

**1 ARTICLE TEMPLATE**

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

**4 Matt Higham<sup>a</sup>, Michael Dumelle<sup>b</sup>, Carly Hammond<sup>c</sup>, Jay Ver Hoef<sup>d</sup>, Jeff Wells<sup>c</sup>**

**5 <sup>a</sup>St. Lawrence University Department of Mathematics, Computer Science, and Statistics**

**6 Canton, NY 13617; <sup>b</sup>United States Environmental Protection Agency Corvallis, OR 97333;**

**7 <sup>c</sup>Alaska Department of Fish and Game Fairbanks, AK 99701; <sup>d</sup>Marine Mammal Laboratory,**

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

**9 Washington 98115**

**10 ARTICLE HISTORY**

Compiled February 1, 2023

**12 ABSTRACT**

Finite population prediction with spatio-temporal modeling can be used to predict a quantity of a finite resource from a sample of data collected over both spatial indices and temporal indices. We develop a spatio-temporal finite population block kriging (st-FPBK) predictor that incorporates an appropriate variance reduction for sampling from a finite population. Through an application to moose surveys in the east-central region of Alaska, we show that the predictor has a substantially smaller standard error compared to a predictor from the purely spatial model that is currently used to analyze moose surveys in the region. A separate simulation study shows that the spatio-temporal predictor is unbiased and that prediction intervals from the st-FPBK predictor attain appropriate coverage. For ecological monitoring surveys completed with some regularity through time, use of st-FPBK could improve precision. Therefore, we also give an R package that ecologists and resource managers could use to incorporate data from past surveys in predicting a quantity from a current survey.

**27 KEYWORDS**

spatial; temporal; kriging; total; resource monitoring

**29 1. Introduction**

**30 1.1. *Background***

31 Spatio-temporal data is indexed by both a spatial index, which we will refer to as  
32 a “site,” and by a temporal index, which we will refer to as a “time point.” Com-  
33 mon examples of spatio-temporal data include infections from a disease in a coun-  
34 try or region collected over a time period (e.g. Martínez-Beneito, López-Quilez, and  
35 Botella-Rocamora 2008; Sahu and Böhning 2022) or climate variables that are recorded  
36 through time at multiple locations (Lemos and Sansó 2009).

37 Models for spatio-temporal data have applications in a wide variety of scientific  
38 fields (see Wikle, Zammit-Mangion, and Cressie 2019, for many examples). One such

---

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

39 application is ecological monitoring of a particular resource, such as animal or plant  
40 abundance, rainfall, concentration of a compound in soil samples, etc.

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

### 48 1.2. *Motivating Example*

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

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

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

## 81 2. Methods

82 We now give details on the development of the spatio-temporal model and subsequently  
83 use this model to develop a finite population correction factor to give a Best-Linear-

84 Unbiased-Predictor (BLUP) and its prediction variance for any linear function of the  
 85 response vector.

### 86 **2.1. Spatio-temporal Model**

87 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  
 88 a spatial site and a time point, where the vector  $\mathbf{s}_i$  contains the coordinates for the  
 89  $i^{th}$  spatial site,  $n_s$  is the number of unique sites,  $t_j$  is the time index for the  $j^{th}$  time  
 90 point, and  $n_t$  is the number of unique time points. If each site is represented at every  
 91 time point, a vector of the  $Y(\mathbf{s}_i, t_j)$ , denoted  $\mathbf{y}(\mathbf{s}_i, t_j)$ , has length  $n_s \cdot n_t \equiv N$ . Then, a  
 92 spatio-temporal model for  $\mathbf{y}(\mathbf{s}_i, t_j)$  is

$$93 \quad \mathbf{y}(\mathbf{s}_i, t_j) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}(\mathbf{s}_i, t_j), \quad (1)$$

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

$$97 \quad \boldsymbol{\epsilon}(\mathbf{s}_i, t_j) = \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}. \quad (2)$$

97 In the spatial component of equation 2 ( $\mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma}$ ), the matrix  $\mathbf{Z}_s$  is an  $N \times n_s$   
 98 matrix of 0's and 1's, where the values in a row corresponding to a data point at  
 99 site  $\mathbf{s}_i$  are 1 in the  $i^{th}$  column and 0 in all other columns.  $\boldsymbol{\delta}$  is a random vector with  
 100 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  
 101 matrix and  $\sigma_\delta^2$  is called the spatial dependent error variance (or spatial partial sill).  
 102 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  
 103 the  $n_s \times n_s$  identity matrix and  $\sigma_\gamma^2$  is called the spatial independent error variance (or  
 104 spatial nugget).

105 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  
 106 0's and 1's, where the values in a row corresponding to a data point at time point  $t_j$   
 107 are 1 in the  $j^{th}$  column and 0 in all other columns.  $\boldsymbol{\tau}$  is a random vector with mean  
 108  $\mathbf{0}$  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  
 109 and  $\sigma_\tau^2$  is called the temporal dependent error variance (or temporal partial sill).  $\boldsymbol{\eta}$  is  
 110 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  
 111  $n_t \times n_t$  identity matrix and  $\sigma_\eta^2$  is called the temporal independent error variance (or  
 112 temporal nugget).

