

## ARTICLE TEMPLATE

# An Application of Spatiotemporal Modeling to Finite Population Abundance Prediction

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>

<sup>a</sup>St. Lawrence University Department of Mathematics, Computer Science, and Statistics Canton, NY 13617; <sup>b</sup>United States Environmental Protection Agency Corvallis, OR 97333;

<sup>c</sup>Alaska Department of Fish and Game Fairbanks, AK 99701; <sup>d</sup>Marine Mammal Laboratory, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration Seattle, Washington 98115

## ARTICLE HISTORY

Compiled January 10, 2023

## ABSTRACT

Insert abstract here.

## KEYWORDS

spatial; temporal; kriging;

## 1. Introduction

### 1.1. *Motivation*

Moose surveys in Alaska and western Canada are performed annually in many regions. The primary goal of these surveys is to predict moose abundance, the total number of moose, in the region. Because of time and money constraints, only some areas (sites) in the region of interest are selected to be in the survey. Biologists fly to these selected sites, count the number of moose, and can then use a spatial statistical model to find a prediction for the finite abundance for that year (Ver Hoef 2008).

Though these surveys are annual, each survey is analysed completely independently of surveys from previous years (e.g. Gasaway et al. 1986; Kellie and DeLong 2006; Boertje et al. 2009; Peters et al. 2014). For example, a model for a survey conducted in the year 2019 only uses counts on sites that were sampled in that year. However, using counts from previous years in a model that incorporates both spatial and temporal correlation (spatiotemporal) could result in a prediction for the realized total or mean that is more precise than predictions from a spatial model using only counts from the most recent survey year.

Though the framework of the motivation is given with an example on moose surveys, this type of analysis could be useful for other monitoring systems with a finite number of sites that are regularly surveyed.

---

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)

## 1.2. Background

- Add paragraph about background of spatiotemporal models

Prediction for a total, a subset of the total, or a mean in a finite number of spatial locations should incorporate a finite population correction to the variance of the predictor (Ver Hoef 2008; Higham et al. 2021b). In the context of ecological monitoring in spatiotemporal prediction, we are often most interested in predicting the total abundance for the most recent year of the survey. In this case, the finite population correction should adjust based on the number of sites surveyed in the most recent year of the survey, so that, for example, the prediction variance is zero if all sites in the sampling frame in the most recent year are surveyed.

The rest of this paper is organized as follows. In Section 2, we couple spatiotemporal modeling with finite population prediction to develop the Best-Linear-Unbiased-Predictor for any linear function of a general response variable, including the total abundance across all sites. In Section 3, we apply the predictor to a moose data set in the TOC region of Alaska. In Section 4, we conduct a brief simulation study to examine the properties of the predictor. Finally, in Section 5, we conclude and give directions for future research.

## 2. Methods

We now give details on the development of the predictor for abundance. We first detail the spatiotemporal model. Because of the heavy use of notation in the spatiotemporal model development, we first introduce a purely spatial model (without temporal variability) and a purely temporal model (without spatial variability). We then build the spatiotemporal model from these two base components and develop a finite population correction factor to give a Best-Linear-Unbiased-Predictor (BLUP) and its prediction variance for any linear function of the response.

### 2.1. Spatial Model

First, we consider a spatial linear model for a response variable  $Y_s(\mathbf{s}_i)$ ,  $i = 1, 2, \dots, n_s$ , where the vector  $\mathbf{s}_i$  contains the coordinates for the  $i^{th}$  spatial site location and  $n_s$  is the number of spatial locations. Then, a spatial model for  $\mathbf{y}_s(\mathbf{s}_i)$ , a vector of the  $Y_s(\mathbf{s}_i)$ , is

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

where  $\mathbf{X}_s$  is a design matrix for the fixed effects and  $\boldsymbol{\beta}_s$  is a parameter vector of fixed effects. The error  $\boldsymbol{\epsilon}_s(\mathbf{s}_i)$  can be decomposed into spatial error and independent error components:

$$\boldsymbol{\epsilon}_s(\mathbf{s}_i) = \mathbf{Z}_s \boldsymbol{\delta} + \mathbf{Z}_s \boldsymbol{\gamma}. \quad (2)$$

In equation 2,  $\mathbf{Z}_s$  is an  $n_s \times n_s$  matrix of 0's and 1's, where the values in a row corresponding to a data point at location  $\mathbf{s}_i$  are 1 in the  $i^{th}$  column and a 0 in all other columns. Note that, without temporal replication,  $\mathbf{Z}_s$  is the identity matrix so is not necessary to include in equation 2.  $\boldsymbol{\delta}$  is a random vector independent of  $\boldsymbol{\gamma}$  with

mean  $\mathbf{0}$  and covariance  $\text{cov}(\boldsymbol{\delta}) = \sigma_\delta^2 \mathbf{R}_s$ , where  $\mathbf{R}_s$  is a spatial correlation matrix and  $\sigma_\delta^2$  is sometimes called the spatial partial sill.  $\boldsymbol{\gamma}$  is also a random vector with mean  $\mathbf{0}$  but has covariance  $\text{cov}(\boldsymbol{\gamma}) = \sigma_\gamma^2 \mathbf{I}_s$ , where  $\mathbf{I}_s$  is the  $n_s \times n_s$  identity matrix and  $\sigma_\gamma^2$  is sometimes called the spatial nugget.

There are many common parameterizations of  $\mathbf{R}_s$ . One common assumption is to assume the covariance function generating  $\mathbf{R}_s$  is stationary and isotropic, depending only on the spatial distance between the data points. For example, the exponential covariance function is defined as follows. For observations at locations  $i$  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  $\phi$  is a spatial range parameter controlling the decay rate of the covariance as distance between two data points increases.

## 2.2. Temporal Model

Next, we consider a temporal linear model for a response variable  $Y_t(t_j)$ ,  $j = 1, 2, \dots, n_t$ , where  $t_j$  contains the time index for the  $j^{th}$  time point and  $n_t$  is the number of time points in the data. Then, a temporal model for  $\mathbf{y}_t(t_j)$ , a vector of the  $Y_t(t_j)$ , is

$$\mathbf{y}_t(t_j) = \mathbf{X}_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t(t_j), \quad (4)$$

where  $\mathbf{X}_t$  is a design matrix for the fixed effects and  $\boldsymbol{\beta}_t$  is a parameter vector of fixed effects. The error  $\boldsymbol{\epsilon}_t(t_j)$  can be decomposed into temporal error and independent error components:

$$\boldsymbol{\epsilon}_t(t_j) = \mathbf{Z}_t \boldsymbol{\tau} + \mathbf{Z}_t \boldsymbol{\eta}. \quad (5)$$

In equation 5,  $\mathbf{Z}_t$  is an  $n_t \times n_t$  matrix of 0's and 1's, where the values in a row corresponding to a data point at time point  $t_j$  are a 1 in the  $j^{th}$  column and a 0 in all other columns. Note that, without spatial replication,  $\mathbf{Z}_t$  is the identity matrix so is not necessary to include in equation 5.  $\boldsymbol{\tau}$  is a random vector independent of  $\boldsymbol{\eta}$  with mean  $\mathbf{0}$  and covariance  $\text{cov}(\boldsymbol{\tau}) = \sigma_\tau^2 \mathbf{R}_t$ , where  $\mathbf{R}_t$  is a temporal correlation matrix and  $\sigma_\tau^2$  is sometimes called the temporal partial sill.  $\boldsymbol{\eta}$  is 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  $n_t \times n_t$  identity matrix and  $\sigma_\eta^2$  is sometimes called the temporal nugget.

