

¹ ARTICLE TEMPLATE

² **An Application of Spatio-temporal Modeling to Finite Population**
³ **Abundance Prediction**

⁴ Matt Higham^a, Michael Dumelle^b, Carly Hammond^e, Jay Ver Hoef^d, Jeff Wells^c

⁵ ^aSt. Lawrence University Department of Mathematics, Computer Science, and Statistics

⁶ Canton, NY 13617; ^bUnited States Environmental Protection Agency Corvallis, OR 97333;

⁷ ^cAlaska Department of Fish and Game Fairbanks, AK 99701; ^dMarine Mammal Laboratory,

⁸ Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration Seattle,

⁹ Washington 98115; ^eBayer Crop Science Creve Coeur, MO 63141

¹⁰ ARTICLE HISTORY

¹¹ Compiled February 6, 2023

¹² ABSTRACT

¹³ Spatio-temporal models can be used to analyze data collected at various spatial
¹⁴ locations throughout multiple time points. However, even with a finite number of
¹⁵ spatial locations, there may be a lack of resources to sample every spatial location
¹⁶ at every time point. We develop a spatio-temporal finite-population block kriging
¹⁷ (ST-FPBK) method to predict a quantity of interest, such as a mean or total,
¹⁸ across a finite number of spatial locations. This ST-FPBK predictor incorporates
¹⁹ an appropriate variance reduction for sampling from a finite population. Through
²⁰ an application to moose surveys in the east-central region of Alaska, we show that
²¹ the predictor has a substantially smaller standard error compared to a predictor
²² from the purely spatial model that is currently used to analyze moose surveys in
²³ the region. We also show how the model can be used to forecast a prediction for
²⁴ abundance in a time point for which spatial locations have not yet been surveyed. A
²⁵ separate simulation study shows that the spatio-temporal predictor is unbiased and
²⁶ that prediction intervals from the ST-FPBK predictor attain appropriate coverage.
²⁷ For ecological monitoring surveys completed with some regularity through time, use
²⁸ of ST-FPBK could improve precision. We also give an R package that ecologists and
²⁹ resource managers could use to incorporate data from past surveys in predicting a
³⁰ quantity from a current survey.

³¹ KEYWORDS

³² spatial; temporal; kriging; total; resource monitoring

³³ 1. Introduction

³⁴ 1.1. *Background*

³⁵ Spatio-temporal data are indexed by both a spatial index, which we will refer to as
³⁶ a “site,” and by a temporal index, which we will refer to as a “time point.” Com-
³⁷ mon examples of spatio-temporal data include infections from a disease in a coun-
³⁸ try or region collected over a time period (e.g. Martínez-Beneito, López-Quilez, and

CONTACT Matt Higham. Email: mhigham@stlawu.edu, Michael Dumelle. Email: Dumelle.Michael@epa.gov, Carly Hammond. Email: carly.hammond@bayer.com, Jay Ver Hoef. Email: jay.verhoef@noaa.gov, Jeff Wells. Email: jeff.wells@alaska.gov

39 Botella-Rocamora 2008; Sahu and Böhning 2022) or climate variables that are recorded
40 through time at multiple locations (Lemos and Sansó 2009).

41 Models for spatio-temporal data have applications in a wide variety of scientific
42 fields (see Wikle, Zammit-Mangion, and Cressie 2019, for many examples). One such
43 application is ecological monitoring of a particular resource, such as animal or plant
44 abundance, rainfall, concentration of a compound in soil samples, etc.

45 In ecological monitoring, we are often interested in prediction of a total or a mean of
46 a particular variable in a finite region at the most recent time point. Ver Hoef (2008)
47 developed Finite Population Block Kriging (FPBK) to predict a linear function of the
48 realized values of a response variable measured at one particular time point in a finite
49 number of sampling units, incorporating a finite population correction to the variance
50 of the predictor. Typically, the linear function is either a mean or a total of the realized
51 values of the response.

52 **1.2. Motivating Example**

53 To motivate the development of the predictor in Section 2, we consider moose surveys,
54 which are performed annually or every other year in many regions of Alaska and west-
55 ern Canada. The most common goal of these surveys is to predict moose abundance,
56 the total number of moose, in some region to inform harvest regulations (Kellie, Col-
57 son, and Reynolds 2019). Because of time and money constraints, only some spatial
58 indices, or sites, in the region of interest are selected to be in the survey at a particular
59 time point. Biologists fly to these selected sites, count the number of moose, and then
60 use FPKB to find a prediction for the finite abundance for that year. These surveys
61 are historically analyzed with software developed by DeLong (2006), which calculates
62 the “GeoSpatial Population Estimator” (GSPE) for a given survey. The GSPE is an
63 application of the FPKB predictor developed by Ver Hoef (2008).

64 Though many of these surveys are completed regularly, most are analyzed com-
65 pletely independently of surveys from previous years (e.g. Gasaway et al. 1986; Kellie
66 and DeLong 2006; Boertje et al. 2009; Peters et al. 2014). For example, a model for
67 a survey conducted in the year 2019 constructs a prediction for total abundance only
68 from counts on sites that were sampled in that year. However, using counts from pre-
69 vious years in a model that incorporates both spatial and temporal (spatio-temporal)
70 correlation while also using a finite population correction factor based on the propor-
71 tion of sites surveyed in the most recent year could result in a prediction for the realized
72 total that is more precise than predictions from a purely spatial model. Shortly, we
73 describe such a predictor.

74 The rest of this paper is organized as follows. In Section 2, we couple spatio-
75 temporal modeling with finite population prediction to develop the Best-Linear-
76 Unbiased-Predictor (BLUP) and its prediction variance for any linear function of a
77 general response variable, including the total abundance across all sites at a particular
78 time point. We call this predictor the ST-FPKB (spatio-temporal Finite Population
79 Block Kriging) predictor. In Section 3, we apply the ST-FPKB to a moose data set
80 in the east-central region of Alaska. In Section 4, we conduct a simulation study to
81 examine the properties of the ST-FPKB predictor and compare its performance to
82 a predictor from a purely spatial model and a simple random sample design-based
83 estimator. Finally, in Section 5, we offer additional thoughts on the application and
84 simulation, and we give directions for future research.

85 **2. Methods**

86 We now give details on the development of the spatio-temporal model and subsequently
 87 use this model to develop a finite population correction factor to give a Best-Linear-
 88 Unbiased-Predictor (BLUP) and its prediction variance for any linear function of the
 89 response vector.

90 **2.1. Spatio-temporal Model**

91 Let $Y(\mathbf{s}_i, t_j)$, $i = 1, 2, \dots, n_s$ and $j = 1, 2, \dots, n_t$, be a random variable indexed by a
 92 spatial site and a time point, where the vector \mathbf{s}_i contains the coordinates for the i^{th}
 93 spatial site, n_s is the number of unique sites, t_j is the time index for the j^{th} time point,
 94 and n_t is the number of unique time points. If each site is represented at every time
 95 point, a vector of the $Y(\mathbf{s}_i, t_j)$, denoted $\mathbf{y}(\mathbf{s}_i, t_j)$, has length $n_s \cdot n_t \equiv N$. Note that, the
 96 above formulation assumes that each site is observed at each time point. We choose to
 97 make this assumption here because doing so ensures cleaner notation throughout the
 98 model development; however, in subsection 2.2, we no longer assume that the response
 99 is recorded at every site-time point combination. Then, a spatio-temporal model for
 100 $\mathbf{y}(\mathbf{s}_i, t_j)$ is

$$\mathbf{y}(\mathbf{s}_i, t_j) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}(\mathbf{s}_i, t_j), \quad (1)$$

101 where \mathbf{X} is a design matrix for the fixed effects and $\boldsymbol{\beta}$ is a parameter vector of fixed
 102 effects. As in Dumelle et al. (2021), we can decompose the error vector $\boldsymbol{\epsilon}(\mathbf{s}_i, t_j)$ into
 103 spatial, temporal, and spatio-temporal components, each of which will be explained
 104 in detail in the subsequent paragraphs:

$$\boldsymbol{\epsilon}(\mathbf{s}_i, t_j) = \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}. \quad (2)$$

105 In the spatial component of equation 2 ($\mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma}$), the matrix \mathbf{Z}_s is an $N \times n_s$
 106 matrix of 0's and 1's, where the values in a row corresponding to a data point at
 107 site \mathbf{s}_i are 1 in the i^{th} column and 0 in all other columns. $\boldsymbol{\delta}$ is a random vector with
 108 mean $\mathbf{0}$ and covariance $\text{cov}(\boldsymbol{\delta}) = \sigma_\delta^2 \mathbf{R}_s$, where \mathbf{R}_s is an $n_s \times n_s$ spatial correlation
 109 matrix and σ_δ^2 is called the spatial dependent error variance (or spatial partial sill).
 110 The random vector $\boldsymbol{\gamma}$ also has mean $\mathbf{0}$ but has covariance $\text{cov}(\boldsymbol{\gamma}) = \sigma_\gamma^2 \mathbf{I}_s$, where \mathbf{I}_s is
 111 the $n_s \times n_s$ identity matrix and σ_γ^2 is called the spatial independent error variance (or
 112 spatial nugget).