113 In the spatio-temporal component of equation 2 ( $\boldsymbol{\omega} + \boldsymbol{\nu}$ ),  $\boldsymbol{\omega}$  is a random vector  
 114 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  
 115 correlation matrix and  $\sigma_\omega^2$  is sometimes called the spatio-temporal dependent error  
 116 variance (or spatio-temporal partial sill).  $\boldsymbol{\nu}$  is also a random vector with mean  $\mathbf{0}$  but  
 117 has covariance  $\text{cov}(\boldsymbol{\nu}) = \sigma_\nu^2 \mathbf{I}_{st}$ , where  $\mathbf{I}_{st}$  is the  $N \times N$  identity matrix and  $\sigma_\nu^2$  is  
 118 sometimes called the spatio-temporal independent error variance (or spatio-temporal  
 119 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

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

where  $\odot$  is the Hadamard product operator.  $\mathbf{R}_s$  can be parameterized in different ways, but one common assumption is to assume the covariance function generating  $\mathbf{R}_s$  is second-order stationary (ie. the covariance between two data points is a function only of the separation vector between two sites) and isotropic (ie. the covariance is a function of the distance only and does not depend on the direction of the separation vector). For example, the exponential covariance function is defined as follows. For observations at sites  $i$  and  $i'$  at  $h_{ii'}$  distance apart, row  $i$  and column  $i'$  of  $\mathbf{R}_s$  is equal to

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

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

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

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

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

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

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

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  $\mathbf{y}(\mathbf{s}_i, t_j)$  has covariance

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

There are a few reasons for why we choose to solely focus on the product-sum model. 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 the covariance matrix in equation 6 is also positive definite (De Cesare, Myers, and Posa 2001; De Iaco, Myers, and Posa 2001). Also, the product-sum model is flexible in its ability to model many kinds of spatial and temporal correlation (De Iaco, Palma, and Posa 2015). Xu and Shu (2015) claim that the product-sum model is the most widely used spatio-temporal model used in practical applications.

150    **2.2. Finite Population Block Kriging**

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

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

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

164    Unbiasedness implies that  $E(\mathbf{q}'\mathbf{y}_o) = E(\mathbf{b}'_a\mathbf{y}_a)$  for all  $\beta$ . So, denoting  $\mathbf{X}_o$  as the  
 165    design matrix for the observed data points and  $\mathbf{X}_a$  as the design matrix for all data  
 166    points,  $\mathbf{q}'\mathbf{X}_o\beta = \mathbf{b}'_a\mathbf{X}_a\beta$  for every  $\beta$ , implying that  $\mathbf{q}'\mathbf{X}_o = \mathbf{b}'_a\mathbf{X}_a$ . Kriging weights  
 167    are then found by finding  $\boldsymbol{\lambda}_o$ , an  $n_o \times 1$  column vector, where  $n_o$  is the number of  
 168    observed data points, such that

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

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

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

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

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

170    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

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

171 where  $\hat{\mathbf{y}}_u = \Sigma_{o,s}\Sigma_{o,o}^{-1}(\tilde{\mathbf{y}}_o - \hat{\mu}_o) + \hat{\mu}_u$  with  $\hat{\mu}_o = \mathbf{X}_o\hat{\beta}$  and  $\hat{\mu}_u = \mathbf{X}_u\hat{\beta}$ .  $\hat{\beta}$  is the generalized  
 172 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  
 173 multiplies the observed data  $\mathbf{y}_o$  with relevant weights from the  $\mathbf{b}_o$  vector, and then  
 174 adds in the kriged predictions  $\hat{\mathbf{y}}_u$  multiplied with relevant weights from the  $\mathbf{b}_u$  vector.

175 The prediction variance of the predictor in equation 11 is

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

176 We call the predictor in equation 11 with  $\Sigma$  in equation 6 the st-FPBK predictor.

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

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

184 where the subscripts *up*, *uc*, *op*, and *oc* denote unobserved sites in past surveys, un-  
 185 observed sites in the current survey, observed sites in past surveys, and observed sites  
 186 in the current survey, respectively.

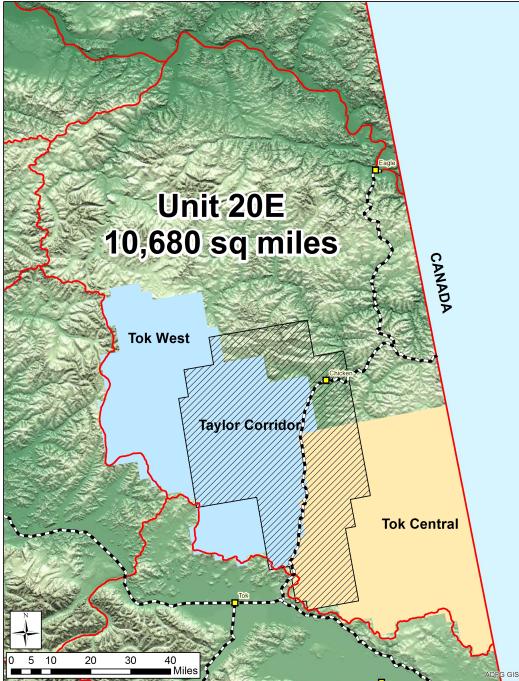
### 187 2.3. Estimation

188 In practical applications, the covariance matrix  $\Sigma$  in equation 6 that is partitioned  
 189 into the various sub-matrices in equations 11 and 12 needs to be estimated from the  
 190 observed data  $\mathbf{y}_o$ . The spatio-temporal model in equation 5 does not have any distri-  
 191 butional assumptions: we only need to specify the mean and variance of  $\mathbf{y}_o$ . Restricted  
 192 Maximum Likelihood (REML) can be used to estimate the covariance parameters in  
 193  $\Sigma$ , 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-  
 194 son 1971; Harville 1977). Even if  $\mathbf{y}_a$  is not multivariate normal, the REML estimator  
 195 for the parameter vector  $\boldsymbol{\theta}$  is still unbiased (Heyde 1994; Cressie and Lahiri 1993).