There are many common parameterizations of  $\mathbf{R}_t$ . One common assumption is to assume the covariance function generating  $\mathbf{R}_t$  is stationary, depending only on the temporal distance between the data points. 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

$$\exp(-m_{jj'}/\rho), \quad (6)$$

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 is greater than zero (Schabenberger and Gotway 2017).

### 2.3. Spatiotemporal Model

We now combine the spatial error components and temporal error components to formulate a model for data collected across both space and time. Let  $Y(\mathbf{s}_i, t_j)$ ,  $i = 1, 2, \dots, n_s$  and  $j = 1, 2, \dots, n_t$ , be a random variable, where  $\mathbf{s}_i$  and  $n_s$  are defined in subsection 2.1 and  $t_j$  and  $n_t$  are defined in subsection 2.2. If each spatial location is represented at every 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 spatiotemporal model for  $\mathbf{y}(\mathbf{s}_i, t_j)$  is

$$\mathbf{y}(\mathbf{s}_i, t_j) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}(\mathbf{s}_i, t_j), \quad (7)$$

where  $\mathbf{X}$  is a design matrix for the fixed effects and  $\boldsymbol{\beta}$  is a parameter vector of fixed effects. The error  $\boldsymbol{\epsilon}(\mathbf{s}_i, t_j)$  can be decomposed into spatial and temporal components, as in Dumelle et al. (2021). A simple model incorporates spatial error and temporal error in  $\boldsymbol{\epsilon}(\mathbf{s}_i, t_j)$  by summing the spatial and temporal errors defined in equation 2 and equation 5:

$$\boldsymbol{\epsilon}(\mathbf{s}_i, t_j) = \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta}. \quad (8)$$

With spatial and temporal replication,  $\mathbf{Z}_s$  is an  $N \times n_s$  matrix while  $\mathbf{Z}_t$  is an  $N \times n_t$  matrix. However, even when the spatial covariance function generating  $\mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma}$  and the temporal covariance function generating  $\mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta}$  are strictly positive definite, the sum of the spatial and temporal components is not necessarily strictly positive definite (Myers and Journel (1990)).

A more flexible option that is always positive definite, given in Dumelle et al. (2021), is the product-sum linear mixed model:

$$\mathbf{y}(\mathbf{s}_i, t_j) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}. \quad (9)$$

In equation 9,  $\boldsymbol{\omega}$  is a random vector of length  $N$  with mean  $\mathbf{0}$  and covariance  $\text{cov}(\boldsymbol{\omega}) = \sigma_\omega^2 \mathbf{R}_{st}$  where  $\mathbf{R}_{st}$  is a spatiotemporal correlation matrix and  $\sigma_\omega^2$  is sometimes called the spatiotemporal partial sill.  $\boldsymbol{\nu}$  is also a random vector of length  $n$  with mean  $\mathbf{0}$  but 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 sometimes called the spatiotemporal nugget.

In the product-sum linear mixed model, the formulation for  $\mathbf{R}_{st}$  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.

If we assume that  $\boldsymbol{\delta}$ ,  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\tau}$ ,  $\boldsymbol{\eta}$ ,  $\boldsymbol{\omega}$ , and  $\boldsymbol{\nu}$  are mutually independent of each other, then

$$\text{var}(\mathbf{y}) \equiv \boldsymbol{\Sigma} = \sigma_\delta^2 \mathbf{Z}_s \mathbf{R}_s \mathbf{Z}'_s + \sigma_\gamma^2 \mathbf{Z}_s \mathbf{I}_s \mathbf{Z}'_s + \sigma_\tau^2 \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}'_t + \sigma_\eta^2 \mathbf{Z}_t \mathbf{I}_t \mathbf{Z}'_t + \sigma_\omega^2 \mathbf{R}_{st} + \sigma_\nu^2 \mathbf{I}_{st}. \quad (10)$$

Note that the model in equation 9 does not have any distributional assumptions: we only need to specify the mean and variance of  $\mathbf{y}$ . However, if we also assume that  $\mathbf{y}$  is multivariate normal (with mean  $\mathbf{X}\boldsymbol{\beta} \equiv \boldsymbol{\mu}$  and variance  $\boldsymbol{\Sigma}$  (Equation 10)), then all model parameters can be easily estimated with Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML).

## 2.4. Finite Population Kriging

The model in equation 9 is for the  $N$ -length vector  $\mathbf{y}$ . However, often we do not have the resources to sample or observe every spatial site in every year. Therefore, we may have an interest in prediction of the response values on sites that were not observed. Throughout this section, let the subscript  $o$  denote data points that were surveyed (both past and present), the subscript  $u$  denote data points that were not surveyed, and the subscript  $a$  denote all observations. Then, we can re-order the response vector so that

$$\mathbf{y}_a = [\mathbf{y}'_u, \mathbf{y}'_o]'. \quad (11)$$

Our primary goal is to use the model developed in equation 9 to find optimal weights  $\mathbf{q}'$  to apply to the observed realizations of  $\mathbf{y}_o$  such that  $\mathbf{q}'\mathbf{y}_o$  is the Best Linear Unbiased Predictor (BLUP) for  $\mathbf{b}'_a\mathbf{y}_a$ . For example, if we are interested in the total of the response across all years, then  $\mathbf{b}_a$  would be a column vector of 1's, so that we are adding up all values of the response for a predictor of total abundance across all spatial sites and time points.

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 design matrix for surveyed data points,  $\mathbf{q}'\mathbf{X}_o\beta = \mathbf{b}'\mathbf{X}\beta$  for every  $\beta$ , implying that  $\mathbf{q}'\mathbf{X}_o = \mathbf{b}'_a\mathbf{X}_a$ .