113 In the temporal component of equation 2 ($\mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta}$), \mathbf{Z}_t is an $N \times n_t$ matrix of
 114 0's and 1's, where the values in a row corresponding to a data point at time point t_j
 115 are 1 in the j^{th} column and 0 in all other columns. $\boldsymbol{\tau}$ is a random vector with mean
 116 $\mathbf{0}$ and covariance $\text{cov}(\boldsymbol{\tau}) = \sigma_\tau^2 \mathbf{R}_t$, where \mathbf{R}_t is an $n_t \times n_t$ temporal correlation matrix
 117 and σ_τ^2 is called the temporal dependent error variance (or temporal partial sill). $\boldsymbol{\eta}$ is
 118 also a random vector with mean $\mathbf{0}$ but has covariance $\text{cov}(\boldsymbol{\eta}) = \sigma_\eta^2 \mathbf{I}_t$, where \mathbf{I}_t is the
 119 $n_t \times n_t$ identity matrix and σ_η^2 is called the temporal independent error variance (or
 120 temporal nugget).

121 In the spatio-temporal component of equation 2 ($\boldsymbol{\omega} + \boldsymbol{\nu}$), $\boldsymbol{\omega}$ is a random vector
 122 with mean $\mathbf{0}$ and covariance $\text{cov}(\boldsymbol{\omega}) = \sigma_\omega^2 \mathbf{R}_{st}$, where \mathbf{R}_{st} is an $N \times N$ spatio-temporal
 123 correlation matrix and σ_ω^2 is sometimes called the spatio-temporal dependent error
 124 variance (or spatio-temporal partial sill). $\boldsymbol{\nu}$ is also a random vector with mean $\mathbf{0}$ but

125 has covariance $\text{cov}(\boldsymbol{\nu}) = \sigma_\nu^2 \mathbf{I}_{st}$, where \mathbf{I}_{st} is the $N \times N$ identity matrix and σ_ν^2 is
 126 sometimes called the spatio-temporal independent error variance (or spatio-temporal
 127 nugget).

Though there are a few types of models for the errors that can be built from 2 by setting certain error variances to 0 (e.g. a sum-with-error model sets $\sigma_\omega^2 = 0$) and/or by allowing \mathbf{R}_{st} to take certain forms, we focus only on the product-sum model (De Cesare, Myers, and Posa 2001; De Iaco, Myers, and Posa 2001). In a common formulation of the product-sum model, \mathbf{R}_{st} is

$$\mathbf{R}_{st} \equiv \mathbf{Z}_s \mathbf{R}_s \mathbf{Z}'_s \odot \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}'_t,$$

128 where \odot is the Hadamard product operator. Note that, in order to save on the number
 129 of parameters, we will assume that the \mathbf{R}_s and \mathbf{R}_t that form \mathbf{R}_{st} are the same as the
 130 \mathbf{R}_s and \mathbf{R}_t associated with $\boldsymbol{\delta}$ and $\boldsymbol{\tau}$, respectively, although this is not necessary in
 131 general. \mathbf{R}_s can be parameterized in different ways, but one common assumption is
 132 to assume the covariance function generating \mathbf{R}_s is second-order stationary (ie. the
 133 covariance between two data points is a function only of the separation vector between
 134 two sites) and isotropic (ie. the covariance is a function of the distance only and does
 135 not depend on the direction of the separation vector). For example, the exponential
 136 covariance function is defined as follows. For observations at sites i and i' at $h_{ii'}$
 137 distance apart, row i and column i' of \mathbf{R}_s is equal to

$$\exp(-h_{ii'}/\phi), \quad (3)$$

138 where $\exp(x)$ is equivalent to e^x and ϕ is a spatial range parameter controlling the
 139 decay rate of the covariance as distance between two sites increases (Cressie 2015).

140 Similarly, one common assumption when parameterizing \mathbf{R}_t is to assume the covariance
 141 function generating \mathbf{R}_t is second-order stationary (ie. the covariance is a function
 142 only of the temporal distance). For example, the exponential covariance function is
 143 defined as follows. For observations at time points j and j' at $m_{jj'}$ units apart, row j
 144 and column j' of \mathbf{R}_t is equal to

$$\exp(-m_{jj'}/\rho), \quad (4)$$

145 where ρ is a temporal range parameter controlling the decay rate of the covariance as
 146 time units between two data points increases. Note that the exponential form of \mathbf{R}_t
 147 is equivalent to an AR(1) time series model if the time points are equally spaced and
 148 the correlation parameter in the AR(1) series is greater than zero (Schabenberger and
 149 Gotway 2017).

150 The product-sum model for $\mathbf{y}(\mathbf{s}_i, t_j)$ is then

$$\mathbf{y}(\mathbf{s}_i, t_j) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s \boldsymbol{\delta} + \mathbf{Z}_s \boldsymbol{\gamma} + \mathbf{Z}_t \boldsymbol{\tau} + \mathbf{Z}_t \boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}, \quad (5)$$

151 where $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$, $\boldsymbol{\eta}$, $\boldsymbol{\omega}$, and $\boldsymbol{\nu}$ are mutually independent, $\mathbf{y}(\mathbf{s}_i, t_j)$ has mean $\mathbf{X}\boldsymbol{\beta}$, and
 152 $\mathbf{y}(\mathbf{s}_i, t_j)$ has covariance

$$\text{var}(\mathbf{y}) \equiv \boldsymbol{\Sigma} = \sigma_\delta^2 \mathbf{Z}_s \mathbf{R}_s \mathbf{Z}'_s + \sigma_\gamma^2 \mathbf{Z}_s \mathbf{I}_s \mathbf{Z}'_s + \sigma_\tau^2 \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}'_t + \sigma_\eta^2 \mathbf{Z}_t \mathbf{I}_t \mathbf{Z}'_t + \sigma_\omega^2 \mathbf{R}_{st} + \sigma_\nu^2 \mathbf{I}_{st}. \quad (6)$$

153 There are a few reasons for why we choose to solely focus on the product-sum model.
 154 First, as long as \mathbf{R}_s and \mathbf{R}_t are positive definite and either $\sigma_\omega^2 > 0$ or $\sigma_\nu^2 > 0$, then

155 the covariance matrix in equation 6 is also positive definite (De Cesare, Myers, and
 156 Posa 2001; De Iaco, Myers, and Posa 2001). Also, the product-sum model is flexible in
 157 its ability to model many kinds of spatial and temporal correlation (De Iaco, Palma,
 158 and Posa 2015; Dumelle et al. 2021). Xu and Shu (2015) claim that the product-sum
 159 model is the most widely used spatio-temporal model used in practical applications.

160 **2.2. Finite Population Block Kriging**

161 The model that we developed in the previous section in equation 5 is for the N -length
 162 vector \mathbf{y} . However, often we do not have the resources to sample or observe every spatial
 163 site during every time point. Therefore, we may have an interest in prediction of the
 164 response values on sites that were not observed, particularly sites in the most recent
 165 time point. Throughout this section, let the subscript o denote data points that were
 166 “observed” or sampled, the subscript u denote data points that were “unobserved” or
 167 not sampled, and the subscript a denote “all” data points. Then, we can re-order the
 168 response vector \mathbf{y} so that

$$\mathbf{y} \equiv \mathbf{y}_a = [\mathbf{y}'_u, \mathbf{y}'_o]'. \quad (7)$$

169 Our primary goal is to use the model developed for \mathbf{y}_a in equation 5 to find optimal
 170 weights \mathbf{q}' to apply to the observed realizations of \mathbf{y}_o such that $\mathbf{q}'\mathbf{y}_o$ is the Best Linear
 171 Unbiased Predictor (BLUP) for $\mathbf{b}'_a\mathbf{y}_a$, a linear function of \mathbf{y}_a . The N -length vector \mathbf{b}'_a
 172 is, for example, a vector of 1's, in which case we would be predicting the total response
 173 across all sites and all time points.

