mathbf{D}_{su} \\ \mathbf{D}'_{su} & \mathbf{D}_{uu} \end{pmatrix}, \tag{6}
$$

167 where $\mathrm{diag}(\boldsymbol{\theta}_3)$ is the diagonal matrix with diagonal elements θ_3, $\boldsymbol{\Sigma}_{ss}$ is the submatrix of $\mathrm{var}(\boldsymbol{\epsilon}) \equiv$

168 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ for the sampled locations, $\boldsymbol{\Sigma}_{su}$ is the covariance of the sampled sites with the unsampled

169 sites, and $\boldsymbol{\Sigma}_{uu}$ is the submatrix of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ for the unsampled locations.

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

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

171 where $\mathbf{b}_s$ and $\mathbf{b}_u$ are subvectors of $\mathbf{b}$ corresponding to the sampled and unsampled locations in D,

172 respectively, and $\hat{\mathbf{z}}_u$ is the BLUP of $\mathbf{z}_u$,

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

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

174 least squares estimator of $\boldsymbol{\beta}$.

8

175   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

$$\mathbf{G} = (\mathbf{D}_{ss})\mathbf{b}_s + \mathbf{D}_{su}\mathbf{b}_u,$$

$$\mathbf{H} = \mathbf{X}'\mathbf{b} - \mathbf{X}'_s(\mathbf{D}_{ss})^{-1}\mathbf{G},$$

$$\mathbf{E} = (\mathbf{X}'_s(\mathbf{D}_{ss})^{-1}\mathbf{X}_s)^{-1}.$$

176 The derivation of equations (7) - (9) can be found in Ver Hoef (2008). Note that the above model

177 assumes that all animals or plants are detected so that there is no difference in the number of

178 observed animals or plants and the true number of animals or plants at a particular site. In the

179 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 (SPEDRA)

180

181 The first proposed model adjusts the observed counts for each sample unit by their estimated

182 detection probabilities prior to spatial modeling. We first motivate our model by considering

183 $W(\mathbf{s}_i)|\{Z(\mathbf{s}_i), P(\mathbf{s}_i)\} \sim \text{Binomial}(Z(\mathbf{s}_i), P(\mathbf{s}_i))$, where $W(\mathbf{s}_i)$ is the observed count, $P(\mathbf{s}_i)$ is

184 the estimated probability of detection of a single animal or plant, and $Z(\mathbf{s}_i)$ is the random variable

185 for the true count at location $\mathbf{s}_i$, defined in (5), which assumes perfect detection. Site s_i may be

186 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),

9

(2), and (3),

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

$$+ \boldsymbol{\mu}_s \boldsymbol{\mu}_s' \odot \mathbf{V}_{ss} + \mathbf{D}_{ss} \odot \mathbf{V}_{ss}, \tag{10}$$

187 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

188 could model $\mathbf{p}_s$ with spatial random effects in much the same way as we did for abundance,

189 it may often be the case that detection is an observation process that is under our control with

190 well-known covariates that mitigate spatial autocorrelation. On the other hand, the abundance of

191 natural resources, like animals or plants, are often inherently aggregated on the landscape with

192 autocorrelated residual error. Additionally, in our application, we have very few nondetections for

193 modeling spatial autocorrelation. Hence, in our simulations, and application, we assume detection

194 variables are spatially independent.

195   We can also write the covariance between the observed sample counts $\mathbf{w}_s$ and the true counts

196 on all of the sites $\mathbf{z}$ as $\text{cov}(\mathbf{w}_s, \mathbf{z}') \equiv \mathbf{R} = \boldsymbol{\pi}_s \odot \begin{pmatrix} \mathbf{D}_{ss} & \mathbf{D}_{su} \end{pmatrix}$.

197   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

198 weights are applied to the observed counts $\mathbf{w}_s$, not the true counts $\mathbf{z}_s$ because the $\mathbf{z}_s$ are unknown.

199 Let $M_{\mathbf{a}} = \text{E}[(\mathbf{a}'\mathbf{w}_s - \mathbf{b}'\mathbf{z})^2]$ be the Mean Square Prediction Error (MSPE) for any particular $\mathbf{a}$. Our

200 goal is to find the BLUP; similar to Ver Hoef (2008), we want the "best" weights $\boldsymbol{\lambda}$ such that

201   1. $\text{E}(\boldsymbol{\lambda}'\mathbf{w}_s) = \text{E}(\mathbf{b}'\mathbf{z})$ and

202   2. $M_{\mathbf{a}} - M_{\boldsymbol{\lambda}}$ is non-negative for all $\mathbf{a} \neq \boldsymbol{\lambda}$.

203   First, we can restrict ourselves to the class of all unbiased estimators such that $\text{E}(\mathbf{a}'\mathbf{w}_s) = \text{E}(\mathbf{b}'\mathbf{z})$

204 for all $\boldsymbol{\beta}$ in the marginal linear model for $\mathbf{z}$ in equation (5). This implies that $\mathbf{a}'(\boldsymbol{\pi}_s \odot \mathbf{X}_s)\boldsymbol{\beta} = \mathbf{b}'\mathbf{X}\boldsymbol{\beta}$

205 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 $\boldsymbol{\lambda}$ that makes $M_{\mathbf{a}} - M_{\boldsymbol{\lambda}}$ non-negative such that the unbiasedness

10

constraint holds. If we minimize the MSPE, then we obtain the prediction equations

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

206 where $\mathbf{L}$ is the LaGrange multiplier in the system of equations. Note the similarity in form between

207 these prediction equations and the prediction equations in standard block kriging and also in FPBK

208 assuming perfect detection in Ver Hoef (2008). Solving for $\boldsymbol{\lambda}$,

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

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

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

210   The above model could also be cast as a hierarchical model, $[\mathbf{z}_s, \mathbf{y}_s, \mathbf{p}_s, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\theta}_3 | \mathbf{w}_s, \mathbf{K}, \mathbf{X}]$,

211 where $\mathbf{K}$ is a vector of 1's and 0's resulting from the sightability trials. A natural approach would

212 be to adopt a Bayesian model with prior distributions and MCMC sampling. However, such an

213 approach can be time consuming and relies on a great amount of statistical expertise. Our goal

214 is to have a fairly automatic way for a non-statistician user to obtain results quickly and reliably.