Kriging weights are then found by finding  $\boldsymbol{\lambda}_o$ , an  $n_o \times 1$  vector, such that

$$E\{(\mathbf{q}'\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)(\mathbf{q}'\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)\} - E\{(\boldsymbol{\lambda}'_o\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)(\boldsymbol{\lambda}'_o\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)\} \quad (12)$$

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

$$\begin{pmatrix} \Sigma_{o,o} & \mathbf{X}_o \\ \mathbf{X}'_o & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ m \end{pmatrix} = \begin{pmatrix} \Sigma_{o,o} & \Sigma_{o,u} \\ \mathbf{X}'_o & \mathbf{X}'_u \end{pmatrix} \begin{pmatrix} \mathbf{b}_o \\ \mathbf{b}_u \end{pmatrix}, \quad (13)$$

where again the subscripts  $o$  and  $u$  denote observed and unobserved data points. For example, letting  $n_o$  denote the number of observed data points,  $\Sigma_{o,o}$  denotes the  $n_o \times n_o$  submatrix of  $\Sigma$  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. Then, the optimal prediction weights are

$$\boldsymbol{\lambda}'_o = \mathbf{b}'_o + \mathbf{b}'_u (\Sigma_{u,o} \Sigma_{o,o}^{-1}) - \mathbf{b}'_u (\Sigma_{u,o} \Sigma_{o,o}^{-1}) \mathbf{X}_o (\mathbf{X}'_o \Sigma_{o,o}^{-1} \mathbf{X}_o)^{-1} \mathbf{X}'_o \Sigma_{o,o}^{-1} + \mathbf{b}'_u \mathbf{X}'_u (\mathbf{X}'_o \Sigma_{o,o}^{-1} \mathbf{X}_o)^{-1} \mathbf{X}_o \Sigma_{o,o}^{-1}. \quad (14)$$

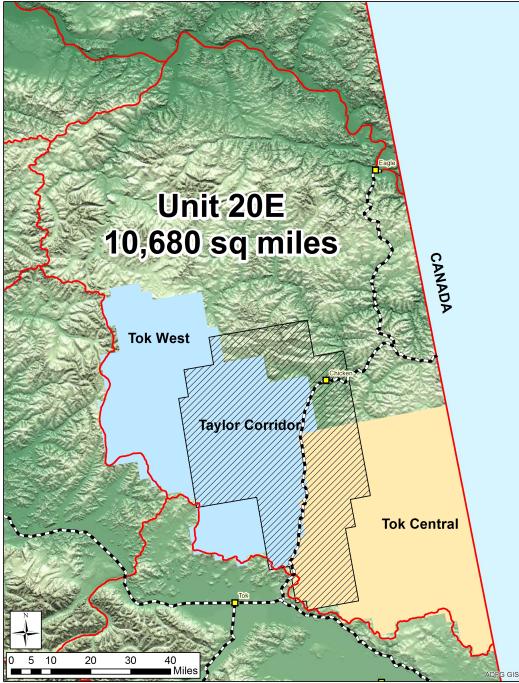
The BLUP for  $\mathbf{b}'_a\mathbf{y}_a$  is

$$\widehat{\mathbf{b}'_a\mathbf{y}_a} = \boldsymbol{\lambda}'_o \mathbf{y}_o, \quad (15)$$

with a prediction variance of

$$E((\boldsymbol{\lambda}'_o \mathbf{y}_o - \mathbf{b}'_a \mathbf{y}_a)(\boldsymbol{\lambda}'_o \mathbf{y}_o - \mathbf{b}'_a \mathbf{y}_a)) = \boldsymbol{\lambda}'_o \Sigma_{o,o} \boldsymbol{\lambda}_o - 2\mathbf{b}'_a \Sigma_{a,o} \boldsymbol{\lambda}_o + \mathbf{b}'_a \Sigma_{a,a} \mathbf{b}_a. \quad (16)$$

A common predictor of interest is the total abundance in the most recent time point of the survey. Then,  $\mathbf{b}_a$  is a vector of 1's and 0's, where the  $k^{th}$  element of  $\mathbf{b}_a$  is a 1



**Figure 1.** A map of the Taylor Corridor in the TOK region of Alaska.

if the  $k^{th}$  element of  $\mathbf{y}_a$  is from the most recent time point of the survey and the  $k^{th}$  element of  $\mathbf{b}_a$  is a 0 otherwise. If we order  $\mathbf{y}_a$  by (1) the unobserved, past data points, (2) the unobserved, current data points, (3) the observed, past data points, and (4) the observed, current data points, then

$$\mathbf{b}_a = [\mathbf{b}'_{up}, \mathbf{b}'_{uc}, \mathbf{b}'_{op}, \mathbf{b}'_{oc}]' = [\mathbf{0}', \mathbf{1}', \mathbf{0}', \mathbf{1}']', \quad (17)$$

where the subscripts  $up$ ,  $uc$ ,  $op$ , and  $oc$  denote unobserved sites in past years, unobserved sites in current years, observed sites in past years, and observed sites in current years, respectively.

$\boldsymbol{\lambda}_o$  can then be rewritten as

$$\boldsymbol{\lambda}'_o = \mathbf{b}'_o + \mathbf{b}'_{uc}(\Sigma_{uc,o}\Sigma_{o,o}^{-1}) - \mathbf{b}'_{uc}(\Sigma_{uc,o}\Sigma_{o,o}^{-1})\mathbf{X}_o(\mathbf{X}'_o\Sigma_{o,o}^{-1}\mathbf{X}_o)^{-1}\mathbf{X}'_o\Sigma_{o,o}^{-1} + \mathbf{b}'_{uc}\mathbf{X}'_{uc}(\mathbf{X}'_o\Sigma_{o,o}^{-1}\mathbf{X}_o)^{-1}\mathbf{X}_o\Sigma_{o,o}^{-1}. \quad (18)$$