174 Unbiasedness implies that $E(\mathbf{q}'\mathbf{y}_o) = E(\mathbf{b}'_a\mathbf{y}_a)$ for all β . So, denoting \mathbf{X}_o as the
 175 design matrix for the observed data points and \mathbf{X}_a as the design matrix for all data
 176 points, $\mathbf{q}'\mathbf{X}_o\beta = \mathbf{b}'_a\mathbf{X}_a\beta$ for every β , implying that $\mathbf{q}'\mathbf{X}_o = \mathbf{b}'_a\mathbf{X}_a$. Kriging weights
 177 are then found by finding λ_o , an $n_o \times 1$ column vector, where n_o is the number of
 178 observed data points, such that

$$E\{(\mathbf{q}'\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)^2\} - E\{(\lambda'_o\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)^2\} \quad (8)$$

179 is greater than 0 for all \mathbf{q}' . The prediction equations are

$$\begin{pmatrix} \Sigma_{o,o} & \mathbf{X}_o \\ \mathbf{X}'_o & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ m \end{pmatrix} = \begin{pmatrix} \Sigma_{o,o} & \Sigma_{o,u} \\ \mathbf{X}'_o & \mathbf{X}'_u \end{pmatrix} \begin{pmatrix} \mathbf{b}_o \\ \mathbf{b}_u \end{pmatrix}, \quad (9)$$

180 where again the subscripts o and u denote observed and unobserved data points. For
 181 example, $\Sigma_{o,o}$ denotes the $n_o \times n_o$ submatrix of Σ (from equation 6) corresponding only
 182 to rows and columns of observed data points and $\Sigma_{u,o}$ denotes the $(N-n_o) \times n_o$ subma-
 183 trix of Σ corresponding to rows of data points that were not observed and columns of
 184 data points that were observed. Solving the prediction equations, the optimal predic-
 185 tion weights that are both unbiased and have the smallest possible prediction variance
 186 compared to any other linear predictor are

$$\lambda'_o = \mathbf{b}'_o + \mathbf{b}'_u [(\Sigma_{u,o}\Sigma_{o,o}^{-1}) - (\Sigma_{u,o}\Sigma_{o,o}^{-1})\mathbf{X}_o\mathbf{W}_o^{-1}\mathbf{X}'_o\Sigma_{o,o}^{-1} + \mathbf{X}'_u\mathbf{W}_o^{-1}\mathbf{X}_o\Sigma_{o,o}^{-1}], \quad (10)$$

187 where $\mathbf{W}_o = \mathbf{X}'_o \Sigma_{o,o}^{-1} \mathbf{X}_o$. The BLUP for $\mathbf{b}'_a \mathbf{y}_a$ is then

$$\widehat{\mathbf{b}'_a \mathbf{y}_a} = \boldsymbol{\lambda}'_o \mathbf{y}_o, \quad (11)$$

which is equivalent to

$$\mathbf{b}'_o \mathbf{y}_o + \mathbf{b}'_u \hat{\mathbf{y}}_u,$$

188 where $\hat{\mathbf{y}}_u = \Sigma_{o,s} \Sigma_{o,o}^{-1} (\mathbf{y}_o - \hat{\mu}_o) + \hat{\mu}_u$ with $\hat{\mu}_o = \mathbf{X}_o \hat{\beta}$ and $\hat{\mu}_u = \mathbf{X}_u \hat{\beta}$. $\hat{\beta}$ is the generalized
189 least squares estimator $(\mathbf{X}'_o \Sigma_{o,o}^{-1} \mathbf{X}_o)^{-1} \mathbf{X}'_o \Sigma_{o,o}^{-1} \mathbf{y}_o$. We can see then that the predictor
190 multiplies the observed data \mathbf{y}_o with relevant weights from the \mathbf{b}_o vector, and then
191 adds in the kriged predictions $\hat{\mathbf{y}}_u$ multiplied with relevant weights from the \mathbf{b}_u vector.

192 The prediction variance of the predictor in equation 11 is

$$E((\boldsymbol{\lambda}'_o \mathbf{y}_o - \mathbf{b}'_a \mathbf{y}_a)(\boldsymbol{\lambda}'_o \mathbf{y}_o - \mathbf{b}'_a \mathbf{y}_a)) = \boldsymbol{\lambda}'_o \Sigma_{o,o} \boldsymbol{\lambda}_o - 2\mathbf{b}'_a \Sigma_{a,o} \boldsymbol{\lambda}_o + \mathbf{b}'_a \Sigma_{a,a} \mathbf{b}_a. \quad (12)$$

193 We call the predictor in equation 11 with Σ in equation 6 the ST-FPBK predictor.

194 A common predictor of interest is the total abundance in the most current time point
195 of the survey. In this scenario, \mathbf{b}_a is a vector of 1's and 0's, where the k^{th} element of
196 \mathbf{b}_a is equal to 1 if the k^{th} element of \mathbf{y}_a is from the most recent time point of the
197 survey and the k^{th} element of \mathbf{b}_a is equal to 0 otherwise. If we order \mathbf{y}_a by (1) the
198 unobserved data points from past surveys, (2) the unobserved data points from the
199 current survey, (3) the observed data points from past surveys, and (4) the observed
200 data points from the current survey, then

$$\mathbf{b}_a = [\mathbf{b}'_{up}, \mathbf{b}'_{uc}, \mathbf{b}'_{op}, \mathbf{b}'_{oc}]' = [\mathbf{0}', \mathbf{1}', \mathbf{0}', \mathbf{1}]', \quad (13)$$

201 where the subscripts *up*, *uc*, *op*, and *oc* denote unobserved sites in past surveys, un-
202 observed sites in the current survey, observed sites in past surveys, and observed sites
203 in the current survey, respectively.

204 2.3. Estimation

205 In practical applications, the covariance matrix Σ in equation 6 that is partitioned
206 into the various sub-matrices in equations 11 and 12 needs to be estimated from the
207 observed data \mathbf{y}_o . The spatio-temporal model in equation 5 does not have any distri-
208 butional assumptions: we only need to specify the mean and variance of \mathbf{y}_o . Restricted
209 Maximum Likelihood (REML) can be used to estimate the covariance parameters in
210 Σ , which we will refer to as $\boldsymbol{\theta} \equiv [\sigma_\delta^2, \sigma_\gamma^2, \phi, \sigma_\tau^2, \sigma_\eta^2, \rho, \sigma_\omega^2, \sigma_\nu^2]'$ (Patterson and Thomp-
211 son 1971; Harville 1977). Even if \mathbf{y}_a is not multivariate normal, the REML estimator
212 for the parameter vector $\boldsymbol{\theta}$ is still unbiased (Heyde 1994; Cressie and Lahiri 1993).

213 However, REML estimation can be computationally burdensome, particularly for
214 large spatio-temporal data sets with many observed sites and time points. Therefore,
215 we use developments from Dumelle et al. (2021) in the application, the simulations
216 described in the next section, and the accompanying R package to speed up estimation
217 of $\boldsymbol{\theta}$.

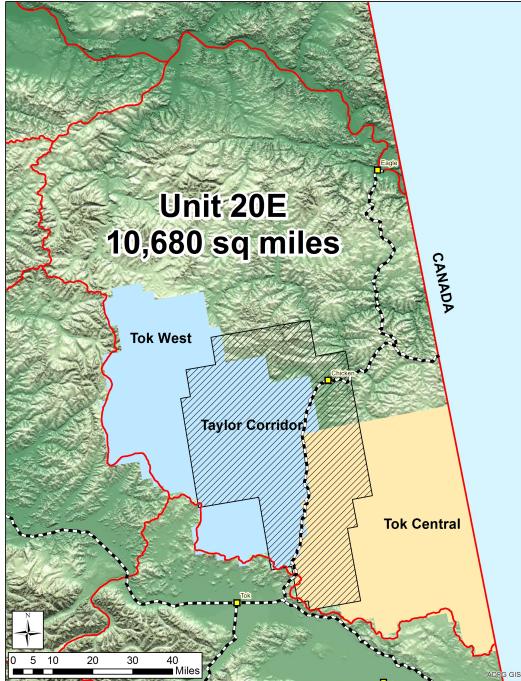


Figure 1. A map of the Taylor Corridor in the east-central region of Alaska.

218 3. Application

219 We now apply the ST-FPBK predictor to a moose data set described below. Moose
 220 surveys throughout Alaska and Canada are often conducted regularly, making them
 221 good candidates for incorporating temporal correlation.

222 3.1. Data Description

223 The Taylor Corridor in the east-central region of Alaska is a popular habitat for moose
 224 and other wildlife. Abundance surveys for moose are performed in the Taylor Corridor
 225 of the Tok region of Alaska annually (Figure 1) so that biologists have an idea about
 226 the abundance of moose each year. In particular, surveys were conducted from 2014
 227 through 2020 in every year except 2016, during which there was not sufficient snow
 228 cover to perform a survey. The spatial sampling frame for our study area consists of
 229 381 sites. There are a total of 7 unique time points represented in the data, including
 230 the missing year of 2016. Therefore, N is 2667.

231 In each year of the survey, a team of biologists selected some of the 381 sites to
 232 survey. The number of sites that were selected varies from a low of 76 in the year
 233 2019 to a high of 90 in the year 2020. Throughout the 7 unique years, some sites were
 234 sampled as many as five different times while others were never sampled at all (Figure
 235 2). Figure 2 and all remaining figure graphics are constructed with the `ggplot2` R
 236 package (Wickham 2016). The number of units sampled throughout all survey years,
 237 n , was 487 units.