215 In several applications that use FPBK so far, dozens of surveys are completed annually, and each

216 one cannot be a carefully tuned analysis. The present approach is fast and robust, relying only on

217 moment assumptions. However, for the interested reader, we have also analyzed these data with a

218 Bayesian hierarchical model (Higham, 2019).

219   2.4   The Logistic Regression Model

220 In order to use the above model, we must have some way to estimate the detection probabilities

221 for the sites in the population of interest. In some animal surveys, there are radiocollared animals

222 and separate sightability trials to model the imperfect detection of animals. Suppose that we have

223 n radiocollared animals to be used for the sightability trials, indexed $i = 1, \ldots, n$. For each i,

224 $K_i \sim \mathrm{Bernoulli}(\pi_i)$, where K_i is the random variable for whether or not a radiocollared animal is

11

225 detected. Also suppose that we have a design matrix $\mathbf{U}$ where each row contains covariates like

226 habitat conditions or time spent surveying for predicting the detection probability in a particular

227 site. Denote $\mathbf{U}_r$ as the design matrix for the sites with the radiocollared animals and $\boldsymbol{\pi}_r$ as the

228 vector of probabilities that a radiocollared animal is sighted. Then we assume that we have the

229 regression model

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

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

231 2.5 Estimation

232 We next discuss estimation of the quantities in (11) and (12). If we denote $\mathbf{U}_s$ as the rows of $\mathbf{U}$

233 corresponding to the sampled sites in the animal count survey, then $\text{logit}(\widehat{\boldsymbol{\pi}}_s) = \mathbf{U}_s \hat{\boldsymbol{\gamma}}$, where $\hat{\boldsymbol{\gamma}}$

234 is the estimator for $\boldsymbol{\gamma}$ in standard logistic regression. The estimated detection probabilities $\widehat{\boldsymbol{\pi}}_s$ are

235 then obtained through the element-wise inverse logit transformation on $\mathbf{U}_s \hat{\boldsymbol{\gamma}}$. The $\widehat{\boldsymbol{\pi}}_s$ take the place

236 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

237 a method other than logistic regression.

238 We investigated using either nonparametric bootstrapping (Efron (1992)) or the delta method

239 (Dorfman (1938), Ver Hoef (2012)) to estimate $\mathbf{V}_{ss}$, the covariance matrix of $\widehat{\boldsymbol{\pi}}_s$, ultimately

240 finding that bootstrapping produced more accurate results in simulations. Letting $\mathbf{U}_{all}$ denote the

241 matrix combining the design matrix $\mathbf{U}_r$ and the sightability response vector $(K_1, K_2, \ldots, K_n)'$,

242 we perform 1400 nonparametric bootstraps of the rows of $\mathbf{U}_{all}$, estimate $\hat{\boldsymbol{\gamma}}_{boot}$ for each bootstrap,

243 calculate $\widehat{\boldsymbol{\pi}}_{s,boot}$, and find the empirical covariance of the 1400 $\widehat{\boldsymbol{\pi}}_{s,boot}$ vectors to obtain an estimator

244 of $\mathbf{V}_{ss}$.

245 We use maximum likelihood to obtain estimates for $\boldsymbol{\mu}_s$, denoted $\hat{\boldsymbol{\mu}}_s$, and the covariance parameters

246 in $\mathbf{D}$, denoted $\hat{\mathbf{D}}$, assuming the exponential covariance structure defined in (4). We use a normal

247 likelihood model for the $\mathbf{w}_s$ with the covariance matrix in equation (10), plugging in $\widehat{\boldsymbol{\pi}}_s$ and $\hat{\mathbf{V}}_{ss}$.

12

₂₄₈ Note that, for ease of estimation, we are using a misspecified normal likelihood for the correlated

₂₄₉ binomial counts. Our estimators for the mean and covariance parameters from misspecified maximum

₂₅₀ likelihood are consistent estimators for the $\boldsymbol{\mu}_s$ and $\boldsymbol{\theta}$ parameters that minimize the Kullback-Leibler

₂₅₁ distance between the misspecified distribution and the true distribution (White, 1982; Kullback

₂₅₂ and Leibler, 1951). Huber and others (1967) and White (1982) give additional examples and

₂₅₃ properties of maximum likelihood estimators under distribution misspecification. Then, we have

₂₅₄ $\hat{\mathbf{D}} = \mathrm{diag}(\hat{\boldsymbol{\theta}}_3) + \hat{\boldsymbol{\Sigma}}$, $\hat{\mathbf{X}}_s^* = (\hat{\boldsymbol{\pi}}_s \odot \mathbf{X}_s)$, and, partitioning $\hat{\mathbf{D}}$ as in equation (6), $\hat{\mathbf{R}} = \hat{\boldsymbol{\pi}}_s \odot$

₂₅₅ $\begin{pmatrix} \hat{\mathbf{D}}_{ss} & \hat{\mathbf{D}}_{su} \end{pmatrix}$. The estimated variance of the observed counts is

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

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

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

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

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

₂₅₆ 2.6 Spatial Population Estimator with Detection: Add then Ratio (SPEDAR)

₂₅₇ The second proposed model uses the FPBK estimator defined in (7) on the observed counts and

₂₅₈ then divides by the estimated mean detection probability. Here, we use the notation **obs** as a

₂₅₉ reminder that kriging uses the observed counts. Using this notation, we can use the standard FPBK