with a prediction variance of

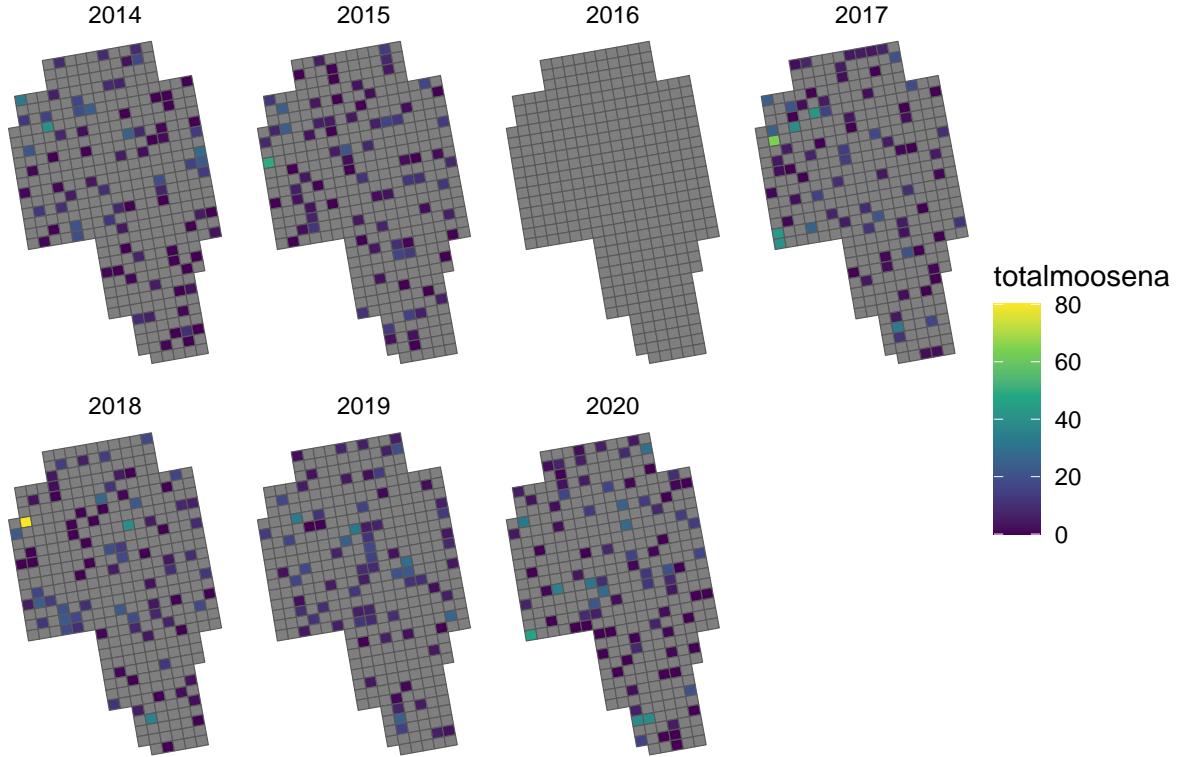
$$\boldsymbol{\lambda}'_o\Sigma_{o,o}\boldsymbol{\lambda}_o - 2\mathbf{b}'_c\Sigma_{c,o}\boldsymbol{\lambda}_o + \mathbf{b}'_c\Sigma_{c,c}\mathbf{b}_c, \quad (19)$$

where  $c$  denotes observations in the most current time point.

### 3. Application

#### 3.1. Data Description

Abundance surveys are performed in the Taylor Corridor of the TOK region of Alaska annually (Figure 1). In particular, surveys were conducted every year from 2014



**Figure 2.** Layout of the spatial sites used to survey moose in the Taylor corridor of the TOK region of Alaska.

through 2020, except for the year 2016, during which there was not sufficient snow cover to perform a survey. There are a total of 381 unique spatial locations, which we refer to as “sites,” and a total of 7 unique time points in the data set, including the missing year of 2016.

In each year of the survey, an aerial team of biologists selects some of the 381 sites to survey. The number of sites in the sampling frame that are selected varies from a low of 76 in the year 2019 to a high of 90 in the year 2020. Throughout the 7 unique time points, some sites are surveyed as many as four or five different times while others are never surveyed (Figure 2).

Before the survey begins in each year, biologists stratify the sites into a “HIGH” stratum and a “LOW” stratum. There are 230 sites in the “HIGH” stratum while there are 151 sites in the “LOW” stratum. The goal of the following analysis is to predict the total abundance of moose across all spatial sites in the year 2020, the most recent year of the survey.

### 3.2. Model Fitting

We fit the product-sum covariance model defined in equation 9 using REML, with stratum as a covariate in the design matrix, an exponential spatial correlation structure defined in 3, and an exponential temporal correlation structure defined in 6. Table ?? 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 spatiotemporal dependent error variance and spatiotemporal independent error variance.

Spatial			Temporal			Spatiotemporal	
$\hat{\sigma}_\delta^2$	$\hat{\sigma}_\gamma^2$	$\hat{\phi}$	$\hat{\sigma}_\tau^2$	$\hat{\sigma}_\eta^2$	$\hat{\rho}$	$\hat{\sigma}_\omega^2$	$\hat{\sigma}_\nu^2$
16.91	3.76	4.44	0.88	0.25	2.29	24.01	30.8

To help interpret what some of these fitted covariance parameter estimates mean, we can construct a fitted covariance plot (Figure 3). Note that the centroids of two sites directly adjacent to one another are about 4 units apart. As the spatial distance between two sites increases (dark colour to light colour), the covariance between the response values 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 units apart, no matter what the temporal distance is. Note that the fact that the covariance is larger than 0 when the temporal distance is 6 and the spatial distance is 0 implies that including surveys before 2014 could improve precision of the predictor for the total abundance in 2020 even more.

The estimated vector of fixed effects, using "HIGH" as the reference group, is  $\beta = (11.26, -9.76)$ . Therefore the overall mean for sites in the "HIGH" stratum is estimated to be 11.26 moose while the overall mean for sites in the "LOW" stratum is estimated to be 1.5 moose.