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

## 201 3. Application

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



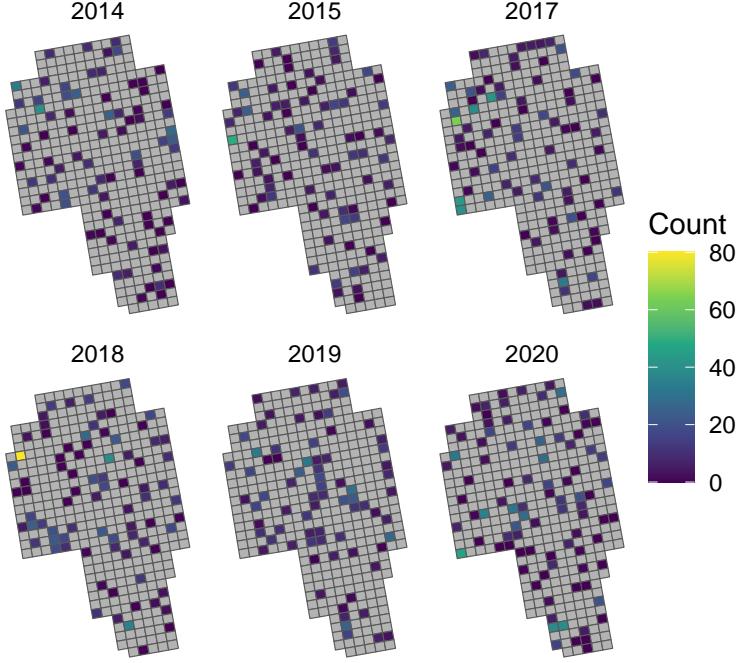
**Figure 1.** A map of the Taylor Corridor in the east-central region of Alaska.

### 205    3.1. Data Description

206    The Taylor Corridor in the east-central region of Alaska is a popular habitat for moose  
 207    and other wildlife. Abundance surveys for moose are performed in the Taylor Corridor  
 208    of the Tok region of Alaska annually (Figure 1) so that biologists have an idea about  
 209    the abundance of moose each year. In particular, surveys were conducted every year  
 210    from 2014 through 2020 in every year except 2016, during which there was not sufficient  
 211    snow cover to perform a survey. The spatial sampling frame for one particular survey  
 212    consists of 381 sites. There are a total of 7 unique time points represented in the data,  
 213    including the missing year of 2016. Therefore,  $N$  is 2667.

214    In each year of the survey, an aerial team of biologists selects some of the 381 sites to  
 215    survey. The number of sites that are selected varies from a low of 76 in the year 2019 to  
 216    a high of 90 in the year 2020. Throughout the 7 unique years, some sites are sampled as  
 217    many as five different times while others are never sampled at all (Figure 2). Figure 2  
 218    and all remaining figure graphics are constructed with the `\texttt{ggplot2 R}` package  
 219    (Wickham 2016). The number of units sampled throughout all survey years,  $n$ , is 487  
 220    units.

221    Before the survey begins in each year, biologists stratify the sites into a “High”  
 222    stratum composed of 230 sites and a “Low” stratum composed of 151 sites. The goal  
 223    of the following analysis is to predict the total abundance of moose across all sites in  
 224    the year 2020, the most recent year of the survey, using stratum as a covariate in the  
 225    spatio-temporal model.



**Figure 2.** Layout of the spatial sites used to survey moose in the Taylor corridor of the TOK region of Alaska, coloured by moose count. The year 2016 is excluded because no survey was performed in that year.

### 226 3.2. Model Fitting

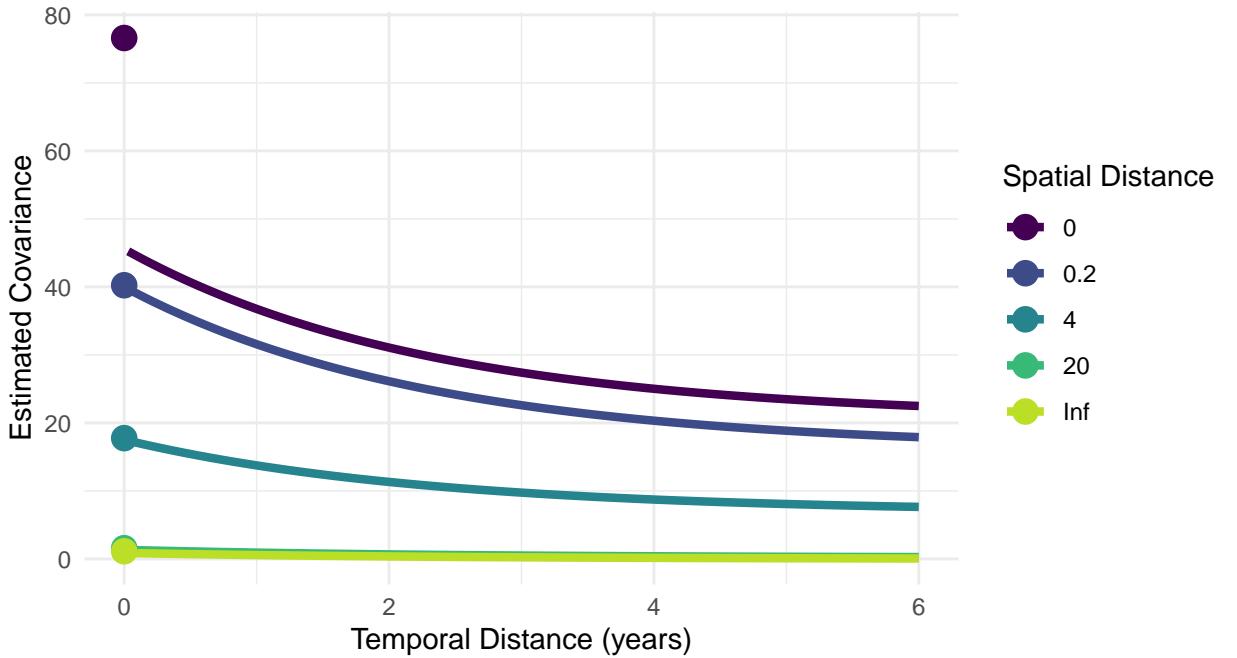
227 We fit the product-sum covariance model defined in equation 5 using REML with  
 228 stratum as a covariate in the design matrix, an exponential spatial correlation structure  
 229 defined in equation 3, and an exponential temporal correlation structure defined in  
 230 equation 4. Table 1 gives 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	
$\hat{\sigma}_\delta^2$	$\hat{\sigma}_\gamma^2$	$\hat{\phi}$	$\hat{\sigma}_\tau^2$	$\hat{\sigma}_\eta^2$	$\hat{\rho}$	$\hat{\sigma}_\omega^2$	$\hat{\sigma}_\nu^2$
16.91	3.76	4.44	0.88	0.25	2.29	24.01	30.8