₂₆₀ in (Ver Hoef, 2008) on the observed counts to obtain

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

₂₆₁ Denote this predictor for the observed counts as $\hat{\tau}_{\mathbf{obs}}$. Also define $\hat{\bar{\pi}} = \sum_{j=1}^{m} \dfrac{\hat{\pi}_j}{m}$, the estimated mean

₂₆₂ detection probability where m is the number of sites in the count survey and j indexes the detection

₂₆₃ probabilities for the sites in the count survey. Then, we adjust estimator (15) by the mean detection

13

²⁶⁴ probability across the sampled sites to obtain the SPEDAR estimator

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

²⁶⁵ We obtain the prediction variance using a plug-in estimator in the scalar form of equation (1),

$$\mathrm{var}(\hat{\tau}_{\mathbf{obs}}\hat{\hat{\pi}}^{-1}) = (\mathrm{E}(\hat{\tau}_{\mathbf{obs}}))^2\,\mathrm{var}(\hat{\hat{\pi}}^{-1}) + \left(\mathrm{E}(\hat{\hat{\pi}}^{-1})\right)^2\mathrm{var}(\hat{\tau}_{\mathbf{obs}}) + \mathrm{var}(\hat{\tau}_{\mathbf{obs}})\,\mathrm{var}(\hat{\hat{\pi}}^{-1}). \tag{17}$$

²⁶⁶ We use bootstrapping to obtain an approximation for $\mathrm{E}(\hat{\hat{\pi}}^{-1})$ and $\mathrm{var}(\hat{\hat{\pi}}^{-1})$, resampling the rows

²⁶⁷ of $\mathbf{U}_{all}$ as in the bootstrapping for SPEDRA estimator. $\hat{\mathrm{E}}(\hat{\hat{\pi}}^{-1})$ and $\widehat{\mathrm{var}}(\hat{\hat{\pi}}^{-1})$ are then the mean

²⁶⁸ and variance of the 1400 bootstrapped quantities of $\hat{\hat{\pi}}^{-1}$. Putting the bootstrap estimates back into

²⁶⁹ equation (17), as well as using the observed counts in equation (7) as an estimate for $\mathrm{E}(\hat{\tau}_{\mathbf{obs}})$ and

²⁷⁰ using the observed counts in equation (9) to estimate $\mathrm{var}(\hat{\tau}_{\mathbf{obs}})$, gives an approximation to the

²⁷¹ prediction variance for the SPEDAR estimator.

²⁷² It is unclear whether SPEDAR is better than SPEDRA in terms of unbiasedness and prediction

²⁷³ variance, but results from distance sampling (Manly et al., 1996; Borchers et al., 1998) and weighted

²⁷⁴ semivariograms (Zimmerman and Ver Hoef, 2017) indicate that there could be substantial reduction

²⁷⁵ in variance at the expense of a little more bias. Therefore, we include both the SPEDAR and the

²⁷⁶ SPEDRA models here in an effort to compare the two in a simulation study in the FPBK setting.

²⁷⁷ Note that we make slightly different assumptions about the nature of the spatial correlation in

²⁷⁸ the counts in the SPEDRA and SPEDAR models. For SPEDRA, we assume that the *total* counts

²⁷⁹ of animals per site are spatially autocorrelated according to some model while, for SPEDAR, we

²⁸⁰ assume that the *observed* counts of animals per site are spatially autocorrelated according to some

²⁸¹ model. These assumptions are slightly different because, for SPEDRA we incorporate detection

²⁸² before performing kriging while, for SPEDAR, we perform kriging on the observed counts. The

²⁸³ implications of these assumptions for which model might be better to use are given in Section 5.

14

3 Simulation Study

285 We ran a Latin Square simulation experiment to cover a wide range of possible detection probabilities.

286 A 3×3 Latin Square with three sightability sample sizes $n_{det} \in \{60, 100, 200\}$, three possible

287 levels of mean detection $\gamma \in \{0.25, 0.5, 0.75\}$, and three possible range parameters $\theta_2 \in \{0.1, 5, 30\}$

288 was replicated for each combination of two population means $\mu \in \{4, 8\}$ and two detection

289 variances $\gamma_1 \in \{1, 4\}$ for low and high detection variances, respectively (Appendix A). γ_{0cat} refers

290 to the level of mean detection (Low has a mean detection of 0.25, Medium has a mean detection of

291 0.5, and High has a mean detection of 0.75). We simulated counts on a regular 1×1 unit grid. The

292 total number of sites was 400, a 20×20 grid. For each simulation setting, we ran 1400 simulations

293 with the partial sill fixed at 2, the nugget fixed at 0.02, and the count survey sample size fixed at

294 100.

295 We simulated spatially correlated normal random variables with the specified mean and covariance

296 parameters assuming the exponential spatial autocorrelation model (rounding to the nearest non-negative

297 integer) as the true counts. We then simulated the true detection probabilities for each site using the

298 logistic regression parameters in Figure 1 with covariates simulated as Unif(0, 1) random variables.

299 The observed counts were then simulated as binomial random variables with sizes equal to the

300 true counts and probabilities equal to the true detection probabilities on a random sample of sites

301 corresponding to the number of sites in the moose count survey. The sightability data were formed

302 by generating n_{det} Bernoulli random variables using the true detection probabilities.

303 The full results of the simulations are given in Appendix A. The coverage across all settings for

304 SPEDRA is 0.882 with a minimum of 0.841 while the coverage across all settings for SPEDAR

305 is 0.911 with a minimum of 0.882. We also included a simple random sample (SRS) estimator

306 in the simulations to compare with SPEDRA and SPEDAR. SRS was calculated as $\frac{N}{m} \sum_{i=1}^{m} \frac{w_i}{\hat{\pi}_i}$,

307 where m denotes the number of sites in the count survey and N denotes the total number of sites

308 in the study region. Results from three of the scenarios are shown in Figure 2. As expected, SRS

15

309 performs similarly to the two spatial estimators when the range parameter is small, as seen in

310 Setting C. On the other hand, in Setting A, where there is a large amount of spatial correlation,

311 SRS performs much more poorly than the spatial estimators. In all settings, SRS had a comparable

312 or worse root Mean Square Prediction Error (rMSPE) than the SPEDRA and SPEDAR estimators

313 so we only consider SPEDRA and SPEDAR henceforth.