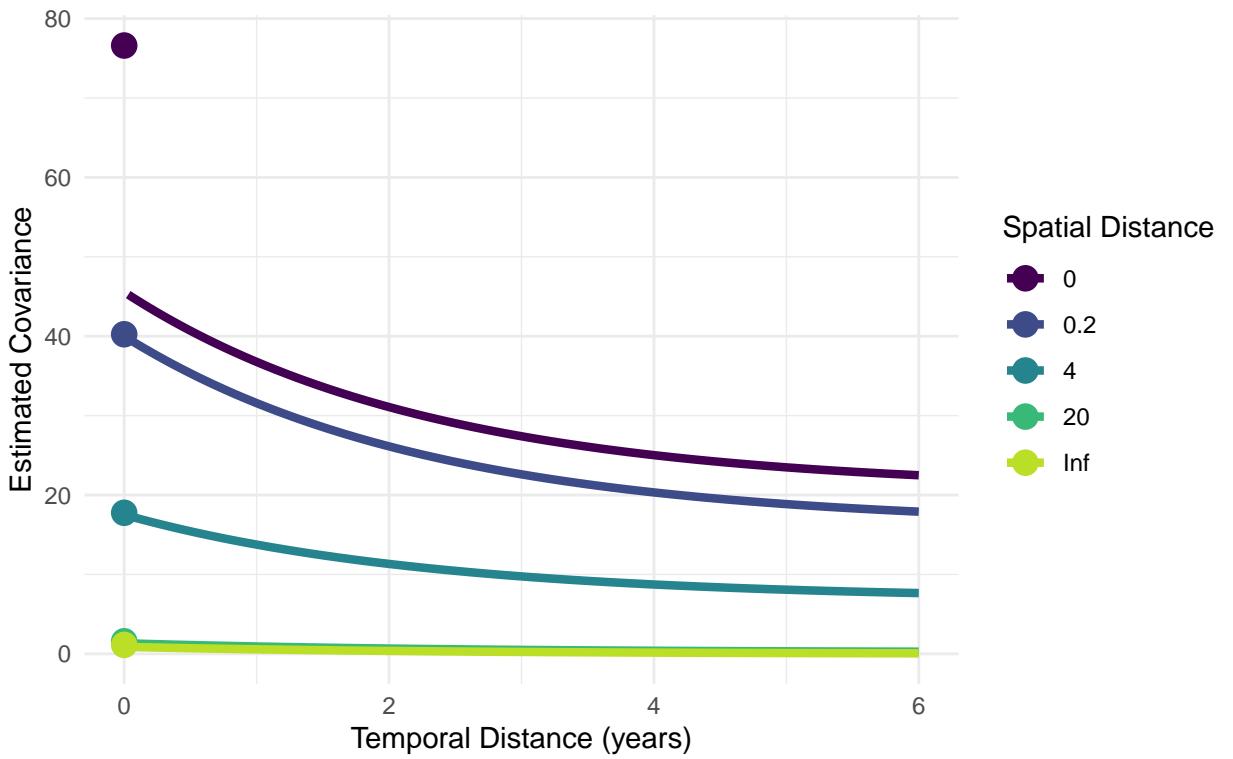
### 3.3. Prediction

We now use the model in subsection 3.2 to predict the total abundance across all sites in the year 2020, the most recent year of the survey. Plugging in estimates of the covariance parameters into equations 18 and 19 and letting elements of  $\mathbf{b}_a$  be 1's for data points in 2020 and 0's otherwise, we obtain a prediction of 2874 moose and a standard error of 234 moose. A 90% normal-based prediction interval for the total abundance in 2020 is (2489, 3259) moose. Sitewise predictions for sites in 2020 are given in the map in Figure 4.

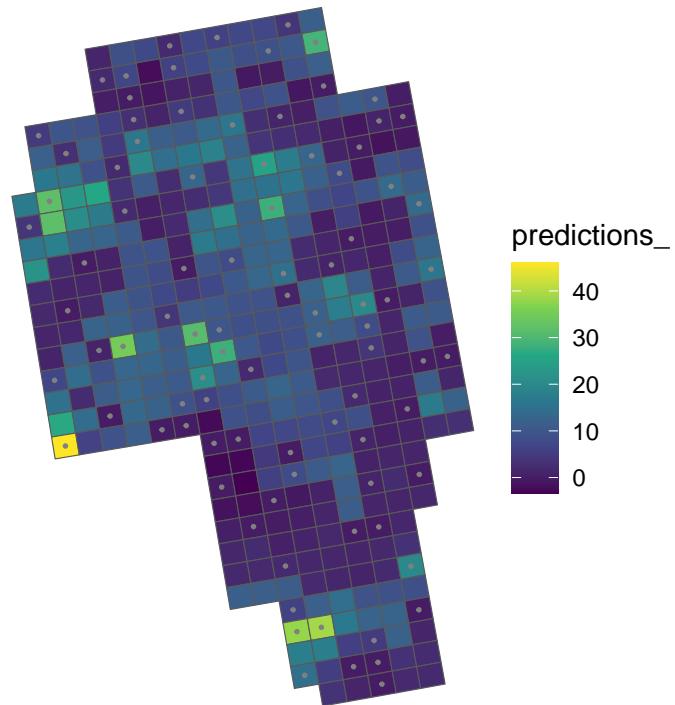
For comparison, we use the spatial `sptotal` package (Higham et al. 2021a) to compute the prediction for the total abundance of moose in the year 2020 with stratum as a covariate (Ver Hoef 2008). We also use the standard stratified random sampling design-based estimator

$$N_h \bar{y}_h + N_l \bar{y}_l$$

where  $\bar{y}_i$  is the sample mean for the data points in 2020 in the  $i^{th}$  stratum,  $N_i$  is the total number of sites in 2020 in the  $i^{th}$  stratum, and  $n_i$  is the number of observed data points in 2020 in the  $i^{th}$  stratum. The stratified random sampling design-based estimator has a variance for the total abundance of  $N_h^2 \cdot (1 - \frac{n_h}{N_h}) \cdot \frac{s_h^2}{n_h} + N_l^2 \cdot (1 - \frac{n_l}{N_l}) \cdot \frac{s_l^2}{n_l}$ , where  $s_i^2$  is the sample variance of the 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 **only** use data from 2020.



**Figure 3.** Estimated covariance of the errors from the fitted parameters in a product-sum model.



**Figure 4.** A map of the predictions for the sites in the year 2020. A site with a grey dot in the center means that the site was sampled in 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 simple random sampling 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 are somewhat similar across the three methods, we see that the spatiotemporal model is most efficient ( $SE = 234$  moose compared to 319 moose for the purely spatial model and 371 moose for the simple random sampling estimator that ignores both spatial and temporal information).

## 4. Simulation

### 4.1. Description

To evaluate performance of the finite population spatiotemporal model, we conduct a small simulation study. We simulate a response vector  $\mathbf{y}$  of length  $N = 1000$  on a  $10 \times 10$  grid of 100 spatial sites on the unit ( $[0, 1] \times [0, 1]$ ) square 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 10 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 spatiotemporal dependent error variance and spatiotemporal independent error variance.