238 Before the survey begins in each year, biologists stratified the sites into a “High”
 239 stratum and a “Low” stratum composed. The goal of the following analysis is to predict
 240 the total abundance of moose across all sites in the year 2020, the most recent year of

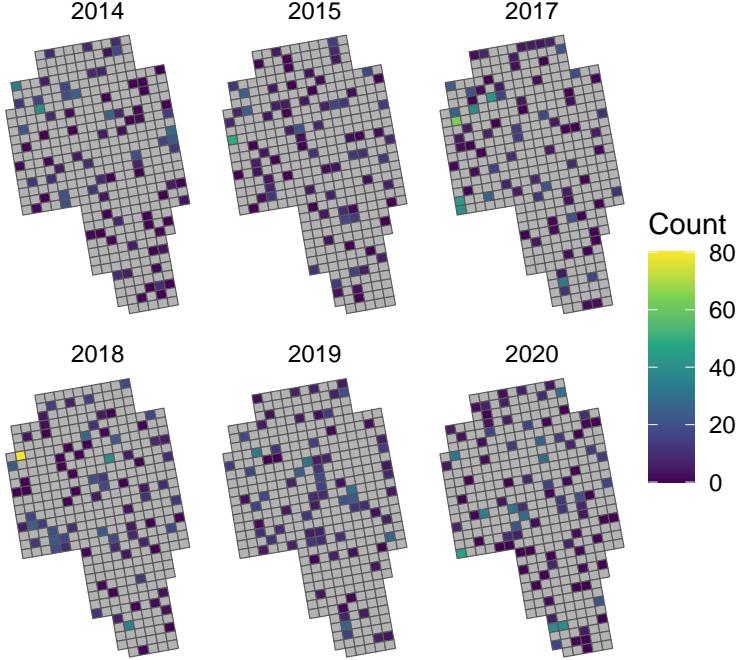


Figure 2. Layout of the spatial sites used to survey moose in the Taylor corridor in eastern-central Alaska, coloured by moose count. Sites coloured grey were not sampled in that year. The year 2016 is excluded because no survey was performed in that year.

the survey, using stratum as a covariate in the spatio-temporal model.

3.2. Model Fitting

We fit the product-sum covariance model defined in equation 5 using REML with stratum as a covariate in the design matrix, an exponential spatial correlation structure defined in equation 3, and an exponential temporal correlation structure defined in equation 4. Table 1 gives the estimated parameters from the model fit.

Table 1. Estimated covariance parameters in the model. $\hat{\sigma}_\delta^2$, $\hat{\sigma}_\gamma^2$, and $\hat{\phi}$ are the spatial dependent error variance, independent error variance, and range parameters, respectively. $\hat{\sigma}_\tau^2$, $\hat{\sigma}_\eta^2$, and $\hat{\rho}$ are the temporal dependent error variance, independent error variance, and range parameters, respectively. $\hat{\sigma}_\omega^2$ and $\hat{\sigma}_\nu^2$ are the spatio-temporal dependent error variance and spatio-temporal independent error variance.

Spatial			Temporal			Spatio-temporal	
$\hat{\sigma}_\delta^2$	$\hat{\sigma}_\gamma^2$	$\hat{\phi}$	$\hat{\sigma}_\tau^2$	$\hat{\sigma}_\eta^2$	$\hat{\rho}$	$\hat{\sigma}_\omega^2$	$\hat{\sigma}_\nu^2$
16.37	7.78	4.51	0.29	0	3.68	25.53	36.47

To help interpret what some of these fitted covariance parameter estimates mean, we can construct a fitted covariance plot (Figure 3). As the spatial distance between two sites increases (dark colour to light colour), the covariance of two errors decreases to 0, with the $\hat{\phi}$ parameter estimate controlling the rate of decay. In fact, the model estimates the covariance to be nearly 0 when two sites are 20 or more kilometers apart, no matter what the temporal distance is. The covariance between two errors that are

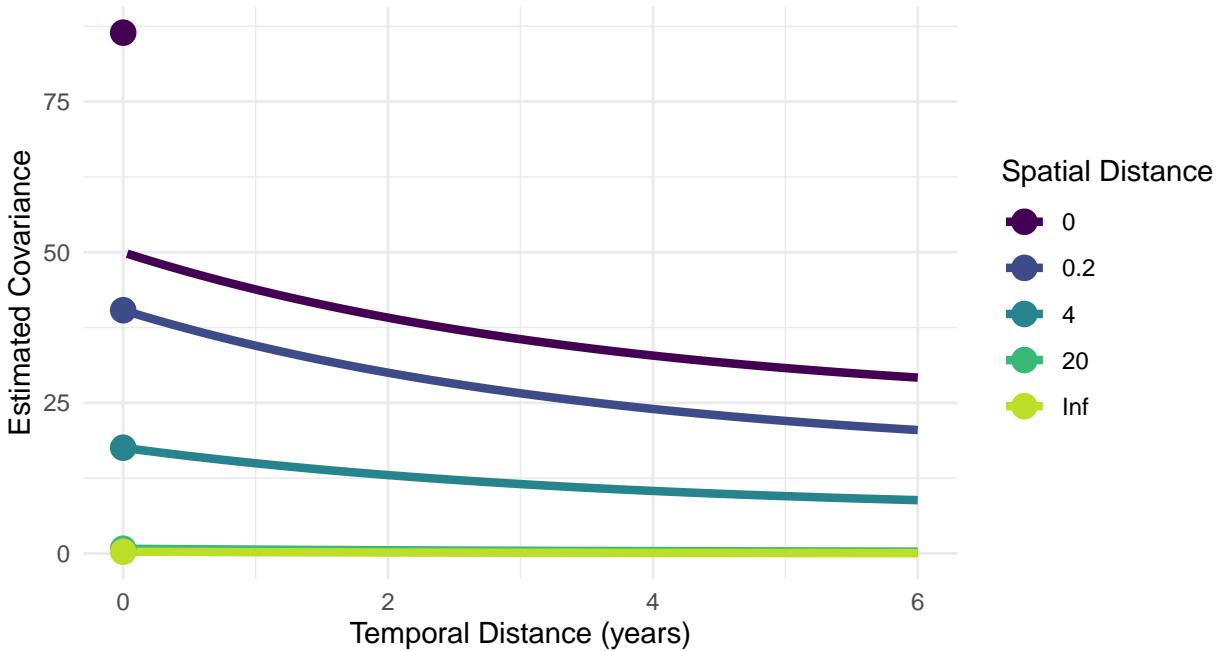


Figure 3. Estimated covariance of the errors from the estimated parameters in a spatio-temporal product-sum model. Distance between two sites is calculated from the site centroids; the centroids of two sites directly adjacent to one another are about 4 spatial kilometers apart.

253 six years apart is still estimated to be positive if the two errors come from the same
 254 site or from adjacent sites.

255 The estimated vector of fixed effects, using “High” as the reference group, is $\hat{\beta} =$
 256 $(9.62, -4.55)$. Therefore, the overall mean for sites in the “High” stratum is estimated
 257 to be 9.62 moose while the overall mean for sites in the “Low” stratum is estimated
 258 to be 5.07 moose.

259 3.3. Prediction

260 We now use the fitted spatio-temporal model with the BLUP from equation 11 and
 261 weights given in equation 13 to predict the total abundance across all sites in the
 262 year 2020, the most recent year of the survey. Plugging in estimates of the covariance
 263 parameters into equations 11 and 12 and letting elements of \mathbf{b}_a be equal to 1 for
 264 data points in 2020 and equal to 0 otherwise, we obtain a prediction of 3001 moose
 265 and a standard error (the square root of the prediction variance) of 217 moose. A 90%
 266 normal-based prediction interval for the total abundance in 2020 is (2644, 3357) moose.
 267 Note that, though the response in this example is a count, a normal-based prediction
 268 interval for the total is still appropriate through an application of the central limit
 269 theorem for dependent data (Smith 1980). Sitewise predictions for sites in 2020 are
 270 given in the map in Figure 4.

271 For comparison, we use the spatial `sptotal` package (Higham et al. 2021) to compute
 272 the spatial FPBK prediction (Ver Hoef 2008) for the total abundance of moose in the
 273 year 2020 with stratum as a covariate. The spatial FPBK predictor is what is currently
 274 implemented in the widely used GSPE software for moose surveys (DeLong 2006).

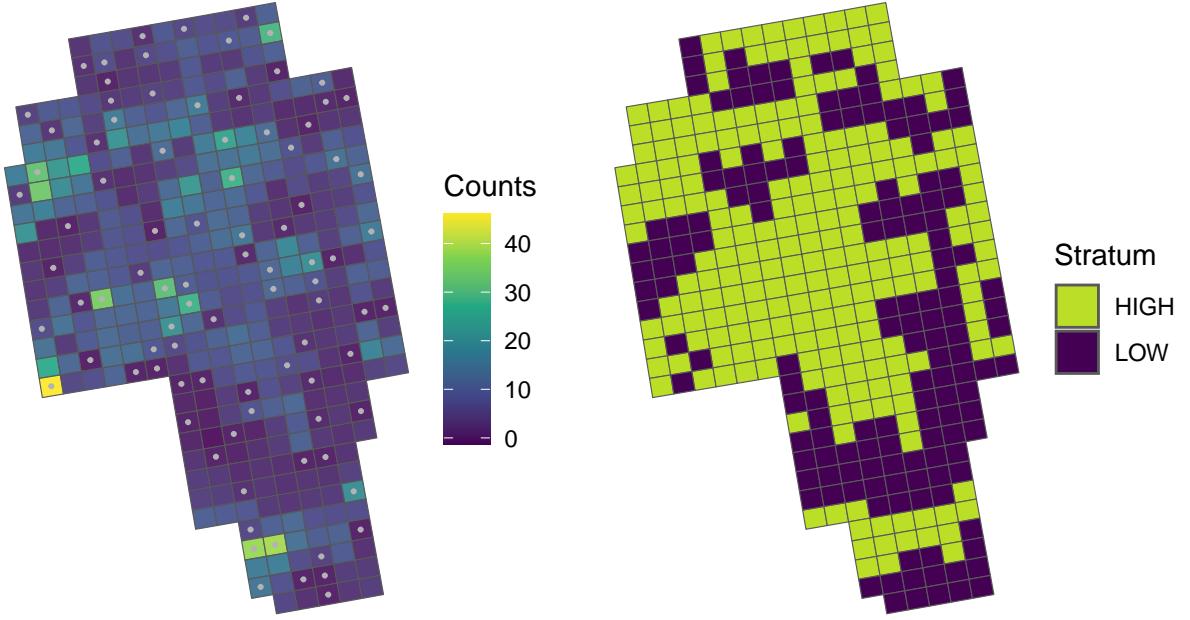


Figure 4. (Left). A map of the predictions for the sites in the year 2020 from the spatio-temporal model. A site with a grey dot in the center means that the site was sampled in 2020. (Right). A map of the stratification of sites in 2020. NOTE: need to update these to be proper sub-figures if we keep the stratification figure.