314 The results from the simulations are summarized as linear models, with the experimental units

315 as the individual simulation settings, in Tables 1, 2, and 3. Table 1 uses SPEDRA rMSPE as the

316 response with the simulation factors as predictors. All of the factors considered are significant at

317 a significance level of 0.05 except for the level of the range parameter, θ_2. Table 2 uses SPEDAR

318 rMSPE as the response with the simulation factors as predictors. The results are similar to those

319 in Table 1, but now the detection variance factor ($\gamma_1 = 4$) is only weakly significant, indicating

320 that, for SPEDAR, the variability in detection is a less important influence on rMSPE than it is for

321 SPEDRA rMSPE. The signs of the point estimates in Table 1 and Table 2 are identical, showing

322 that rMSPE tends to increase for: (1) larger means, (2) larger ranges, (3) lower mean detection, (4)

323 lower detection variability, and (5) lower detection sample size.

324 Table 3 uses the efficiency, rMSPE of SPEDRA divided by rMSPE of SPEDAR, as the response

325 with the simulation factors as predictors. An efficiency greater than 1 indicates that SPEDAR is

326 better than SPEDRA while an efficiency less than 1 indicates that SPEDRA is better than SPEDAR.

327 SPEDRA does better than SPEDAR for higher levels of detection variability and higher levels

328 of mean detection. The form of the detection function therefore drives the change in efficiency

329 between SPEDRA and SPEDAR. As detection gets smaller, the efficiency increases due to the

330 instabilities in dividing by small individual detections.

16

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 ten years and future changes are not expected to improve the situation (Park et al., 2012).

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 non-randomly 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 finite population block kriging 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.

17

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. Additionally, 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 cross-correlation 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 $\widehat{\boldsymbol{\pi}}_s$ vectors for each stratum will typically be positively correlated. Additionally, 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

18

377 moderate amount of spatial autocorrelation in the high stratum (Figure 4). We also checked the

378 assumption of stationarity in the high stratum by comparing variograms for different quadrants of

379 the study area. In this case, there was no strong violation of stationarity in the high stratum while

380 the low stratum had too few data points to construct multiple semi-variograms for different regions.

381 Note that each semi-variogram uses the naive $\hat{z}_i = w_i/\widehat{\pi}_i$ as observations to compute the empirical

382 variogram. Therefore, particularly in the low stratum where there is a large abundance of zeroes,

383 the variogram appears to fit the "observations" poorly. The zeroes are not inflated at all in our naive

384 "observations," which drives up the variance in the empirical variogram while our fitted variogram

385 model using maximum likelihood estimation accounts for some inflation of the zero counts.

386 Figure 5 shows a few sites with very high predicted counts in the high stratum while most

387 other sites are predicted to have few or no moose. The SPEDRA estimator yielded a total estimate

388 of 3658 moose while SPEDAR estimator yielded a total estimate of 3645. Approximate 90%

389 normal-based prediction intervals for the total number of moose at the Togiak Wildlife Refuge

390 were (2839, 4478) moose using SPEDRA and (2785, 4505) moose using SPEDAR. In this case,

391 both the estimates and the standard errors were approximately equal for the two estimators.

4.4 Application Simulation

393 In addition to the full simulation study in Section 3, we also ran 2000 simulations for a stratified

394 scenario with counts and detection similar to the March 2017 Togiak survey, where we wanted

395 to incorporate simulated data that is zero-inflated as well as overdispersed for the animal counts.

396 Therefore, we simulated the true counts on each site as spatially correlated negative binomial

397 random variables. The procedure follows that of Madsen and Birkes (2013) except that, in order to

398 save simulation time, accounting for ties in the discrete negative binomial data is ignored so that

399 the spatial correlation between the counts only approximately follows the specified correlation.

400 SPEDRA and SPEDAR had similar biases (31.6 and 29.3, respectively), rMSPEs (524 and 513),

401 coverages (0.906 and 0.897), and median confidence interval lengths (1664 and 1626) for the

19

402 March 2017 Togiak stratified simulation setting. Because detection probabilities were estimated to
403 be identical across all sites, SPEDRA and SPEDAR should be similar (though not identical) so the
404 comparable results of the application simulation are not surprising.

405 **5 Discussion**

406 We have developed a spatial model that allows incorporation of imperfect detection in population
407 abundance prediction. The model that incorporates detection probability sitewise (SPEDRA)
408 outperforms the model that divides the predicted observed count total by the mean detection
409 (SPEDAR) in simulation settings. Using detection data from radiocollared animals, we have
410 applied the model to predict moose abundance in the Togiak National Wildlife Refuge. Both
411 SPEDRA and SPEDAR yield similar predictions and standard errors for the Togiak moose data.

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

422   SPEDRA outperformed SPEDAR for rMSPE in almost all simulation scenarios (Appendix A).
423 Additionally, a major difference in the two estimators is that SPEDRA assumes stationarity in the
424 true counts while SPEDAR assumes stationarity in the observed counts. Checking this assumption
425 is difficult because we never actually observe the true counts. However, we believe that it is more
426 reasonable to assume stationarity in the true counts, not the observed counts. For example, suppose

20

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 occuring 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.

21

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 et al., 2019). The datasets are in the "data" folder with titles MarchMoose for the count survey and MarchSight for the sightability trials.