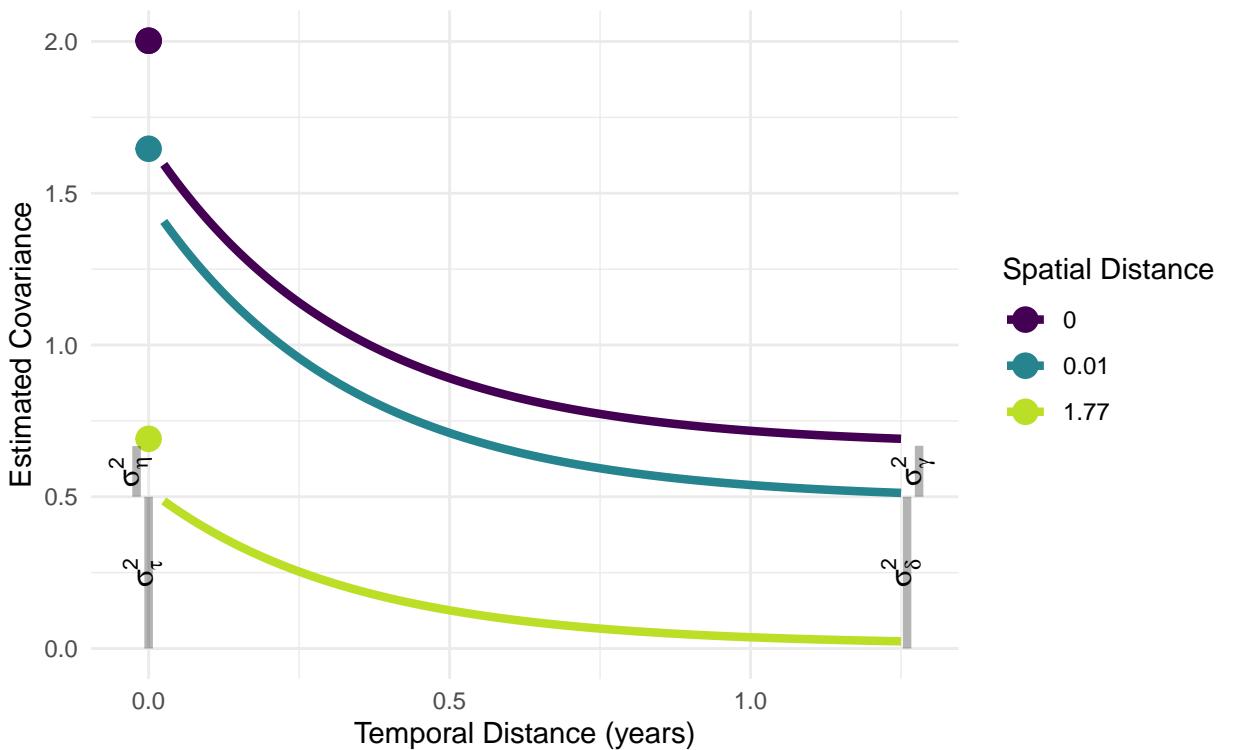
scenario	Spatial			Temporal			Spatiotemporal	
	$\sigma_\delta^2$	$\sigma_\gamma^2$	$\phi$	$\sigma_\tau^2$	$\sigma_\eta^2$	$\rho$	$\sigma_\omega^2$	$\sigma_\nu^2$
spatiotemporal	0.5	0.17	0.47	0.5	0.17	0.33	0.50	0.17
temporal iev	0	0	0.47	0	1.50	0	0.25	0.25
spatial	0	0	0.47	0	0	0	1.50	0.50
independent	0	0	0	0	0	0	0	2.00

The three scenarios in the table correspond to (1) **spatiotemporal**: a setting where there is spatiotemporal covariance in the random errors, (2) **spatial**: a setting where there is spatial covariance within a particular time point but errors in different time points are not correlated even when the errors come from the same spatial site, and (3) **independent**: all errors are independent regardless of spatial index and time index. In all scenarios, the total variance (summing all six variance parameters) is equal to 2.

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 6. 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 observations in space ( $\sqrt{2} = 1.414$ ) and the maximum distance between two observations in time (1). A value of 0 for  $\phi$  (or  $\rho$ ) sets the  $\mathbf{R}_s$  matrix (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 spatiotemporal scenario.

Each of these three scenarios is replicated for two different sample sizes:  $n = 250$



**Figure 5.** The model covariance used in the simulations for the spatiotemporal scenario. Covariance is approximately 0 for errors from observations that are  $\sqrt{2}$  distance units apart in space and 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.

and  $n = 500$ . A simple random sample of the 1000 total observations 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. The resulting response 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 model performs when the covariance is misspecified, as the resulting response variable no longer follows a tractable covariance function.

Therefore, the simulation study has 12 total settings coming from a  $3 \times 2 \times 2$  factorial design. For each setting, we simulate 1000 realizations. For each realization, we predict the total response for the “most recent” time point (when the time index is equal to 1 on the interval  $[0, 1]$ ), which we will henceforth call the “current total” using three methods. The first method uses the finite population kriging described in subsection 2.4 with the model covariance in equation 10. The second method is a spatial model fit with the `sptotal` R package (Higham et al. 2021a) that only uses data from the most recent time point. This method corresponds to how moose surveys are often currently analyzed, assuming that abundance is correlated across space in the survey but ignoring data from all previous surveys. Both of the first two methods estimate model parameters with Restricted Maximum Likelihood (REML). The third method uses a simple random sample (SRS) estimator with data from the most recent time point. The SRS estimator for the total is  $\bar{y} \cdot \frac{100}{n_1}$ , where  $\bar{y}$  is the sample mean of the response in the most recent time point and  $n_1$  is the number of sampled locations in the most recent time point. The variance is  $n_1^2 \cdot \frac{\hat{\sigma}^2}{100} \cdot (1 - \frac{n_1}{100})$ , where  $\hat{\sigma}^2$  is the sample variance of the response variable in the most recent time point.

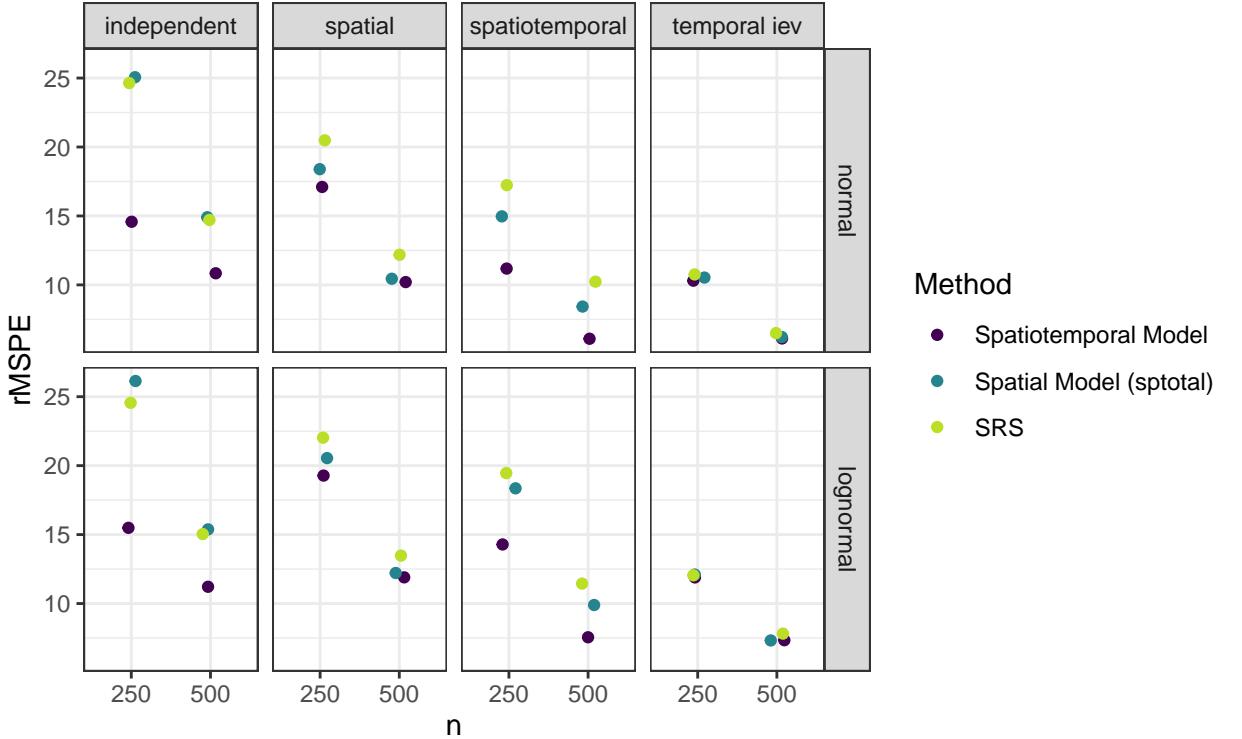
Note that the SRS method gives an estimator (not a predictor) and corresponding confidence interval (not a prediction interval) because the estimator treats the observed data as fixed, not as a random realization from a process (Brus 2021; Dumelle et al. 2022). However, in the remaining text and tables,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