We also use the stratified random sampling design-based estimator

$$\sum_{i=1}^2 N_i \cdot \bar{y}_i$$

where \bar{y}_i is the sample mean for the observed data in 2020 in the i^{th} stratum and N_i is the total number of sites in 2020 in the i^{th} stratum. The stratified random sampling design-based estimator has a variance for the total abundance of

$$\sum_{i=1}^2 N_i^2 \cdot \left(1 - \frac{n_i}{N_i}\right) \cdot \frac{s_i^2}{n_i},$$

where s_i^2 is the sample variance of the observed data points in 2020 in the i^{th} stratum and n_i is the number of observed data points in 2020 in the i^{th} stratum. Both the purely spatial model fit with `sptotal` and the stratified random sampling design-based estimator use data only from 2020.

For the purely spatial model, the prediction for the total number of moose in 2020 in the region is 2870 moose with a standard error of 319 moose. For the stratified random sampling design-based estimator, the estimated total number of moose in 2020 in the region is 2853 moose with a standard error of 371 moose. While the predictions for the total moose abundance are similar across the three methods, we see that the spatio-temporal model is most efficient ($SE = 217$ moose compared to 319 moose for the purely spatial model that ignores previous surveys and 371 moose for the stratified random sampling design-based estimator that ignores both previous surveys and spatial correlation in the current survey).

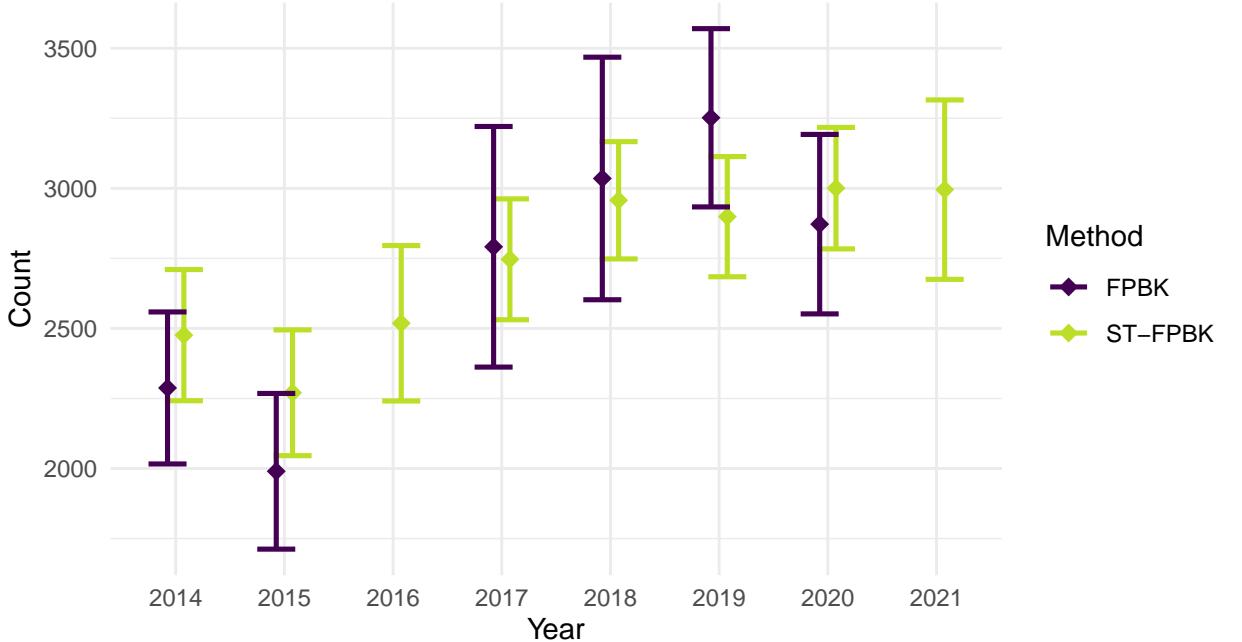


Figure 5. Moose abundance predictions for the Taylor Corridor from 2014 through 2021 with the spatial FPBK predictor and the ST-FPBK predictor. Predictions are given with a diamond symbol; the bars surrounding each prediction are standard error bars. Because surveys were not conducted in 2016 and 2021, there is no spatial FPBK predictor and the standard errors for the ST-FPBK predictor for those years is larger than the standard errors in the other years. The stratification scheme used for 2016 and 2021 in the ST-FPBK analysis was the same scheme used in 2015 and 2020, respectively.

288 In addition to making a prediction for the abundance in the most recent survey,
 289 we can also use the spatio-temporal model to backcast predictions for the abundance
 290 in past survey years, interpolate predictions for years during which a survey was not
 291 completed, and forecast predictions for future years. For example, in the Taylor Cor-
 292 ridor surveys, there was no survey conducted in the year 2016 because of insufficient
 293 snow cover. Leveraging the temporal structure of the st-FPBK predictor, we can still
 294 construct a prediction and corresponding standard error though, as expected, this
 295 standard error is larger than the standard errors of years where a survey was com-
 296 pleted (Figure 5). Also, in Figure 5, we see a forecasted prediction and correspond-
 297 ing standard error for the abundance in 2021. Again, the standard error associated with
 298 the forecasted prediction is larger than the standard errors for the years with surveys.

299 **4. Simulation**

300 **4.1. Description**

301 To evaluate performance of the ST-FPBK predictor, we conduct a simulation study.
 302 We simulate a response vector \mathbf{y} of length $N = 1000$ on a 10×10 grid of 100 spatial
 303 sites on the unit square ($[0, 1] \times [0, 1]$) and 10 equally-spaced time points in the interval
 304 $[0, 1]$, so that each spatial site has a response value at each time point. \mathbf{y} is multivariate
 305 normal with mean $\mathbf{0}$ and product-sum covariance matrix Σ defined in equation 6 with
 306 the covariance parameters given in Table 2.

Table 2. Covariance parameters used to simulate data. σ_δ^2 , σ_γ^2 , and ϕ are the spatial dependent error variance, independent error variance, and range parameters, respectively. σ_τ^2 , σ_η^2 , and ρ are the temporal dependent error variance, independent error variance, and range parameters, respectively. σ_ω^2 and σ_ν^2 are the spatio-temporal dependent error variance and spatio-temporal independent error variance. Note that both ϕ (and ρ) appear in \mathbf{R}_{st} ; therefore, their values can change the underlying covariance even when σ_δ^2 (and σ_τ^2) are equal to 0.

scenario	Spatial			Temporal			Spatio-temporal	
	σ_δ^2	σ_γ^2	ϕ	σ_τ^2	σ_η^2	ρ	σ_ω^2	σ_ν^2
all-dev	0.5	0.17	0.47	0.5	0.17	0.33	0.50	0.17
t-iev	0	0	0.47	0	1.50	0	0.25	0.25
spt-iev	0	0	0	0	0	0	0	2.00

307 The three scenarios in the table correspond to (1) **all-dev**: a scenario where a
 308 substantial proportion of the overall variance comes from the spatial, temporal, and
 309 spatio-temporal dependent error variance parameters σ_δ^2 , σ_τ^2 , and σ_ω^2 ; (2) **t-iev**: a sce-
 310 nario where there the overall variance is dominated by the temporal independent error
 311 variance parameter, σ_η^2 ; and (3) **spt-iev**: a scenario where all of the variability comes
 312 from σ_ν^2 so that errors are independent regardless of spatial and time indices. In all
 313 scenarios, summing all six variance parameters gives a total variance equal to two.

314 Both \mathbf{R}_s and \mathbf{R}_t are generated from the exponential correlation function with ϕ
 315 and ρ as the range parameters in equations 3 and 4. The values 0.471 and 0.3333 are
 316 chosen for ϕ and ρ , respectively, so that the effective ranges, 3ϕ and 3ρ , are equal
 317 to the maximum distance between two data points in space ($\sqrt{2} = 1.414$) and the
 318 maximum distance between two data points in time (1). A value of 0 for ϕ (or ρ) sets
 319 the \mathbf{R}_s (or the \mathbf{R}_t) matrix to the identity matrix. Figure 6 shows the model covariance
 320 of the errors used to generate data for the “all-dev” scenario.

321 Each of these three scenarios is replicated for two different sample sizes: $n = 250$
 322 and $n = 500$. A simple random sample is chosen from the 1000 total data points.