22

References

Aderman, A., Higham, M., Ver Hoef, J., and Madsen, L. (2019), "Togiak National Wildlife Refuge Data," *Zenodo*.

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., and 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," *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., and 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., and Hedley, S. L. (1998), "Horvitz-Thompson estimators for double-platform line transect surveys," *Biometrics*, 1221–1237.

Boveng, P. L., Bengtson, J. L., Withrow, D. E., Cesarone, J. C., Simpkins, M. A., Frost, K. J., and 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., and Laake, J. (2001), *Introduction to Distance Sampling*, Oxford: Oxford University Press.

23

— (2004), *Advanced Distance Sampling: Estimating Abundance of Biological Populations*, Oxford: Oxford University Press.

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

Chiles, J.-P. and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: John Wiley & Sons.

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

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

DeLong, R. A. (2006), "Geospatial population estimator software user's guide," Tech. rep., 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. and Simmons, M. (1987), "Calibrating population indices by double sampling," *The Journal of Wildlife Management*, 665–675.

Efron, B. (1992), "Bootstrap methods: another look at the jackknife," in *Breakthroughs in Statistics*, Springer, pp. 569–593.

Fattorini, L., Corona, P., Chirici, G., and 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.

24

Gasaway, W. C., DuBois, S. D., Reed, D. J., and Harbo, S. J. (1986), "Estimating moose population parameters from aerial surveys," Tech. rep., University of Alaska. Institute of Arctic Biology.

Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P. (2010), *Handbook of spatial statistics*, CRC press.

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

Gould, W. R. and Pollock, K. H. (2002), "CaptureRecapture Methods," in *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd, pp. 243–251.

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

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

Higham (2019), "Spatial Prediction for Finite Populations with Ecological Applications," Ph.D. thesis, Oregon State University.

Horvitz, D. G. and 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. and others (1967), "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, University of California Press, vol. 1, pp. 221–233.

Kellie, K. A. and DeLong, R. A. (2006), "Geospatial survey operations manual," Tech. rep., Alaska Department of Fish and Game.

25

Kellner, K. F. and Swihart, R. K. (2014), "Accounting for imperfect detection in ecology: a quantitative review," *PLoS One*, 9.

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

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

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

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

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

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

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

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

McCrea, R. S. and Morgan, B. J. (2014), *Analysis of capture-recapture data*, CRC Press.

26

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

Park, H., Yabuki, H., and Ohata, T. (2012), "Analysis of satellite and model datasets for variability and trends in Arctic snow extent and depth, 19482006," *Polar Science*, 6, 23–37.

Peters, W., Hebblewhite, M., Smith, K. G., Webb, S. M., Webb, N., Russell, M., Stambaugh, C., and 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., and 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. and Young, K. V. (2008), "A hierarchical model for spatial capturerecapture data," *Ecology*, 89, 2281–2289.

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

Sólymos, P., Lele, S., and 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 Jr, D. L. and Olsen, A. R. (2003), "Variance estimation for spatially balanced samples of environmental resources," *Environmetrics*, 14, 593–610.

27

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

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

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

581 — (2008), "Spatial methods for plot-based sampling of wildlife populations," *Environmental and*
582 *Ecological Statistics*, 15, 3–13.

583 — (2012), "Who invented the delta method?" *The American Statistician*, 66, 124–127.

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

587 Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., and Fortin, M.-J. (2018), "Spatial
588 autoregressive models for statistical inference from ecological data," *Ecological Monographs*,
589 88, 36–59.

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

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

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

28

Table 1: Linear model results with SPEDRA rMSPE as the response and the simulation settings as experimental units. γ_{0cat} 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 $\theta_2 = 0.1$, γ_{0cat} = High, logistic regression slope $\gamma_1 = 1$, and detection sample size $n_{det} = 60$.

Parameter	Estimate	Std. Error	p-value
Intercept	176.61	34.55	< 0.0001
$\mu = 8$	146.28	23.03	< 0.0001
$\theta_2 = 30$	11.58	28.21	0.6846
$\theta_2 = 5$	7.83	28.21	0.7834
$\gamma_{0cat} = $ Low	261.83	28.21	< 0.0001
$\gamma_{0cat} = $ Medium	89.33	28.21	0.0038
$\gamma_1 = 4$	-90.72	23.03	0.0005
$n_{det} = 100$	-72.50	28.21	0.0160
$n_{det} = 200$	-138.33	28.21	< 0.0001

29

Table 2: Linear model results with SPEDAR rMSPE as the response and the simulation settings as experimental units. γ_{0cat} 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 $\theta_2 = 0.1$, γ_{0cat} = High, logistic regression slope $\gamma_1 = 1$, and detection sample size $n_{det} = 60$.

| Parameter	Estimate	Std. Error	p-value
Intercept	176.92	33.13	< 0.0001
$\mu = 8$	169.94	22.08	< 0.0001
$\theta_2 = 30$	17.17	27.05	0.5310
$\theta_2 = 5$	13.50	27.05	0.6217
γ_{0cat} = Low	256.83	27.05	0.0000
γ_{0cat} = Medium	89.58	27.05	0.0026
$\gamma_1 = 4$	-42.50	22.08	0.0649
$n_{det} = 100$	-82.50	27.05	0.0051
$n_{det} = 200$	-162.08	27.05	< 0.0001