For each method, we record the root mean squared prediction error (rMSPE),  $\sqrt{(\sum_{i=1}^{1000} (\hat{T}_i - T_i)^2)/1000}$ , where  $\hat{T}_i$  and  $T_i$  are the predicted and realized current totals, respectively, in the  $i^{th}$  iteration. We also create a normal-based 90% prediction interval for the realized current total and record both the average prediction interval length and the proportion of iterations that the prediction interval covers the realized current total.

## 4.2. Results

Tables 4, 5, and 6 in Section 6 give the rMSPE, bias and interval coverage of the three predictors in all 12 simulation settings. In Figure 6, we see that the spatiotemporal model outperforms both the purely spatial model and the simple random sample design-based estimator in all 12 settings. In general, rMSPE improvement is larger for the smaller sample size. In general, as  $n$  approaches  $N$ , rMSPE for all approaches should become smaller and be exactly equal to 0 when  $n = N$ .

We also see that, in general, there are moderate gains in rMSPE for the spatiotem-



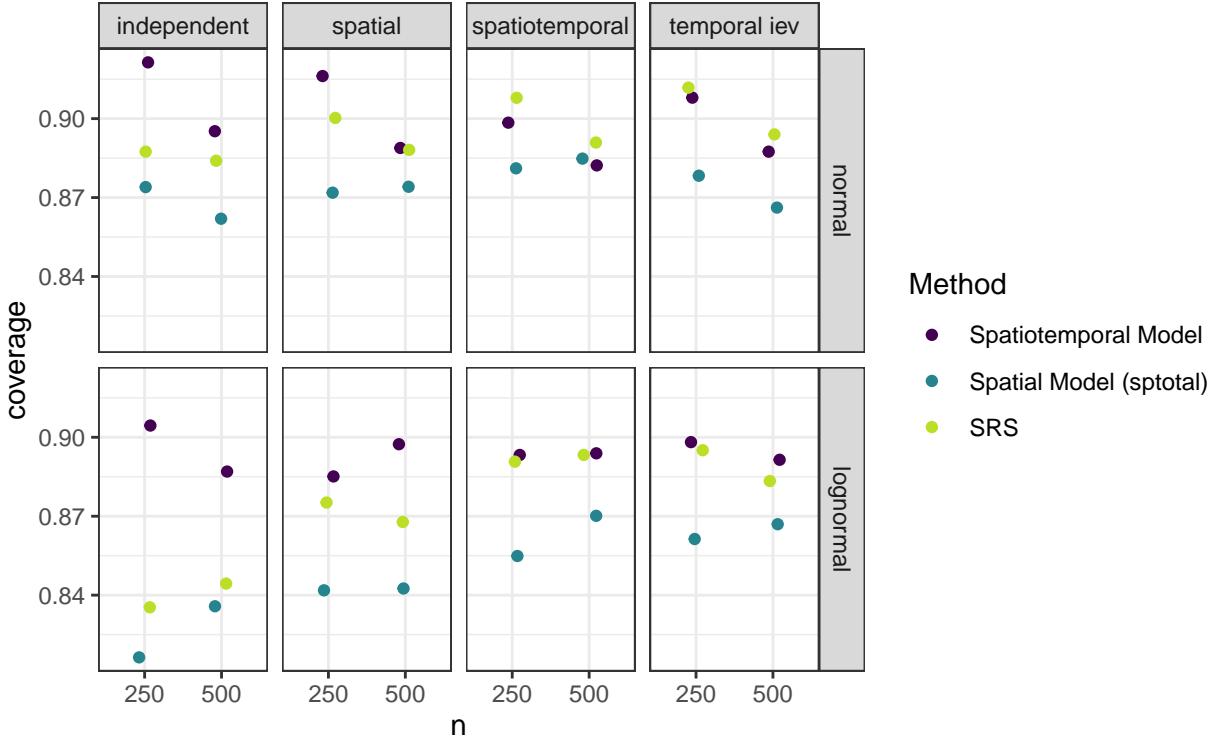
**Figure 6.** root Mean-Squared-Prediction-Error for all simulation settings. The spatiotemporal model has the smallest rMSPE in all settings tested, though it is similar to the rMSPE of the other two methods in the spatial scenario, where response values within a time point are correlated across space but are uncorrelated with all response values from other time points.

poral model are made in the “spatial” simulation setting. In this setting, response values across space are only correlated within a given time point; therefore, the spatiotemporal model only does marginally better than the other two approaches because the data from the other time points are not correlated with the data in the most current time point. However, in both the independent setting and the spatial setting, the spatiotemporal model still outperforms the purely spatial model and the simple random sample design-based estimator because the spatiotemporal can still use the response values collected in previous time points to estimate the fixed effects structure of the model.

We see the smallest gains in rMSPE for the spatiotemporal model in the “temporal iev” scenario. In this setting, most of the variability in the response comes from  $\sigma_\eta^2$ , the temporal independent error variance. Therefore, data collected in different time points is not correlated, and, different time points can have very different realized totals. As expected, the spatiotemporal model performs no better than a purely spatial model for this scenario, but we can also say that the added complexity of the spatiotemporal model does not make it perform worse than the purely spatial model.

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 4.

Figure 7 shows the interval coverage for the normal-based prediction intervals, where the nominal level is 0.90. We see that the spatiotemporal model predictor for the current total has approximate 90% coverage in all settings tested. The spatial model



**Figure 7.** Prediction interval coverage for all simulation settings, where the prediction intervals are normal-based and the nominal level is 0.90. The predictor from the spatiotemporal model has close to appropriate coverage in all settings tested.

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

## 5. Discussion

- substantial reduction of se in the application (and, presumably, the simulations).
  - if there is a very high amount of spatial correlation but a low amount of temporal correlation (as in the second scenario), then gains are minimal.
- normal-based-related limitations
- Bayesian approach, and its drawbacks benefits to bayes: - models counts as counts (better for predicting on one particular site) - model more expandable (lots of zeroes eg.) benefits to ours: - much quicker - assessed via a small simulation study - easier for practitioner to fit (especially given software - no convergence checks, priors to be tweaked, etc.)

