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

Received: date / Accepted: date

Abstract

Spatio-temporal models can be used to analyze data collected at various spatial locations throughout multiple time points. However, even with a finite number of spatial locations, there may be insufficient resources to collect data from every spatial location at every time point. We develop a spatio-temporal finite-population block kriging (ST-FPBK) method to predict a quantity of interest, such as a mean or total, across a finite number of spatial locations. This ST-FPBK predictor 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. We also show how the model can be used to forecast a prediction for abundance in a time point for which spatial locations have not yet been surveyed. 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. 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.

1 Introduction

1.1 Background

Spatio-temporal data are indexed by both a spatial index, which we will refer to as a "site," and by a temporal index, which we will refer to as a "time point." Common examples of spatio-temporal data include infections from a disease in a region collected over a time period (e.g. Martínez-Beneito, López-Quilez, and Botella-Rocamora 2008; Sahu and Böhning 2022) or climate variables that are recorded through time at multiple locations (Lemos and Sansó 2009). Models for spatio-temporal data can incorporate rich dependence structures and have

applications in a wide variety of scientific fields (Cressie and Wikle 2015; Montero, Fernández-Avilés, and Mateu 2015; Wikle, Zammit-Mangion, and Cressie 2019; Chen, Genton, and Sun 2021; Porcu, Furrer, and Nychka 2021).

One such application is ecological monitoring of a particular resource, such as animal or plant counts, rainfall, concentration of a compound in soil samples, etc. A common goal of ecological monitoring is to predict the total abundance of a resource at the most recent time point using data collected in subsets of the region at multiple time points. Often, the region of interest is divided into a finite number of areal sampling units and data is collected everywhere within a subset of these areal sampling units at each time point. The remaining areal units are then left unsampled. We refer to the goal of predicting abundance in the most recent time point as *spatio-temporal abundance prediction*.

Classical sampling techniques can be used for spatio-temporal abundance prediction: inference in classical sampling is based on the mechanism in which units are randomly selected to be in the sample. For example, Wang and Zhu (2019) use a spatio-temporally balanced sampling design to construct estimators for finite quantities, like abundance. Panel sampling designs can also be used when data are sampled over multiple time points (Urquhart 2012). However, model-assisted techniques and purely model-based techniques can provide increased precision in predicting finite quantities, at the cost of placing additional assumptions on the model generating the observed data (Breidt and Opsomer 2017; Dumelle et al. 2022).

A few common ways to formally model spatio-temporal covariance include the product model, the sum model, and the product-sum model. If we consider $\text{Cov}(\mathbf{h}_s, h_t)$ to be the covariance between two data points with spatial separation $\mathbf{h}_s$ and temporal separation h_t, then a product model for this covariance (also known as a "separable" model) is $\text{Cov}_s(\mathbf{h}_s)\text{Cov}_t(h_t)$, where $\text{Cov}_s(\mathbf{h}_s)$ only depends on the spatial separation and $\text{Cov}_t(h_t)$ only depends on the temporal separation (Posa 1993; Stein 2005; Gneiting, Genton, and Guttorp 2006). On the other hand, a sum model for $\text{Cov}(\mathbf{h}_s, h_t)$ adds together the spatial and temporal covariance as $\text{Cov}_s(\mathbf{h}_s) + \text{Cov}_t(h_t)$ (Rouhani and Hall 1989). A product-sum model combines the product model and the sum model:

$$\text{Cov}(\mathbf{h}_s, h_t) = w_1\text{Cov}_s(\mathbf{h}_s)\text{Cov}_t(h_t) + w_2\text{Cov}_s(\mathbf{h}_s) + w_3\text{Cov}_t(h_t), \qquad (1)$$

where w_1, w_2, and w_3 are non-negative weights (De Cesare, Myers, and Posa 2001; S. De Iaco, Myers, and Posa 2002). In general, the product-sum model is more flexible than the product model and the sum model.