30

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. γ_{0cat} 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 $\theta_2 = 0.1$, γ_{0cat} = High, logistic regression slope $\gamma_1 = 1$, and detection sample size $n_{det} = 60$.
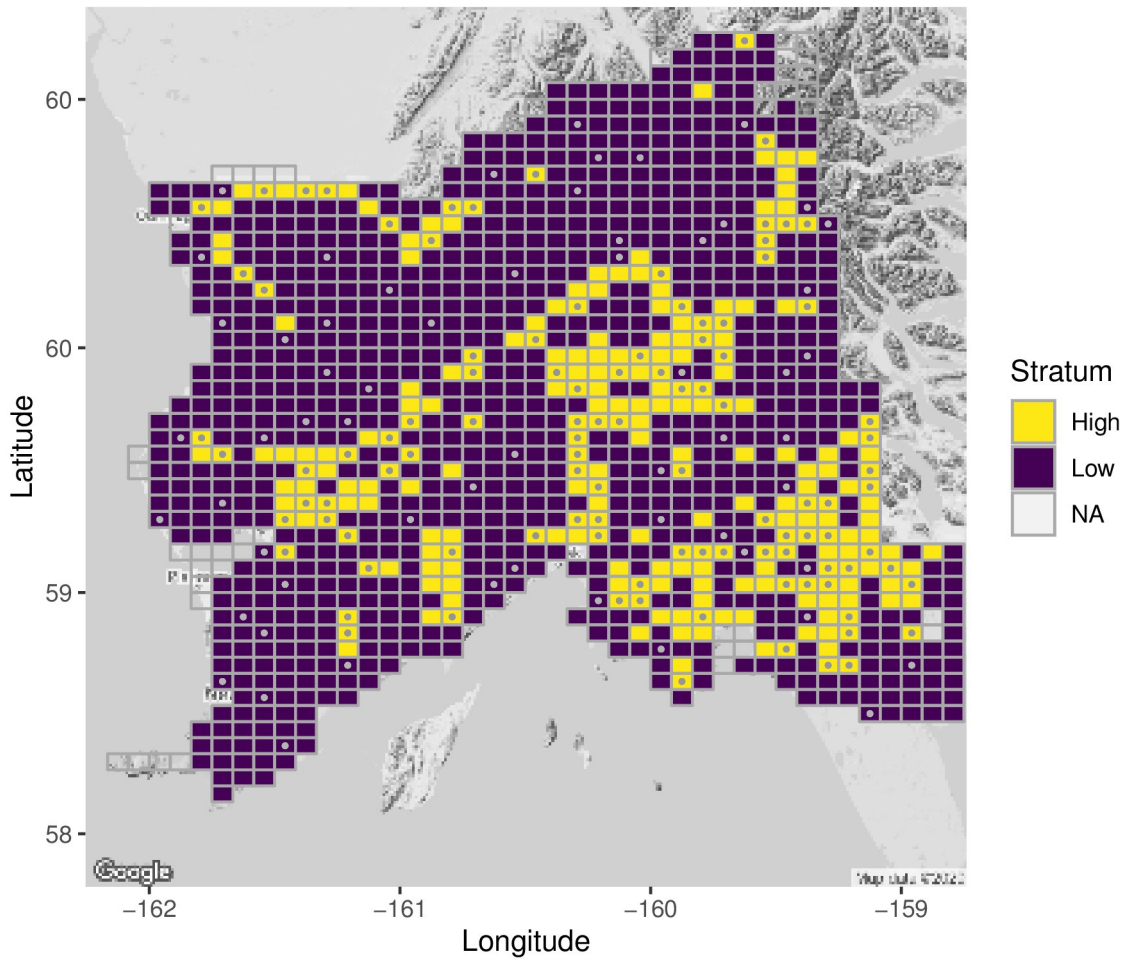
Parameter	Estimate	Std. Error	p-value
Intercept	0.961	0.032	< 0.0001
$\mu = 8$	-0.030	0.021	0.1721
$\theta_2 = 30$	-0.038	0.026	0.1601
$\theta_2 = 5$	-0.027	0.026	0.3097
γ_{0cat} = Low	0.106	0.026	0.0004
γ_{0cat} = Medium	0.052	0.026	0.0573
$\gamma_1 = 4$	-0.195	0.021	< 0.0001
$n_{det} = 100$	0.001	0.026	0.9735
$n_{det} = 200$	0.035	0.026	0.1996