323 Finally, the simulation experiment is repeated for a skewed response variable. To
 324 create the skewed response variable, a normally-distributed response is simulated ac-
 325 cording to the parameters given in Table 2, except that each of the variance parameters
 326 (not including ϕ and ρ) is divided by 2.89 so that the total variance is equal to 0.6931.
 327 This variable is then exponentiated so that the total variance after exponentiation is
 328 equal to 2. Note that, not only does exponentiation result in a right-skewed response
 329 variable, but exponentiating also allows for an assessment of how the ST-FPBK pre-
 330 dictor performs when the covariance is mis-specified, as the resulting response variable
 331 is now simulated with an intractable covariance function that is not used in the model
 332 fitting.

333 Therefore, the simulation study has 12 total settings coming from a $3 \times 2 \times 2$ (scenario
 334 \times sample size \times distribution shape) factorial design. For each setting, we simulate 1000
 335 realizations of the response vector \mathbf{y} . For each realization, we use three methods to
 336 predict the total response for the “most current” time point, which is when the time
 337 index is equal to 1 on the interval $[0, 1]$). We will henceforth call this “total response
 338 for the most current time point quantity” the “current total.”

339 The first method uses the ST-FPBK predictor in equation 11 with the spatio-
 340 temporal model covariance in equation 6. REML estimation with the observed data
 341 \mathbf{y}_o is used to obtain estimates for the covariance parameter vector $\boldsymbol{\theta}$. The second
 342 method is the FPBK spatial model fit with the **sptotal** R package (Higham et al.
 343 2021) that only uses data from the most current time point.

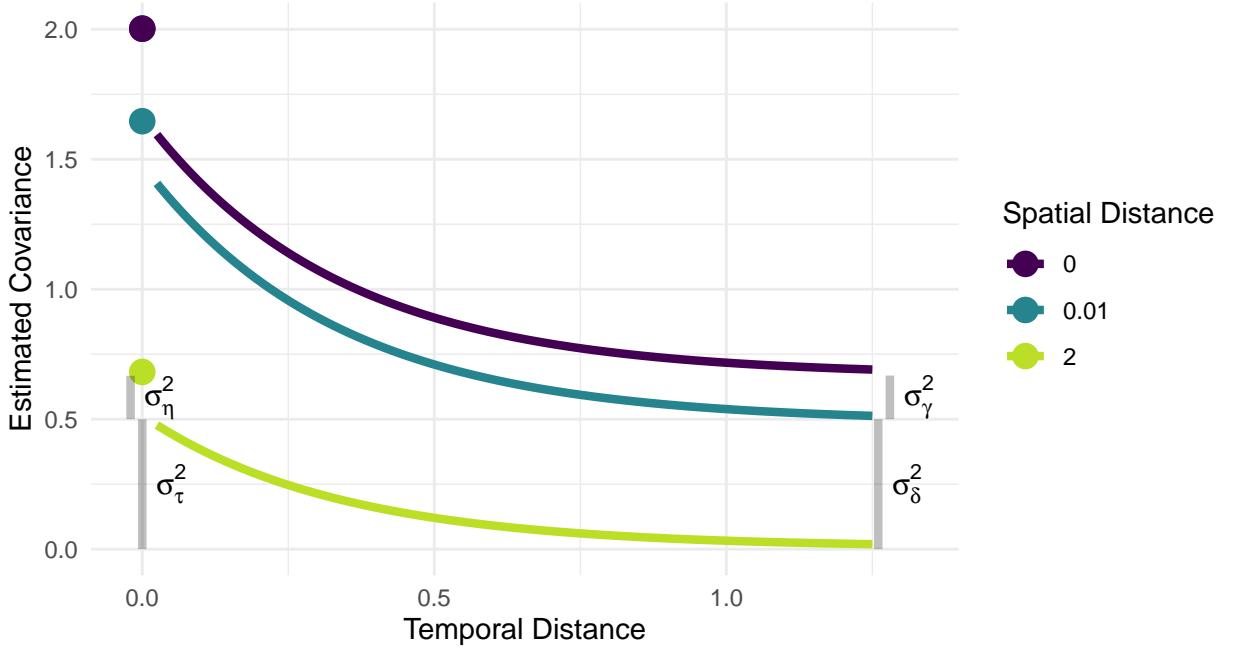


Figure 6. The model covariance used in the simulations for the spatio-temporal scenario. Covariance is approximately 0 for errors from data points that are $\sqrt{2}$ distance units apart in space and 1 distance unit apart in time. The spatial dependent error variance (σ_δ^2), spatial independent error variance (σ_γ^2), temporal dependent error variance (σ_τ^2), and temporal independent error variance (σ_η^2) are shown with grey lines.

344 The third method uses a simple random sample (SRS) design-based estimator with
 345 data from the most current time point. The SRS design-based estimator for the total
 346 is $100 \cdot \bar{y}$, where \bar{y} is the sample mean of the response in the most current time point.
 347 The variance of the estimator (Lohr 2021) is $100^2 \cdot \frac{s^2}{n_1} \cdot (1 - \frac{n_1}{100})$, where s^2 is the sample
 348 variance of the response variable in the most current time point and n_1 is the number
 349 of sampled locations in the most current time point.

350 The SRS method gives an estimator, not a predictor, and a corresponding confidence
 351 interval, not a prediction interval, because the SRS design-based estimator treats the
 352 observed data as fixed, not as a random realization from a process (Brus 2021; Dumelle
 353 et al. 2022). However, in the remaining text and tables, we refer to the “current
 354 total” response quantity obtained from the three methods as a “prediction” and to
 355 the corresponding interval as a “prediction interval” to limit unnecessarily verbose
 356 text and tables.

357 For each method, we calculate the root-mean-squared-prediction-error (rMSPE)
 358 as $\sqrt{\frac{1}{1000}(\sum_{i=1}^{1000}(T_i - \hat{T}_i)^2)}$, where T_i and \hat{T}_i are the realized and predicted current
 359 totals, respectively, in the i^{th} iteration. Bias is recorded as $\frac{1}{1000} \sum_{i=1}^{1000}(T_i - \hat{T}_i)$. We
 360 also create a normal-based 90% prediction interval for the realized current total and
 361 record $\frac{1}{1000} \sum_{i=1}^{1000} I(LB_i < T_i < UB_i)$, where $I(LB_i < T_i < UB_i)$ is an indicator
 362 variable that is equal to 1 if the realized total in iteration i , T_i , is between the lower
 363 bound, LB_i , and the upper bound, UB_i , of the i^{th} prediction interval.

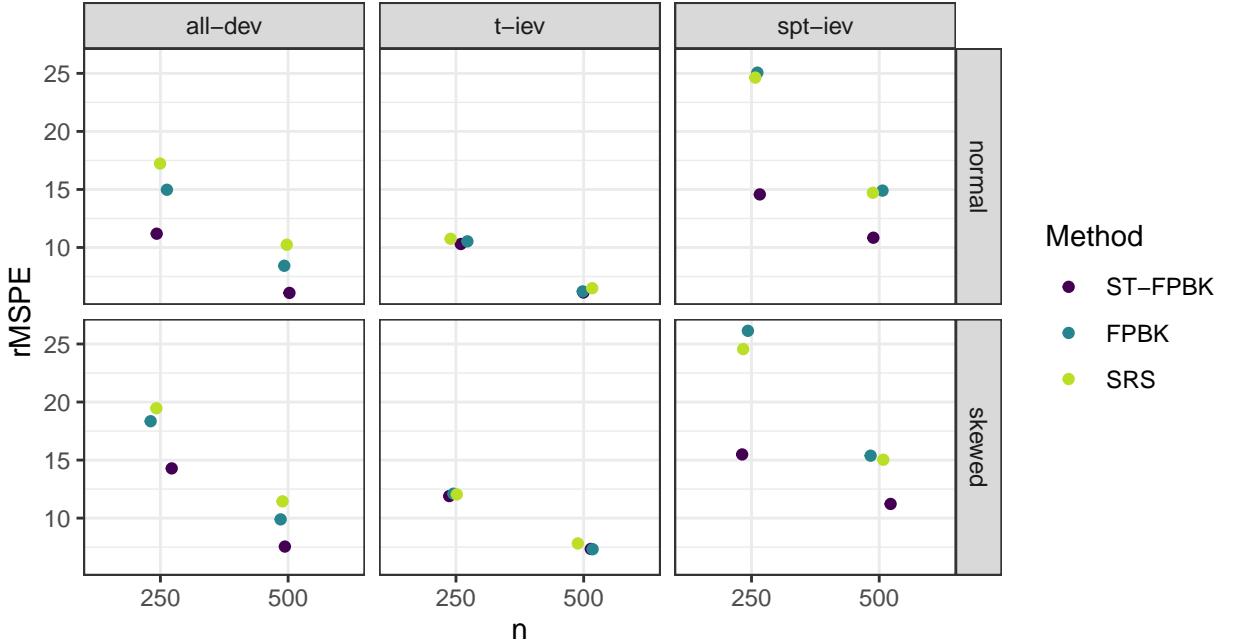


Figure 7. root-mean-squared-prediction-error (rMSPE) for all simulation settings. The ST-FPBK predictor has the smallest rMSPE in all settings tested, though it is similar to the rMSPE of the other two methods in the t-iev scenario.