231 To help interpret what some of these fitted covariance parameter estimates mean,  
 232 we can construct a fitted covariance plot (Figure 3). As the spatial distance between  
 233 two sites increases (dark colour to light colour), the covariance of two errors decreases  
 234 to 0, with the  $\hat{\phi}$  parameter estimate controlling the rate of decay. In fact, the model  
 235 estimates the covariance to be nearly 0 when two sites are 20 or more kilometers apart,  
 236 no matter what the temporal distance is. The covariance between two errors that are  
 237 six years apart is still estimated to be positive if the two errors come from the same  
 238 site or from adjacent sites.

239 The estimated vector of fixed effects, using “High” as the reference group, is  $\hat{\beta} =$   
 240  $(11.26, -9.76)$ . Therefore, the overall mean for sites in the “High” stratum is estimated



**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.

241 to be 11.26 moose while the overall mean for sites in the “Low” stratum is estimated  
 242 to be 1.5 moose.

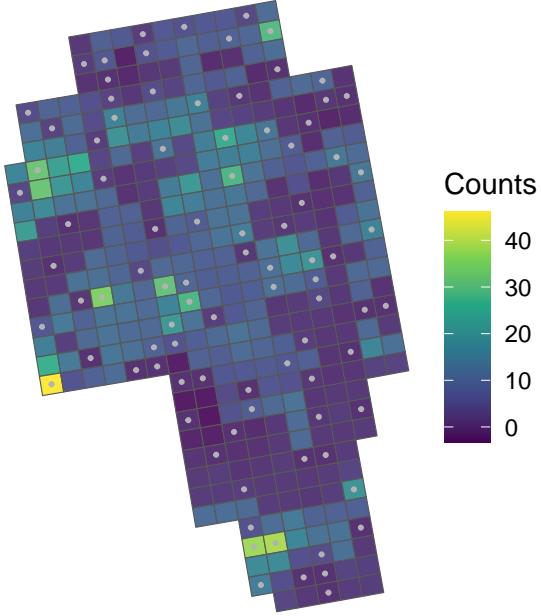
243 **3.3. Prediction**

244 We now use the fitted spatio-temporal model with the BLUP from equation 11 and  
 245 weights given in equation 13 to predict the total abundance across all sites in the  
 246 year 2020, the most recent year of the survey. Plugging in estimates of the covariance  
 247 parameters into equations 11 and 12 and letting elements of  $\mathbf{b}_a$  be equal to 1 for  
 248 data points in 2020 and equal to 0 otherwise, we obtain a prediction of 2874 moose  
 249 and a standard error (the square root of the prediction variance) of 234 moose. A 90%  
 250 normal-based prediction interval for the total abundance in 2020 is (2489, 3259) moose.  
 251 Note that, though the response in this example is a count, a normal-based prediction  
 252 interval for the total is still appropriate through an application of the central limit  
 253 theorem for dependent data (Smith 1980). Sitewise predictions for sites in 2020 are  
 254 given in the map in Figure 4.

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

We also use the stratified random sampling design-based estimator

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



**Figure 4.** 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.

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 ( $SE = 234$  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).

## 272 4. Simulation

### 273 4.1. Description

274 To evaluate performance of the st-FPBK predictor, we conduct a simulation study.  
 275 We simulate a response vector  $\mathbf{y}$  of length  $N = 1000$  on a  $10 \times 10$  grid of 100 spatial

sites on the unit square ( $[0, 1] \times [0, 1]$ ) and 10 equally-spaced time points in the interval  $[0, 1]$ , so that each spatial site has a response value at each time point.  $\mathbf{y}$  is multivariate normal with mean  $\mathbf{0}$  and product-sum covariance matrix  $\Sigma$  defined in equation 6 with the covariance parameters given in Table 2.