31

env_2654_figure1.eps
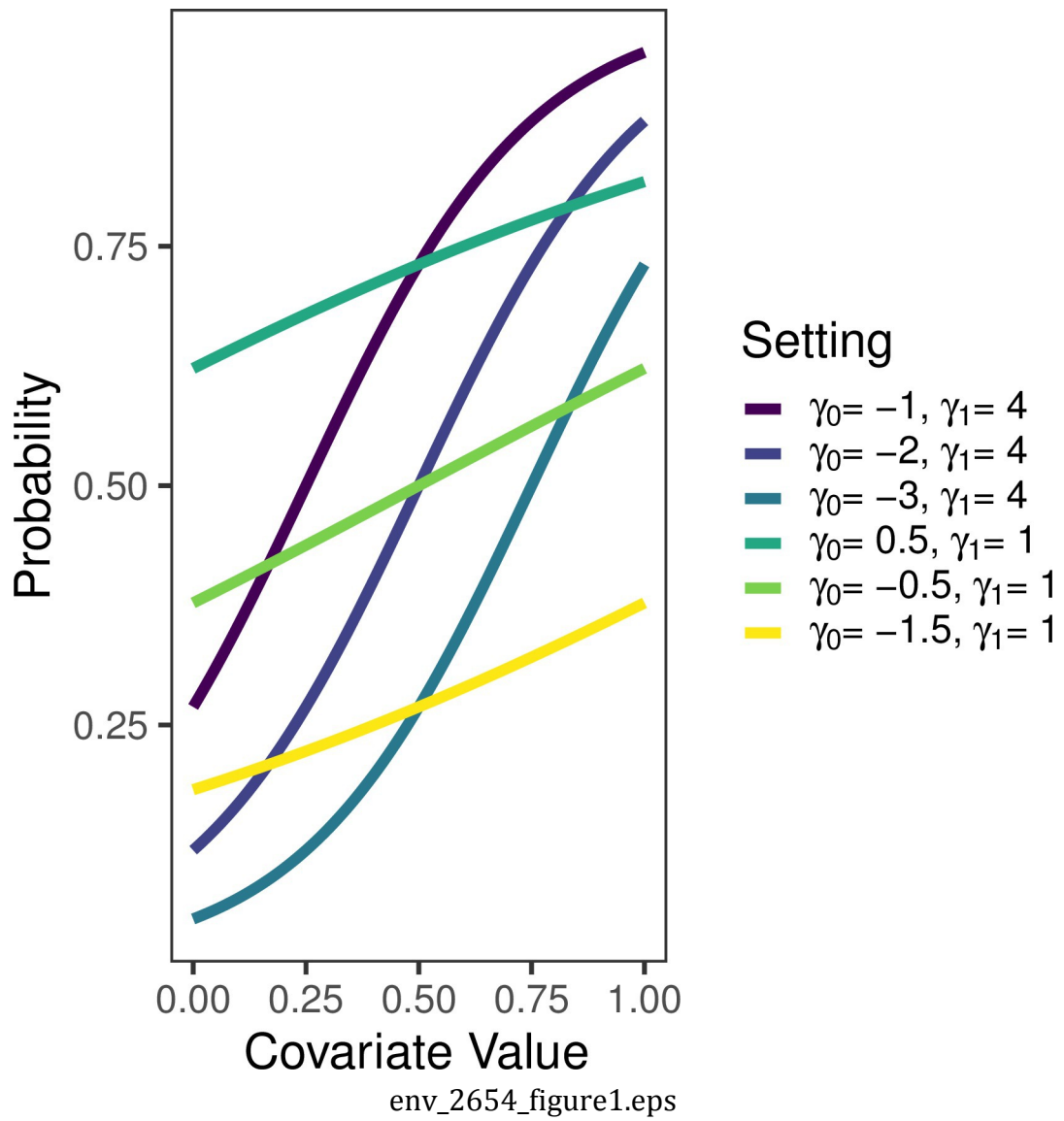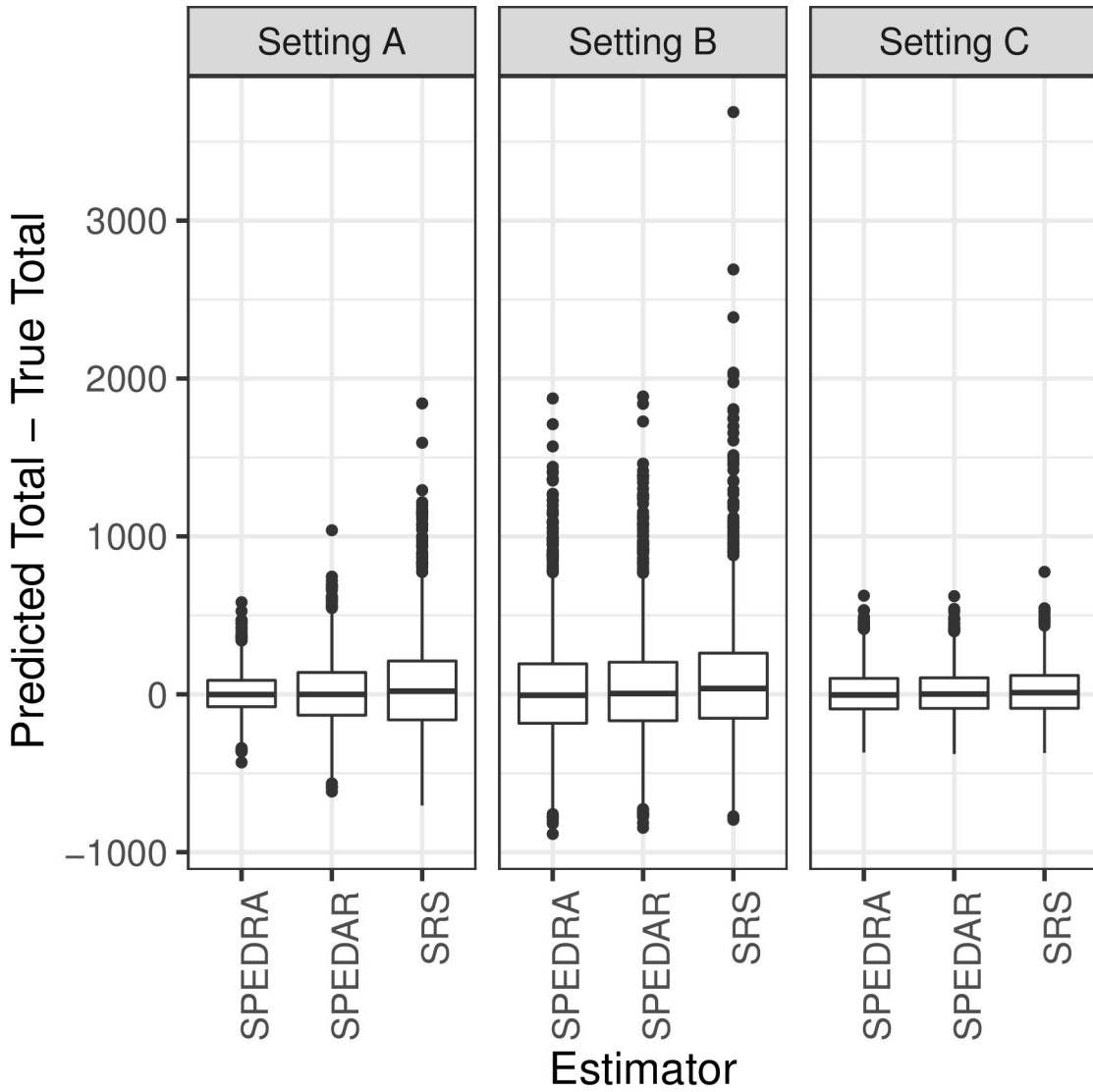
env_2654_figure2.eps