364 4.2. Results

365 Tables A1, A2, and A3 in the Appendix give the rMSPE, bias, and interval coverage
 366 of the three methods in all 12 simulation settings. In Figure 7, we see that the ST-
 367 FPBK predictor outperforms both the purely spatial FPBK predictor and the simple
 368 random sample design-based estimator in all of the “all-dev” and “spt-iev” scenarios.
 369 In general, rMSPE improvement is larger for the smaller sample size.

370 We see little gains in rMSPE for the ST-FPBK predictor in the “t-iev” scenario.
 371 This setting was chosen to explore how the spatio-temporal model would perform
 372 when most of the variability in the response comes from σ_η^2 , which allows for data
 373 collected in different time points to be uncorrelated, and, for different time points to
 374 have very different realized totals. As expected, the ST-FPBK predictor performs no
 375 better than a purely spatial model or the SRS design-based estimator for this scenario;
 376 however, we can also say that the added complexity of the spatio-temporal model is
 377 not detrimental.

378 All methods appear relatively unbiased in all simulation settings: Table A2 shows
 379 that the bias of each method is small compared to the squares of the rMSPE values
 380 given in Table A1.

381 Figure 8 shows the interval coverage for the normal-based prediction intervals
 382 (Smith 1980), where the nominal level is 0.90. We see that the ST-FPBK predictor
 383 for the current total has approximate 90% coverage in all settings tested. The
 384 spatial model and the SRS design-based estimator have lower than nominal coverage
 385 in some settings because of the small sample size used (recall that the $n = 250$ ob-
 386 served samples span 10 unique time points so that, on average, the spatial model and
 387 SRS design-based estimator only have 25 observed responses to use in the current time
 388 point).

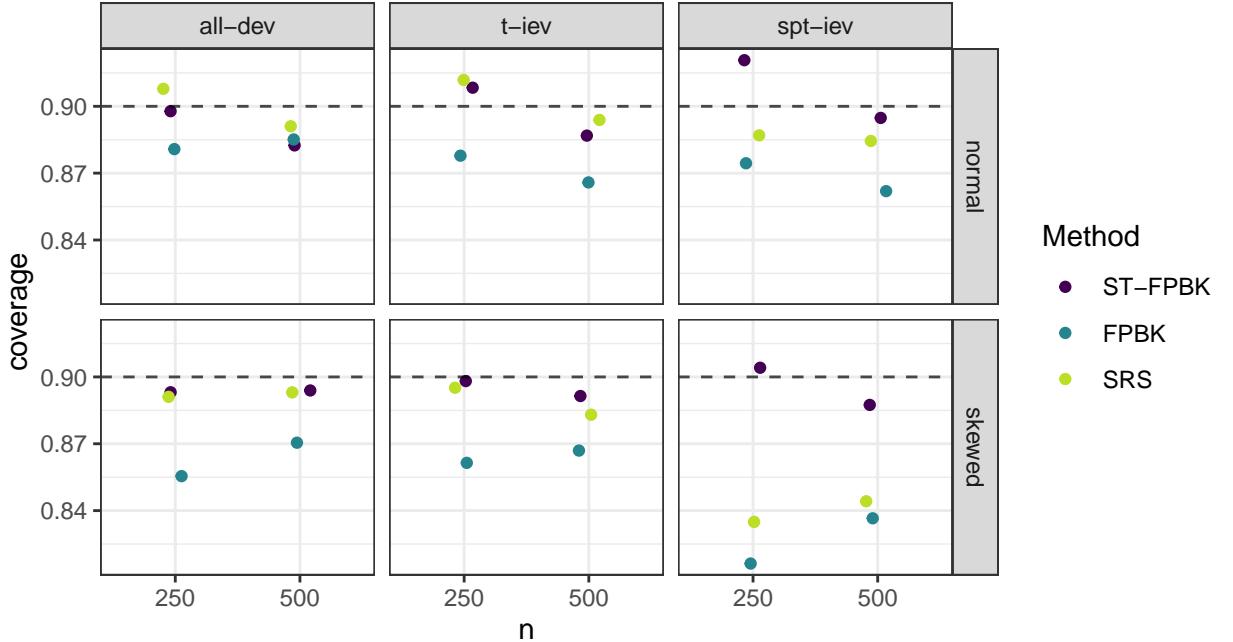


Figure 8. Prediction interval coverage for all simulation settings, where the prediction intervals are normal-based and the nominal level is 0.90. The ST-FPBK predictor has close to appropriate coverage in all settings tested.

389 5. Discussion

390 We see in the moose application in Section 3 that there is substantial reduction in the
 391 standard error of the predictor for the total moose abundance in 2020 when incorpo-
 392 rating data from surveys in previous years. In the simulation study in 4, we find that
 393 the ST-FPBK predictor has lower rMSPE than the FPKB predictor from a purely
 394 spatial model and an SRS design-based estimator in many settings. The ST-FPBK
 395 predictor is less beneficial when the temporal independent error variance contributes
 396 a large proportion to the overall variance. Additionally, the ST-FPBK predictor main-
 397 tains appropriate interval coverage in all settings tested, even when the covariance for
 398 the errors is mis-specified.

399 An additional possible benefit of using the ST-FPBK predictor compared to a purely
 400 spatial FPKB predictor is the potential for forecasting abundance before a survey is
 401 completed. In Figure 5, we see the forecasted prediction for abundance in the year
 402 2021. While there is a (presumed) loss in precision by constructing a prediction for a
 403 year that has no observed samples, the prediction could still be useful to wildlife man-
 404 agers for decision-making before a survey from that year is completed and analyzed.
 405 Constructing a prediction for years or time points at which a survey is not completed
 406 can be applied to other contexts as well, including temporal interpolation (e.g., the
 407 year 2016 in Figure 5).

408 The ability to predict the abundance (or other quantity) in time points that were
 409 not surveyed also allows biologists to investigate how much efficiency is lost from, for
 410 example, sampling every other year instead of every year. These types of surveys are
 411 often expensive, so perhaps the drop in efficiency from sampling every other year is
 412 worth the cost of completing those surveys annually.

413 We would also like to give our perception of the benefits and drawbacks of our

approach with that of Schmidt et al. (2022), who use a hierarchical Bayesian model with spatial radial basis functions that are estimated per year and with time as a trend component in the fixed effects. The benefits of our approach include a faster fitting time, as there is no need to construct and implement the time-consuming Markov chain Monte Carlo sampler. Therefore, our approach is easier to assess in a simulation study, which would be too time-prohibitive for the Bayesian model. Biometricians could also use simulation with our approach to answer various questions given proposed values of covariance parameters like how much efficiency would drop if a survey was only conducted every other year. We argue that our approach is simpler overall for a practitioner to use and could be integrated more readily with the current GSPE software. Finally, our approach allows for temporal interpolation and forecasting while the estimation of the spatial radial basis functions in Schmidt et al. (2022) for each time point do not allow for inference outside of the time points observed.

The Bayesian approach by Schmidt et al. (2022), however, offers features that would be harder to implement in our approach. Their method allows for incorporation of more levels in the Bayesian hierarchical model, including allowing for imperfect detection of animals from a separate detectability survey. Additionally, the Bayesian hierarchical model can use a Poisson or negative binomial model for the counts. Therefore, an appropriate prediction interval for the response on one particular site could be constructed. On the other hand, for our approach, we rely on the central limit theorem for dependent data to form a prediction interval for the total, which would not apply for a prediction interval for the response on just one site.

We have developed a finite population block kriging predictor for spatio-temporal data, which adjusts the variance of the predictor to be appropriate for sampling from a finite population. The resulting predictor is generally at least as good as the predictor from a purely spatial model, and, is often much better. Monitoring programs that use regularly scheduled surveys should consider incorporating data from past surveys to improve precision in the predictor for the most current survey.

Future work in this area includes developing a frequentist model for which imperfect detection of units through time is incorporated into the predictor or how best to select sites to sample for future surveys given proposed values for the spatio-temporal covariance parameters. Additionally, for moose surveys in particular, updating the GSPE software to include analysis for spatio-temporal data could be useful for practitioners. Though we recognize that doing so would be a substantial undertaking, the R package that we provide could be a useful starting point for the integration.

449 Data and Code Availability

450 This manuscript has a supplementary R package that contains all of the data and code
451 used in its creation, with the exception of the shapefile used to make the maps in some
452 of the figures (which cannot be released due to Alaska Department of Fish and Game
453 policy). The supplementary R package is hosted on GitHub and can be found at
454 <https://github.com/highamm/FPSpatioTemp>.

455 6. Acknowledgements

456 The views expressed in this manuscript are those of the authors and do not neces-
457 sarily represent the views or policies of the U.S. Environmental Protection Agency or

458 the National Oceanic and Atmospheric Administration. Any mention of trade names,
 459 products, or services does not imply an endorsement by the U.S. government, the
 460 U.S. Environmental Protection Agency, or the National Oceanic and Atmospheric
 461 Administration. The U.S. Environmental Protection Agency and National Oceanic
 462 and Atmospheric Administration do not endorse any commercial products, services,
 463 or enterprises.