These spatio-temporal models can be used for abundance prediction in a Bayesian framework as well as in a frequentist framework. Within the Bayesian framework, spatio-temporal modeling is often used as part of a hierarchical model for the response variable: Conn et al. (2015) offer a review of such methods as well as issues with using spatio-temporal modeling for abundance prediction for a few different hierarchical model structures. Bayesian hierarchical models have been used for spatio-temporal abundance prediction by Ver Hoef and Jansen (2007) for harbor seals in Alaska, Sauer and Link (2011) for birds in

North America, Davy, Squires, and Zimmerling (2021) for bats in Canada, Adde et al. (2020) for waterfowl in Canada, and Schmidt et al. (2022) for moose in Alaska. Additionally, Ross, Hooten, and Koons (2012) use Integrated Nested Laplace Approximation (INLA) to model spatio-temporal abundance data in a hierarchical framework. While a Bayesian hierarchical models allow for a large amount of flexibility in the model formulation and are thus powerful tools for statistical inference, Conn et al. (2015) note that "care must be taken to tailor models to match the study population and survey data available." We discuss Bayesian hierarchical models for abundance prediction more in Section 5.

Spatio-temporal predictions can also be constructed in a frequentist framework. As a couple of recent examples, Breivik et al. (2021) use a Gaussian model to predict abundance of cod in the Barents Sea, and Stock et al. (2020) use random forests, generalized additive models, and Gaussian Markov random fields to predict bycatch of a few species of fish. While a prediction for the total abundance is fairly straightforward to obtain in the frequentist setting (by making predictions for the response at each unobserved site and then summing the observed response values with the predictions), obtaining the prediction variance is more challenging. If the number of sites in the region of interest is finite, then the spatio-temporal abundance prediction variance should incorporate a finite population correction such that, if all spatial sites are sampled in the most recent time point, the prediction variance for the total abundance in the most recent time point is equal to 0.

Our primary goal in this paper is to develop such a finite-population-prediction-variance correction. We extend the work of Ver Hoef (2008), who developed Finite Population Block Kriging (FPBK) to predict a linear function of the realized values of a response variable (such as the total abundance) for a survey at one particular time point, incorporating a finite population correction to the variance of the predictor. In this paper, we broaden the approach of Ver Hoef (2008) to the spatio-temporal context, appropriately adjusting the finite-population-prediction-variance correction and allowing for the covariance matrix of the response vector to include spatio-temporal structure.

1.2 Motivating Example

To motivate the development of the predictor in Section 2, we consider moose surveys, which are performed annually or every other year in many regions of Alaska and western Canada. The most common goal of these surveys is to predict moose abundance, the total number of moose, in some finite region to inform harvest regulations (Kellie, Colson, and Reynolds 2019). The region of interest is divided into a finite number of areal polygons. Because of time and money constraints, only some polygons, or sites, in the region of interest are selected to be in the survey at a particular time point. Biologists fly to these selected polygons, count the number of moose, and then use FPBK to find a prediction for the abundance for that year. Note that the interest is in predicting the realized total abundance so that, if there were enough resources to survey every site in a year, we would know the abundance in that year exactly. These

surveys are historically analyzed with the "WinfoNet" site developed by DeLong (2006), which calculates the "GeoSpatial Population Estimator" (GSPE) for a given survey. The GSPE is an application of the FPBK predictor developed by Ver Hoef (2008).

Though many of these surveys are completed regularly, most are analyzed 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 constructs a prediction for total abundance only from counts on sites that were sampled in that year. However, using counts from previous years in a model that incorporates both spatial and temporal (spatio-temporal) correlation while also using a finite population correction factor based on the proportion of sites surveyed in the most recent year could result in a prediction for the realized total that is more precise than predictions from a purely spatial model. Shortly, we describe such a predictor.

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

2 Methods

We now give details on the development of the spatio-temporal model and subsequently use this model to extend the approach of Ver Hoef (2008), developing a finite population correction factor to give a Best-Linear-Unbiased-Predictor (BLUP) and its prediction variance for any linear function of the response vector.