**Table 2.** Covariance parameters used to simulate data.  $\sigma_\delta^2$ ,  $\sigma_\gamma^2$ , and  $\phi$  are the spatial dependent error variance, independent error variance, and range parameters, respectively.  $\sigma_\tau^2$ ,  $\sigma_\eta^2$ , and  $\rho$  are the temporal dependent error variance, independent error variance, and range parameters, respectively.  $\sigma_\omega^2$  and  $\sigma_\nu^2$  are the spatio-temporal dependent error variance and spatio-temporal independent error variance.

scenario	Spatial			Temporal			Spatio-temporal	
	$\sigma_\delta^2$	$\sigma_\gamma^2$	$\phi$	$\sigma_\tau^2$	$\sigma_\eta^2$	$\rho$	$\sigma_\omega^2$	$\sigma_\nu^2$
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

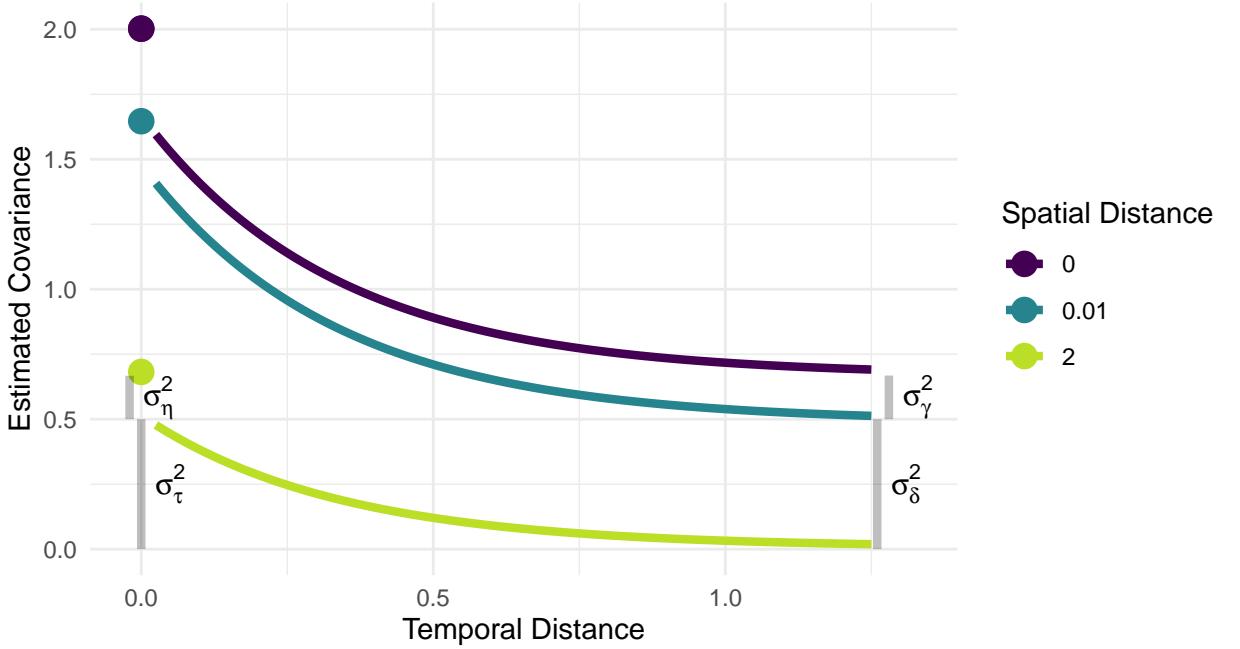
The three scenarios in the table correspond to (1) **all-dev**: a scenario where a substantial proportion of the overall variance comes from the spatial, temporal, and spatio-temporal dependent error variance parameters  $\sigma_\delta^2$ ,  $\sigma_\tau^2$ , and  $\sigma_\omega^2$ ; (2) **t-iev**: a scenario where there the overall variance is dominated by the temporal independent error variance parameter,  $\sigma_\eta^2$ ; and (3) **spt-iev**: a scenario where all of the variability comes from  $\sigma_\nu^2$  so that errors are independent regardless of spatial and time indices. In all scenarios, summing all six variance parameters gives a total variance equal to two.

Both  $\mathbf{R}_s$  and  $\mathbf{R}_t$  are generated from the exponential correlation function with  $\phi$  and  $\rho$  as the range parameters in equations 3 and 4. The values 0.471 and 0.3333 are chosen for  $\phi$  and  $\rho$ , respectively, so that the effective ranges,  $3\phi$  and  $3\rho$ , are equal to the maximum distance between two data points in space ( $\sqrt{2} = 1.414$ ) and the maximum distance between two data points in time (1). A value of 0 for  $\phi$  (or  $\rho$ ) sets the  $\mathbf{R}_s$  (or the  $\mathbf{R}_t$ ) matrix to the identity matrix. Figure 5 shows the model covariance of the errors used to generate data for the “all-dev” scenario.

Each of these three scenarios is replicated for two different sample sizes:  $n = 250$  and  $n = 500$ . A simple random sample of the 1000 total data points is used to select units to be in the sample.

Finally, the simulation experiment is repeated for a skewed response variable. To create the skewed response variable, a normally-distributed response is simulated according to the parameters given in Table 2, except that each of the variance parameters (not including  $\phi$  and  $\rho$ ) is divided by 2.89 so that the total variance is equal to 0.6931. This variable is then exponentiated so that the total variance after exponentiation is equal to 2. Note that, not only does exponentiation result in a right-skewed response variable, but exponentiating also allows for an assessment of how the st-FPBK predictor performs when the covariance is mis-specified, as the resulting response variable is now simulated with an intractable covariance function form that is not used in the model fitting.

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



**Figure 5.** 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 or 1 distance unit apart in time. The spatial dependent error variance ( $\sigma_\delta^2$ ), spatial independent error variance ( $\sigma_\gamma^2$ ), temporal dependent error variance ( $\sigma_\tau^2$ ), and temporal independent error variance ( $\sigma_\eta^2$ ) are shown with grey lines.