464 **Appendix**

Table A1. root-mean-squared-prediction-error (rMSPE) for the ST-FPBK predictor, the FPBK predictor, and the SRS estimator for each of the 12 simulation settings. In all settings, the rMSPE for the ST-FPBK predictor is approximately equal to or lower than the rMSPE for the other two methods.

scenario	n	Response Type	rMSPE		
			ST-FPBK	FPBK	SRS
spt-iev	250	normal	14.58	25.06	24.64
t-iev	250	normal	10.31	10.53	10.76
all-dev	250	normal	11.18	14.97	17.23
spt-iev	500	normal	10.84	14.91	14.71
t-iev	500	normal	6.12	6.22	6.50
all-dev	500	normal	6.09	8.43	10.24
spt-iev	250	skewed	15.49	26.14	24.56
t-iev	250	skewed	11.89	12.10	12.05
all-dev	250	skewed	14.28	18.35	19.46
spt-iev	500	skewed	11.22	15.38	15.04
t-iev	500	skewed	7.34	7.32	7.82
all-dev	500	skewed	7.55	9.89	11.45

Table A2. Bias (Realized Current Total - Predicted Current Total) for the ST-FPBK predictor, the FPBK predictor, and the SRS estimator for each of the 12 simulation settings. In all settings, all methods appear fairly unbiased.

Simulation Setting			Bias		
scenario	n	Response Type	ST-FPBK	FPBK	SRS
spt-iev	250	normal	0.73	1.38	1.55
	250	normal	0.44	0.39	0.47
	250	normal	0.48	0.27	0.45
t-iev	500	normal	0.46	0.60	0.67
	500	normal	0.15	0.14	0.07
	500	normal	0.04	0.07	0.04
all-dev	250	skewed	0.36	0.56	1.48
	250	skewed	0.33	0.22	0.41
	250	skewed	-0.07	-0.85	-0.49
spt-iev	500	skewed	0.32	0.29	0.66
	500	skewed	0.24	0.15	0.08
	500	skewed	-0.10	-0.39	-0.37

Table A3. Prediction interval coverage for the ST-FPBK predictor, the FPBK predictor, and the SRS for each of the 12 simulation settings. All intervals are normal-based and have a nominal coverage level of 0.90.

Simulation Setting			Coverage		
scenario	n	Response Type	ST-FPBK	FPBK	SRS
spt-iev	250	normal	0.92	0.87	0.89
	250	normal	0.91	0.88	0.91
	250	normal	0.90	0.88	0.91
t-iev	500	normal	0.90	0.86	0.88
	500	normal	0.89	0.87	0.89
	500	normal	0.88	0.89	0.89
all-dev	250	skewed	0.90	0.82	0.84
	250	skewed	0.90	0.86	0.90
	250	skewed	0.89	0.86	0.89
spt-iev	500	skewed	0.89	0.84	0.84
	500	skewed	0.89	0.87	0.88
	500	skewed	0.89	0.87	0.89

465 **References**

- 466 Boertje, Rodney D, Mark A Keech, Donald D Young, Kalin A Kellie, and C Tom Seaton.
467 2009. "Managing for elevated yield of moose in interior Alaska." *The Journal of Wildlife
468 Management* 73 (3): 314–327.
- 469 Brus, Dick J. 2021. "Statistical approaches for spatial sample survey: Persistent misconceptions
470 and new developments." *European Journal of Soil Science* 72 (2): 686–703.
- 471 Cressie, Noel. 2015. *Statistics for spatial data - revised edition*. John Wiley & Sons.
- 472 Cressie, Noel, and Soumendra Nath Lahiri. 1993. "The asymptotic distribution of REML
473 estimators." *Journal of multivariate analysis* 45 (2): 217–233.
- 474 De Cesare, Luigi, DE Myers, and D Posa. 2001. "Product-sum covariance for space-time model-
475 ing: an environmental application." *Environmetrics: The official journal of the International
476 Environmetrics Society* 12 (1): 11–23.
- 477 De Iaco, S, M Palma, and D Posa. 2015. "Spatio-temporal geostatistical modeling for French
478 fertility predictions." *Spatial Statistics* 14: 546–562.
- 479 De Iaco, Sandra, Donald E Myers, and Donato Posa. 2001. "Space–time analysis using a
480 general product–sum model." *Statistics & Probability Letters* 52 (1): 21–28.
- 481 DeLong, Robert A. 2006. *Geospatial population estimator software user's guide*. Alaska De-
482 partment of Fish and Game, Division of Wildlife Conservation.
- 483 Dumelle, Michael, Matt Higham, Jay M Ver Hoef, Anthony R Olsen, and Lisa Madsen. 2022.
484 "A comparison of design-based and model-based approaches for finite population spatial
485 sampling and inference." *Methods in Ecology and Evolution* 13 (9): 2018–2029.
- 486 Dumelle, Michael, Jay M Ver Hoef, Claudio Fuentes, and Alix Gitelman. 2021. "A linear mixed
487 model formulation for spatio-temporal random processes with computational advances for
488 the product, sum, and product–sum covariance functions." *Spatial Statistics* 43: 100510.
- 489 Gasaway, William C, Stephen D DuBois, Daniel J Reed, and Samuel J Harbo. 1986. *Estimating
490 moose population parameters from aerial surveys*. Technical Report. University of Alaska.
491 Institute of Arctic Biology.
- 492 Harville, David A. 1977. "Maximum likelihood approaches to variance component estimation
493 and to related problems." *Journal of the American statistical association* 72 (358): 320–338.
- 494 Heyde, CC. 1994. "A quasi-likelihood approach to the REML estimating equations." *Statistics
495 & Probability Letters* 21 (5): 381–384.
- 496 Higham, Matt, Jay Ver Hoef, Bryce Frank, and Michael Dumelle. 2021. *sptotal: Predicting
497 Totals and Weighted Sums from Spatial Data*. R package version 1.0.0, <https://highamm.github.io/sptotal/index.html>.
- 498 Kellie, Kalin A, and Robert A DeLong. 2006. *Geospatial survey operations manual*. Technical
499 Report. Alaska Department of Fish and Game.
- 500 Kellie, Kalin Ann, Cassidy E Colson, and Joel H Reynolds. 2019. *Challenges to Monitoring
501 Moose in Alaska*. Alaska Department of Fish and Game, Division of Wildlife Conservation
502 Juneau
- 503 Lemos, Ricardo T, and Bruno Sansó. 2009. "A spatio-temporal model for mean, anomaly, and
504 trend fields of North Atlantic sea surface temperature." *Journal of the American Statistical
505 Association* 104 (485): 5–18.
- 506 Lohr, Sharon L. 2021. *Sampling: design and analysis*. Chapman and Hall/CRC.
- 507 Martínez-Beneito, Miguel A, Antonio López-Quilez, and Paloma Botella-Rocamora. 2008. "An
508 autoregressive approach to spatio-temporal disease mapping." *Statistics in medicine* 27 (15):
509 2874–2889.
- 510 Patterson, H Desmond, and Robin Thompson. 1971. "Recovery of inter-block information
511 when block sizes are unequal." *Biometrika* 58 (3): 545–554.
- 512 Peters, Wibke, Mark Hebblewhite, Kirby G Smith, Shevenell M Webb, Nathan Webb, Mike
513 Russell, Curtis Stambaugh, and Robert B Anderson. 2014. "Contrasting aerial moose pop-
514 ulation estimation methods and evaluating sightability in west-central Alberta, Canada."
515 *Wildlife Society Bulletin* 38 (3): 639–649.
- 516 Sahu, Sujit K, and Dankmar Böhning. 2022. "Bayesian spatio-temporal joint disease mapping

- 518 of Covid-19 cases and deaths in local authorities of England.” *Spatial Statistics* 49: 100519.
- 519 Schabenberger, Oliver, and Carol A Gotway. 2017. *Statistical methods for spatial data analysis:*
- 520 *Texts in statistical science*. Chapman and Hall/CRC.
- 521 Schmidt, Joshua H, Matthew D Cameron, Kyle Joly, Jordan M Pruszenski, Joel H Reynolds,
- 522 and Mathew S Sorum. 2022. “Bayesian spatial modeling of moose count data: increasing
- 523 estimator efficiency and exploring ecological hypotheses.” *The Journal of Wildlife Management*
- 524 e22220.
- 525 Smith, Tony E. 1980. “A central limit theorem for spatial samples.” *Geographical Analysis* 12
- 526 (4): 299–324.
- 527 Ver Hoef, Jay M. 2008. “Spatial methods for plot-based sampling of wildlife populations.”
- 528 *Environmental and Ecological Statistics* 15 (1): 3–13.
- 529 Wickham, Hadley. 2016. “Data analysis.” In *ggplot2*, 189–201. Springer.
- 530 Wikle, Christopher K, Andrew Zammit-Mangion, and Noel Cressie. 2019. *Spatio-temporal*
- 531 *Statistics with R*. Chapman and Hall/CRC.
- 532 Xu, Jiaqi, and Hong Shu. 2015. “Spatio-temporal kriging based on the product-sum model:
- 533 some computational aspects.” *Earth Science Informatics* 8 (3): 639–648.