Togiak Map of Stratified Sites

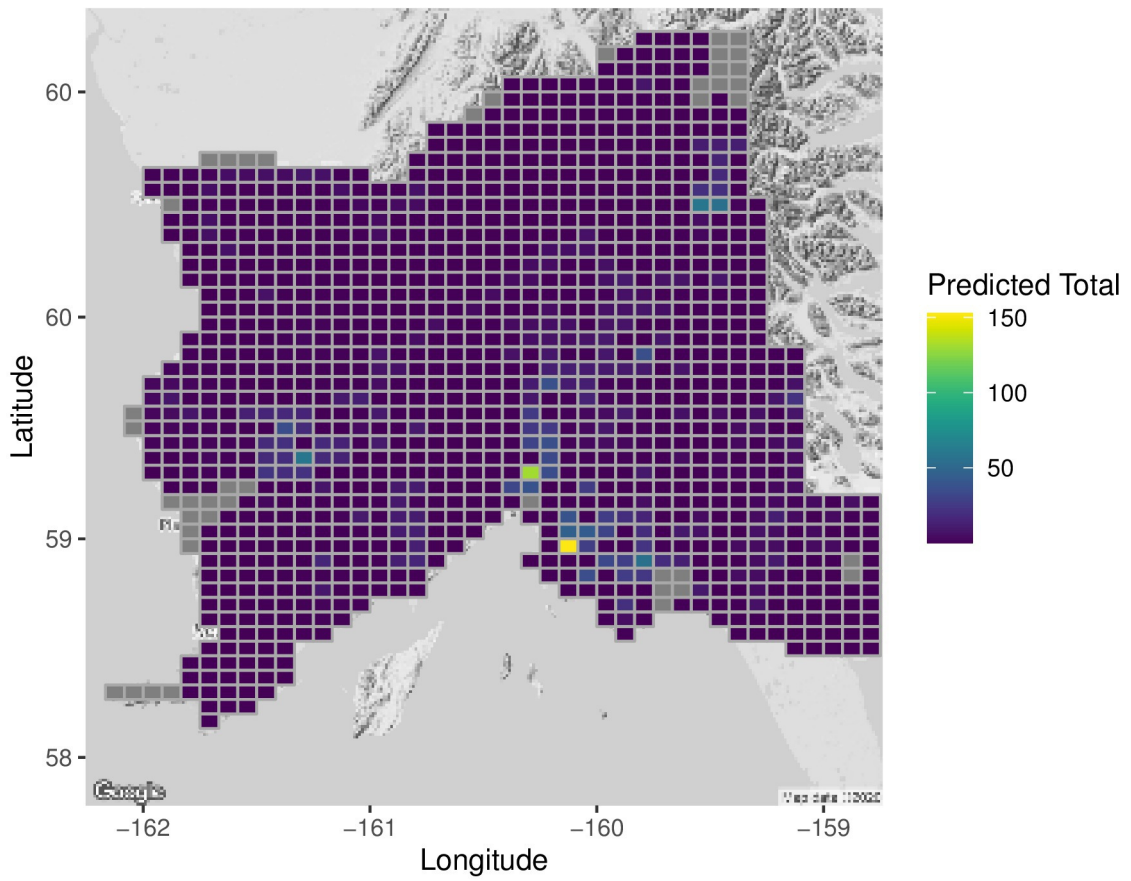env_2654_figure3.eps

High Stratum Semivariogram · Low Stratum Semivariogram
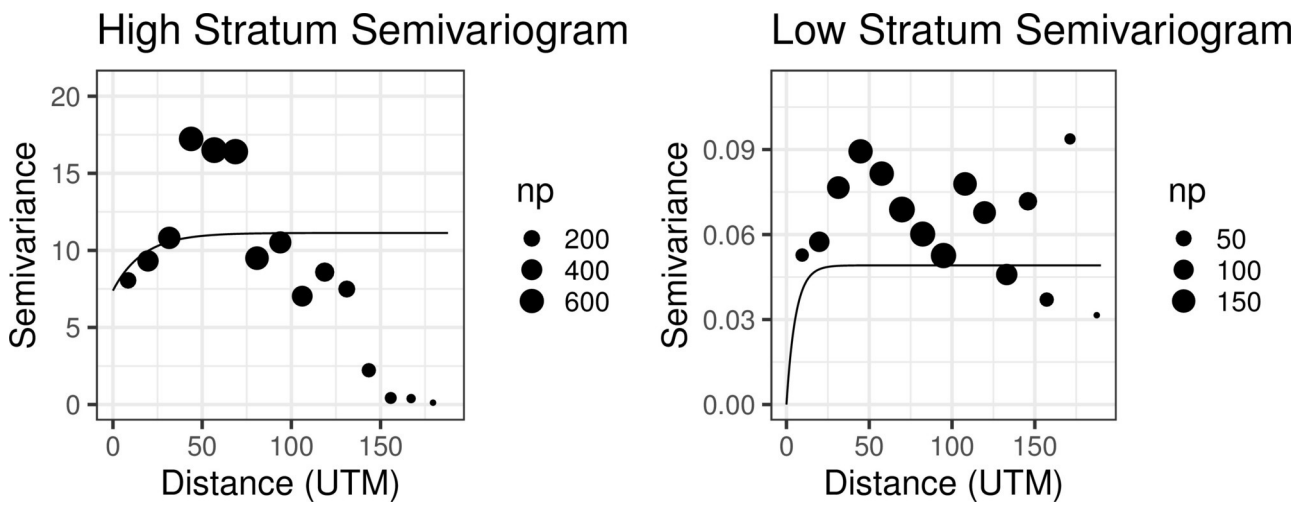
env_2654_figure4.eps

Togiak Map of Kriged Predictions

env_2654_figure5.eps