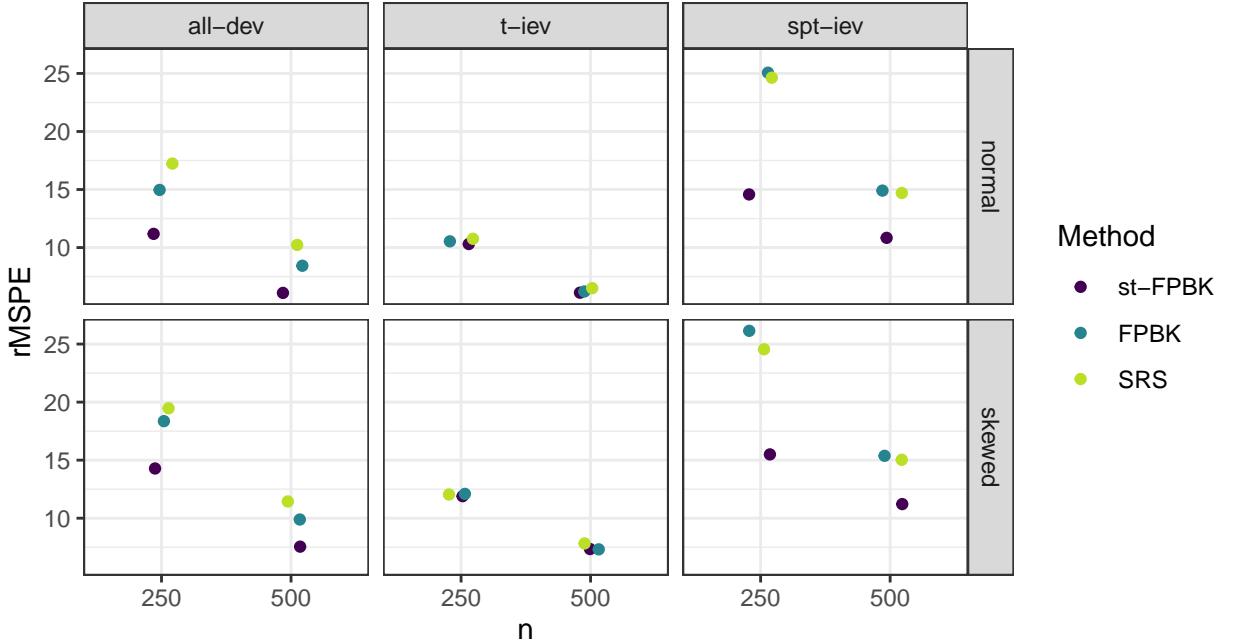
312 for the most current time point quantity” the “current total.”

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

318 The third method uses a simple random sample (SRS) design-based estimator with  
 319 data from the most current time point. The SRS design-based estimator for the total  
 320 is  $100 \cdot \bar{y}$ , where  $\bar{y}$  is the sample mean of the response in the most current time point.  
 321 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  
 322 variance of the response variable in the most current time point and  $n_1$  is the number  
 323 of sampled locations in the most current time point.

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

331 For each method, we calculate the root-mean-squared-prediction-error (rMSPE)  
 332 as  $\frac{1}{1000} \sqrt{(\sum_{i=1}^{1000} (T_i - \hat{T}_i)^2)}$ , where  $T_i$  and  $\hat{T}_i$  are the realized and predicted current  
 333 totals, respectively, in the  $i^{th}$  iteration. Bias is recorded as  $\frac{1}{1000} \sum_{i=1}^{1000} (T_i - \hat{T}_i)$ . We  
 334 also create a normal-based 90% prediction interval for the realized current total and  
 335 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



**Figure 6.** 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.

variable that is equal to 1 if the realized total in iteration  $i$ ,  $T_i$ , is between the lower bound,  $LB_i$ , and the upper bound,  $UB_i$ , of the  $i^{th}$  prediction interval.