2.1 Spatio-temporal Model

Let $Y(\mathbf{s}_i, t_j)$, $i = 1, 2, \ldots, n_s$ and $j = 1, 2, \ldots, n_t$, be a random variable indexed by a spatial site and a time point, where the vector $\mathbf{s}_i$ contains the coordinates for the i^{th} spatial site, n_s is the number of unique sites, t_j is the time index for the j^{th} time point, and n_t is the number of unique time points. If each site is represented at every time point, a vector of the $Y(\mathbf{s}_i, t_j)$, denoted $\mathbf{y}$, has length $n_s \cdot n_t \equiv N$. Then, a spatio-temporal model for $\mathbf{y}$ is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2}$$

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

$$\boldsymbol{\epsilon} = \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}. \tag{3}$$

In the spatial component of equation 3 ($\mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma}$), the matrix $\mathbf{Z}_s$ is an $N \times n_s$ matrix of 0's and 1's, where the values in a row corresponding to a data point at site $\mathbf{s}_i$ are 1 in the i^{th} column and 0 in all other columns. $\boldsymbol{\delta}$ is a random vector with mean $\mathbf{0}$ and covariance $\text{cov}(\boldsymbol{\delta}) = \sigma^2_\delta\mathbf{R}_s$, where $\mathbf{R}_s$ is an $n_s \times n_s$ spatial correlation matrix and σ^2_δ is called the spatial dependent error variance (or spatial partial sill). The random vector $\boldsymbol{\gamma}$ also has mean $\mathbf{0}$ but has covariance $\text{cov}(\boldsymbol{\gamma}) = \sigma^2_\gamma\mathbf{I}_s$, where $\mathbf{I}_s$ is the $n_s \times n_s$ identity matrix and σ^2_γ is called the spatial independent error variance (or spatial nugget).

In the temporal component of equation 3 ($\mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta}$), $\mathbf{Z}_t$ is an $N \times n_t$ matrix of 0's and 1's, where the values in a row corresponding to a data point at time point t_j are 1 in the j^{th} column and 0 in all other columns. $\boldsymbol{\tau}$ is a random vector with mean $\mathbf{0}$ and covariance $\text{cov}(\boldsymbol{\tau}) = \sigma^2_\tau\mathbf{R}_t$, where $\mathbf{R}_t$ is an $n_t \times n_t$ temporal correlation matrix and σ^2_τ is called the temporal dependent error variance (or temporal partial sill). $\boldsymbol{\eta}$ is also a random vector with mean $\mathbf{0}$ but has covariance $\text{cov}(\boldsymbol{\eta}) = \sigma^2_\eta\mathbf{I}_t$, where $\mathbf{I}_t$ is the $n_t \times n_t$ identity matrix and σ^2_η is called the temporal independent error variance (or temporal nugget).

In the spatio-temporal component of equation 3 ($\boldsymbol{\omega}+\boldsymbol{\nu}$), $\boldsymbol{\omega}$ is a random vector with mean $\mathbf{0}$ and covariance $\text{cov}(\boldsymbol{\omega}) = \sigma^2_\omega\mathbf{R}_{st}$, where $\mathbf{R}_{st}$ is an $N \times N$ spatio-temporal correlation matrix and σ^2_ω is sometimes called the spatio-temporal dependent error variance (or spatio-temporal partial sill). $\boldsymbol{\nu}$ is also a random vector with mean $\mathbf{0}$ but has covariance $\text{cov}(\boldsymbol{\nu}) = \sigma^2_\nu\mathbf{I}_{st}$, where $\mathbf{I}_{st}$ is the $N \times N$ identity matrix and σ^2_ν is sometimes called the spatio-temporal independent error variance (or spatio-temporal nugget).

Though there are a few types of models for the errors that can be built from equation 3 by setting certain error variances to 0 (e.g. a sum-with-error model sets $\sigma^2_\omega = 0$) and/or by allowing $\mathbf{R}_{st}$ to take certain forms (e.g. a separable model sets all covariance parameters equal to 0 except σ^2_ω and uses the structure for $\mathbf{R}_{st}$ given in equation 4 below), we focus only on the product-sum model (De Cesare, Myers, and Posa 2001; Sandra De Iaco, Myers, and Posa 2001). In a common formulation of the product-sum model, $\mathbf{R}_{st}$ is