An additional possible benefit of the spatiotemporal model compared to a purely spatial model is the potential for forecasting abundance before a survey is done. 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 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 biologists. We might also consider using such an approach if there is a year during which a survey cannot be completed for logistical reasons. For example, in the TOC region of Alaska, a moose survey was not conducted at all in the year 2016 because there was insufficient snow cover for the survey. The spatiotemporal could still be applied to get a prediction for moose abundance using survey data from previous years.

**Table 3.** Results from analysis on the TOC 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

There is a substantial drop in efficiency when 2020 is not surveyed, but some monitoring programs may deem this worthy, or, might find this useful if a survey cannot be done one year because of poor conditions, changes in personnel, etc. (and it might drop more or not as much depending how the temporal correlation structure).

- take-home message: monitoring programs that use regularly-scheduled surveys might consider incorporating time into their analysis to improve precision of predictors for the mean or total.

## 6. Appendix

**Table 4.** root Mean Squared Prediction Error (rMSPE) for the spatiotemporal model, the spatial model, and the simple random sample estimator for each of the 12 simulation settings. In all settings, the rMSPE for the spatiotemporal model is lower than the rMSPE for the other two methods.

Simulation Setting			rMSPE		
scenario	n	Response Type	spatiotemporal	spatial	srs
independent	250	normal	14.58	25.06	24.64
	250	normal	17.10	18.39	20.49
	250	normal	10.31	10.53	10.76
	250	normal	11.18	14.97	17.23
spatial	500	normal	10.84	14.91	14.71
	500	normal	10.21	10.45	12.19
	500	normal	6.12	6.22	6.50
	500	normal	6.09	8.43	10.24
temporal iev	250	lognormal	15.49	26.14	24.56
	250	lognormal	19.28	20.55	22.03
	250	lognormal	11.89	12.10	12.05
	250	lognormal	14.28	18.35	19.46
spatiotemporal	500	lognormal	11.22	15.38	15.04
	500	lognormal	11.90	12.21	13.48
	500	lognormal	7.34	7.32	7.82
	500	lognormal	7.55	9.89	11.45

**Table 5.** Bias (Realized Current Total - Predicted Current Total) for the spatiotemporal model, the spatial model, and the simple random sample estimator for each of the 12 simulation settings. In all settings, all methods appear fairly unbiased.

Simulation Setting			Bias		
scenario	n	Response Type	spatiotemporal	spatial	srs
independent	250	normal	0.73	1.38	1.55
	250	normal	0.57	0.52	0.76
	250	normal	0.44	0.39	0.47
	250	normal	0.48	0.27	0.45
spatial	500	normal	0.46	0.60	0.67
	500	normal	0.22	0.18	0.04
	500	normal	0.15	0.14	0.07
	500	normal	0.04	0.07	0.04
temporal iev	250	lognormal	0.36	0.56	1.48
	250	lognormal	0.01	-0.19	0.26
	250	lognormal	0.33	0.22	0.41
	250	lognormal	-0.07	-0.85	-0.49
spatiotemporal	500	lognormal	0.32	0.29	0.66
	500	lognormal	0.16	-0.13	0.08
	500	lognormal	0.24	0.15	0.08
	500	lognormal	-0.10	-0.39	-0.37

**Table 6.** Prediction interval coverage for the spatiotemporal model, the spatial model, and the simple random sample estimator 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	spatiotemporal	spatial	srs
independent	250	normal	0.92	0.87	0.89
	250	normal	0.92	0.87	0.90
	250	normal	0.91	0.88	0.91
	250	normal	0.90	0.88	0.91
spatial	500	normal	0.90	0.86	0.88
	500	normal	0.89	0.87	0.89
	500	normal	0.89	0.87	0.89
	500	normal	0.88	0.89	0.89
temporal iev	250	lognormal	0.90	0.82	0.84
	250	lognormal	0.89	0.84	0.88
	250	lognormal	0.90	0.86	0.90
	250	lognormal	0.89	0.86	0.89
spatiotemporal	500	lognormal	0.89	0.84	0.84
	500	lognormal	0.90	0.84	0.87
	500	lognormal	0.89	0.87	0.88
	500	lognormal	0.89	0.87	0.89

## References

- Boertje, Rodney D, Mark A Keech, Donald D Young, Kalin A Kellie, and C Tom Seaton. 2009. “Managing for elevated yield of moose in interior Alaska.” *The Journal of Wildlife Management* 73 (3): 314–327.
- Brus, Dick J. 2021. “Statistical approaches for spatial sample survey: Persistent misconceptions and new developments.” *European Journal of Soil Science* 72 (2): 686–703.
- Dumelle, Michael, Matt Higham, Jay M Ver Hoef, Anthony R Olsen, and Lisa Madsen. 2022. “A comparison of design-based and model-based approaches for finite population spatial sampling and inference.” *Methods in Ecology and Evolution* 13 (9): 2018–2029.
- Dumelle, Michael, Jay M Ver Hoef, Claudio Fuentes, and Alix Gitelman. 2021. “A linear mixed model formulation for spatio-temporal random processes with computational advances for the product, sum, and product–sum covariance functions.” *Spatial Statistics* 43: 100510.
- Gasaway, William C, Stephen D DuBois, Daniel J Reed, and Samuel J Harbo. 1986. *Estimating moose population parameters from aerial surveys*. Technical Report. University of Alaska Institute of Arctic Biology.
- Higham, Matt, Jay Ver Hoef, Bryce Frank, and Michael Dumelle. 2021a. *sptotal: Predicting Totals and Weighted Sums from Spatial Data*. R package version 1.0.0, <https://highamm.github.io/sptotal/index.html>.
- Higham, Matt, Jay Ver Hoef, Lisa Madsen, and Andy Aderman. 2021b. “Adjusting a finite population block kriging estimator for imperfect detection.” *Environmetrics* 32 (1): e2654.
- Kellie, Kalin A, and Robert A DeLong. 2006. *Geospatial survey operations manual*. Technical Report. Alaska Department of Fish and Game.
- Myers, Donald E, and Andre Journel. 1990. “Variograms with zonal anisotropies and nonin-

- vertible kriging systems.” *Mathematical Geology* 22 (7): 779–785.
- Peters, Wibke, Mark Hebblewhite, Kirby G Smith, Shevenell M Webb, Nathan Webb, Mike Russell, Curtis Stambaugh, and Robert B Anderson. 2014. “Contrasting aerial moose population estimation methods and evaluating sightability in west-central Alberta, Canada.” *Wildlife Society Bulletin* 38 (3): 639–649.
- Schabenberger, Oliver, and Carol A Gotway. 2017. *Statistical methods for spatial data analysis: Texts in statistical science*. Chapman and Hall/CRC.
- Ver Hoef, Jay M. 2008. “Spatial methods for plot-based sampling of wildlife populations.” *Environmental and Ecological Statistics* 15 (1): 3–13.