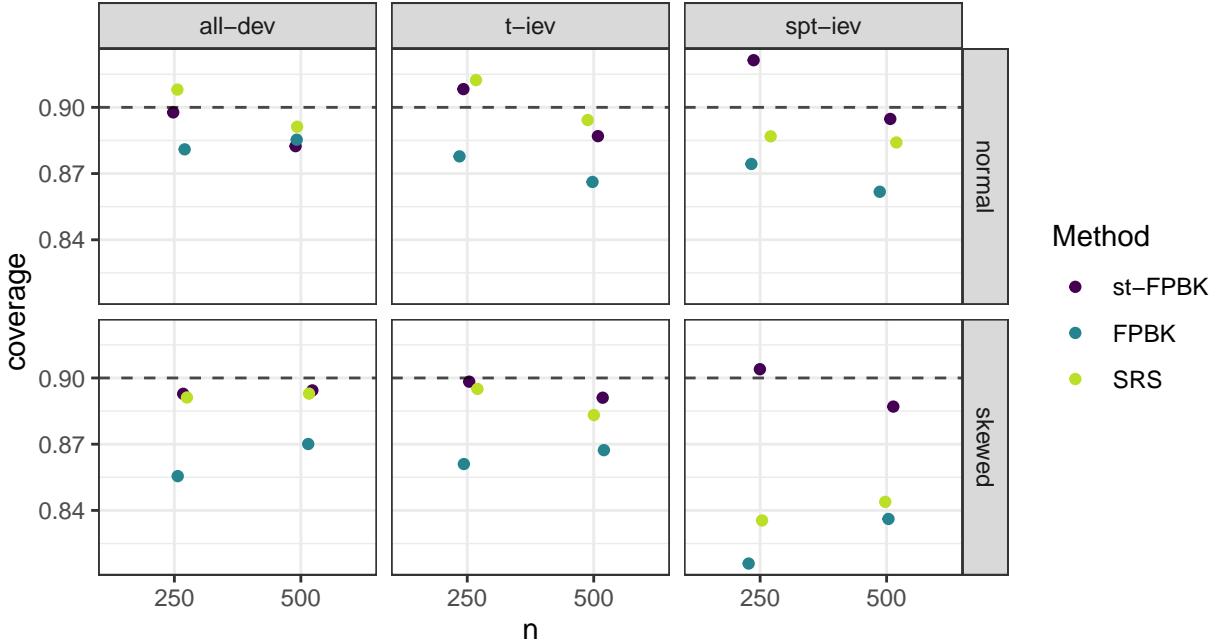
#### 4.2. Results

Tables 4, 5, and 6 in Section 6 give the rMSPE, bias, and interval coverage of the three methods in all 12 simulation settings. In Figure 6, we see that the st-FPBK predictor outperforms both the purely spatial FPKB predictor and the simple random sample design-based estimator in all of the “all-dev” and “spt-iev” scenarios. In general, rMSPE improvement is larger for the smaller sample size.

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

All methods appear relatively unbiased in all simulation settings: Table 5 shows that the bias of each method is small compared to the squares of the rMSPE values given in Table 4.

Figure 7 shows the interval coverage for the normal-based prediction intervals (Smith 1980), where the nominal level is 0.90. We see that the st-FPBK predictor for the current total has approximate 90% coverage in all settings tested. The spatial model and the SRS design-based estimator have lower than nominal coverage in some settings because of the small sample size used (recall that the  $n = 250$  observed



**Figure 7.** 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.

samples span 10 unique time points so that, on average, the spatial model and SRS design-based estimator only have 25 observed responses to use in the current time point).

## 5. Discussion

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

An additional possible benefit of using the st-FPBK predictor compared to a purely spatial FPBK predictor is the potential for forecasting abundance before a survey is completed. For example, in our application, we can refit the model without any of the observed counts from the 2020 survey and examine the prediction for the total abundance in the year 2020. Table 3 compares the model with the 2020 data used and the model without the 2020 data used. We see that, while there is a substantial loss in precision by excluding the 2020 data (as we would expect), the prediction is not very different from the prediction with the 2020 data included. And, the prediction interval might be narrow enough to still be useful to wildlife management. We could also consider using such an approach if there is a year during which a survey cannot

383 be completed for logistical reasons. For example, in the Tok region of Alaska, a moose  
384 survey was not conducted at all in the year 2016 because there was insufficient snow  
385 cover for the survey. The spatio-temporal predictor could still be applied to get a  
386 prediction for moose abundance using survey data from previous years.

**Table 3.** Results from analysis on the Tok moose survey data with the 2020 survey included and excluded. We can see that, even with 2020 data excluded, we can obtain a prediction for moose abundance in 2020, though there is substantial loss in precision.

	Prediction	SE	90% LB	90% UB
2020 Data Included	2874	234	2489	3259
2020 Data Excluded	2757	417	2071	3444

387 We would also like to give our perception of the benefits and drawbacks of our  
388 approach with that of Schmidt et al. (2022), who use a hierarchical Bayesian model  
389 to predict total abundance, among other quantities of interest. The benefits of our  
390 frequentist approach include a faster fitting time, as there is no need to construct and  
391 implement the time-consuming Markov chain Monte Carlo sampler. Therefore, our  
392 approach is easier to assess in a simulation study, which would be too time-prohibitive  
393 for the Bayesian model. Biometricalians could also use simulation with our approach to  
394 answer various questions given proposed values of covariance parameters like how much  
395 efficiency would drop if a survey was only conducted every other year. Additionally,  
396 we argue that our approach is simpler overall for a practitioner to use and could be  
397 integrated more readily with the current GSPE software.

398 The Bayesian approach by Schmidt et al. (2022), however, offers features that would  
399 be harder to implement in our approach. In general, the model is more flexible, and  
400 allows for incorporation of more levels in the Bayesian hierarchical model, including  
401 allowing for imperfect detection of animals from a separate detectability survey. Addi-  
402 tionally, the Bayesian hierarchical model can use a Poisson or negative binomial model  
403 for the counts. Therefore, an appropriate prediction interval for the response on one  
404 particular site could be constructed. On the other hand, for our approach, we rely  
405 on the central limit theorem for dependent data to form a prediction interval for the  
406 total, which would not apply for a prediction interval for the response on just one site.

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

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

420 **6. Appendix**

**Table 4.** 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.

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

**Table 5.** 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 6.** 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

## 421 References

- 422 Boertje, Rodney D, Mark A Keech, Donald D Young, Kalin A Kellie, and C Tom Seaton.  
423 2009. “Managing for elevated yield of moose in interior Alaska.” *The Journal of Wildlife  
424 Management* 73 (3): 314–327.
- 425 Brus, Dick J. 2021. “Statistical approaches for spatial sample survey: Persistent misconceptions  
426 and new developments.” *European Journal of Soil Science* 72 (2): 686–703.
- 427 Cressie, Noel. 2015. *Statistics for spatial data - revised edition*. John Wiley & Sons.
- 428 Cressie, Noel, and Soumendra Nath Lahiri. 1993. “The asymptotic distribution of REML  
429 estimators.” *Journal of multivariate analysis* 45 (2): 217–233.
- 430 De Cesare, Luigi, DE Myers, and D Posa. 2001. “Product-sum covariance for space-time model-  
431 ing: an environmental application.” *Environmetrics: The official journal of the International  
432 Environmetrics Society* 12 (1): 11–23.
- 433 De Iaco, S, M Palma, and D Posa. 2015. “Spatio-temporal geostatistical modeling for French  
434 fertility predictions.” *Spatial Statistics* 14: 546–562.
- 435 De Iaco, Sandra, Donald E Myers, and Donato Posa. 2001. “Space-time analysis using a  
436 general product-sum model.” *Statistics & Probability Letters* 52 (1): 21–28.
- 437 DeLong, Robert A. 2006. *Geospatial population estimator software user’s guide*. Alaska De-  
438 partment of Fish and Game, Division of Wildlife Conservation.
- 439 Dumelle, Michael, Matt Higham, Jay M Ver Hoef, Anthony R Olsen, and Lisa Madsen. 2022.  
440 “A comparison of design-based and model-based approaches for finite population spatial  
441 sampling and inference.” *Methods in Ecology and Evolution* 13 (9): 2018–2029.
- 442 Dumelle, Michael, Jay M Ver Hoef, Claudio Fuentes, and Alix Gitelman. 2021. “A linear mixed  
443 model formulation for spatio-temporal random processes with computational advances for  
444 the product, sum, and product-sum covariance functions.” *Spatial Statistics* 43: 100510.
- 445 Gasaway, William C, Stephen D DuBois, Daniel J Reed, and Samuel J Harbo. 1986. *Estimating  
446 moose population parameters from aerial surveys*. Technical Report. University of Alaska.  
447 Institute of Arctic Biology.
- 448 Harville, David A. 1977. “Maximum likelihood approaches to variance component estimation

- 449 and to related problems.” *Journal of the American statistical association* 72 (358): 320–338.
- 450 Heyde, CC. 1994. “A quasi-likelihood approach to the REML estimating equations.” *Statistics  
& Probability Letters* 21 (5): 381–384.
- 452 Higham, Matt, Jay Ver Hoef, Bryce Frank, and Michael Dumelle. 2021. *sptotal: Predicting  
Totals and Weighted Sums from Spatial Data*. R package version 1.0.0, <https://highamm.github.io/sptotal/index.html>.
- 454 Kellie, Kalin A, and Robert A DeLong. 2006. *Geospatial survey operations manual*. Technical Report. Alaska Department of Fish and Game.
- 456 Kellie, Kalin Ann, Cassidy E Colson, and Joel H Reynolds. 2019. *Challenges to Monitoring  
Moose in Alaska*. Alaska Department of Fish and Game, Division of Wildlife Conservation  
Juneau . . . .
- 458 Lemos, Ricardo T, and Bruno Sansó. 2009. “A spatio-temporal model for mean, anomaly, and  
trend fields of North Atlantic sea surface temperature.” *Journal of the American Statistical  
Association* 104 (485): 5–18.
- 460 Lohr, Sharon L. 2021. *Sampling: design and analysis*. Chapman and Hall/CRC.
- 462 Martínez-Beneito, Miguel A, Antonio López-Quilez, and Paloma Botella-Rocamora. 2008. “An  
autoregressive approach to spatio-temporal disease mapping.” *Statistics in medicine* 27 (15):  
2874–2889.
- 464 Patterson, H Desmond, and Robin Thompson. 1971. “Recovery of inter-block information  
when block sizes are unequal.” *Biometrika* 58 (3): 545–554.
- 466 Peters, Wibke, Mark Hebblewhite, Kirby G Smith, Shevenell M Webb, Nathan Webb, Mike  
Russell, Curtis Stambaugh, and Robert B Anderson. 2014. “Contrasting aerial moose pop-  
ulation estimation methods and evaluating sightability in west-central Alberta, Canada.”  
*Wildlife Society Bulletin* 38 (3): 639–649.
- 468 Sahu, Sujit K, and Dankmar Böhning. 2022. “Bayesian spatio-temporal joint disease mapping  
of Covid-19 cases and deaths in local authorities of England.” *Spatial Statistics* 49: 100519.
- 470 Schabenberger, Oliver, and Carol A Gotway. 2017. *Statistical methods for spatial data analysis:  
Texts in statistical science*. Chapman and Hall/CRC.
- 472 Schmidt, Joshua H, Matthew D Cameron, Kyle Joly, Jordan M Pruszenski, Joel H Reynolds,  
and Mathew S Sorum. 2022. “Bayesian spatial modeling of moose count data: increasing  
estimator efficiency and exploring ecological hypotheses.” *The Journal of Wildlife Manage-  
ment* e22220.
- 474 Smith, Tony E. 1980. “A central limit theorem for spatial samples.” *Geographical Analysis* 12  
(4): 299–324.
- 476 Ver Hoef, Jay M. 2008. “Spatial methods for plot-based sampling of wildlife populations.”  
*Environmental and Ecological Statistics* 15 (1): 3–13.
- 478 Wickham, Hadley. 2016. “Data analysis.” In *ggplot2*, 189–201. Springer.
- 480 Wikle, Christopher K, Andrew Zammit-Mangion, and Noel Cressie. 2019. *Spatio-temporal  
Statistics with R*. Chapman and Hall/CRC.
- 482 Xu, Jiaqi, and Hong Shu. 2015. “Spatio-temporal kriging based on the product-sum model:  
some computational aspects.” *Earth Science Informatics* 8 (3): 639–648.