$$\mathbf{R}_{st} \equiv \mathbf{Z}_s\mathbf{R}_s\mathbf{Z}_s' \odot \mathbf{Z}_t\mathbf{R}_t\mathbf{Z}_t', \tag{4}$$

where $\odot$ is the Hadamard product operator. Note that, in order to save on the number of parameters, we will assume that the $\mathbf{R}_s$ and $\mathbf{R}_t$ that form $\mathbf{R}_{st}$ are the same as the $\mathbf{R}_s$ and $\mathbf{R}_t$ associated with $\boldsymbol{\delta}$ and $\boldsymbol{\tau}$, respectively, although this is not necessary in general. $\mathbf{R}_s$ 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

$$\exp(-h_{ii'}/\phi), \tag{5}$$

where $\exp(x)$ is equivalent to e^x and ϕ 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

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

where ρ is a temporal range parameter controlling the decay rate of the covariance as the number of 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 (Hamilton 2020; Ver Hoef, London, and Boveng 2010).

The product-sum model for $\mathbf{y}$ is then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\delta} + \mathbf{Z}_s\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\omega} + \boldsymbol{\nu}, \tag{7}$$

where $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\tau}$, $\boldsymbol{\eta}$, $\boldsymbol{\omega}$, and $\boldsymbol{\nu}$ are mutually independent, $\mathbf{y}$ has mean $\mathbf{X}\boldsymbol{\beta}$, and $\mathbf{y}$ has covariance

$$\text{var}(\mathbf{y}) \equiv \boldsymbol{\Sigma} = \sigma^2_\delta \mathbf{Z}_s\mathbf{R}_s\mathbf{Z}'_s + \sigma^2_\gamma \mathbf{Z}_s\mathbf{I}_s\mathbf{Z}'_s + \sigma^2_\tau \mathbf{Z}_t\mathbf{R}_t\mathbf{Z}'_t + \sigma^2_\eta \mathbf{Z}_t\mathbf{I}_t\mathbf{Z}'_t + \sigma^2_\omega \mathbf{R}_{st} + \sigma^2_\nu \mathbf{I}_{st}. \tag{8}$$

There are a few reasons 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 $\sigma^2_\omega > 0$ or $\sigma^2_\nu > 0$, then the covariance matrix in equation 8 is also positive definite (De Cesare, Myers, and Posa 2001; Sandra De Iaco, Myers, and Posa 2001). Also, the product-sum model is non-separable, giving it more flexibility than a separable (product) model for the covariance (S. De Iaco, Palma, and Posa 2015; Dumelle et al. 2021). Xu and Shu (2015) claim that the product-sum model is widely used in practical applications.

2.2 Finite Population Block Kriging

The model that we developed in the previous section in equation 7 is for the N-length vector $\mathbf{y}$. However, often we do not have the resources to sample or observe every spatial site during every time point. Therefore, we may have an interest in prediction of the response values on sites that were not observed,

particularly sites in the most recent time point. Throughout this section, let the subscript o denote data points that were "observed" or sampled, the subscript u denote data points that were "unobserved" or not sampled, and the subscript a denote "all" data points. Then, we can re-order the response vector $\mathbf{y}$ so that

$$\mathbf{y} \equiv \mathbf{y}_a = [\mathbf{y}'_u, \mathbf{y}'_o]'. \tag{9}$$

Our primary goal is to use the model developed for $\mathbf{y}_a$ in equation 7 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 $\mathbf{b}'_a\mathbf{y}_a$, a linear function of $\mathbf{y}_a$. The N-length vector $\mathbf{b}'_a$ might be, for example, a vector of 1's, in which case we would be predicting the total response across all sites and all time points. More typical would be the case where $\mathbf{b}_a$ has 1's for all sites in the current time point and all other elements are 0, in which case we would be predicting the total response in the current time point.

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

$$E\{(\mathbf{q}'\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)^2\} - E\{(\boldsymbol{\lambda}'_o\mathbf{y}_o - \mathbf{b}'_a\mathbf{y}_a)^2\} \tag{10}$$

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

$$\begin{pmatrix} \boldsymbol{\Sigma}_{o,o} & \mathbf{X}_o \\ \mathbf{X}'_o & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{o,o} & \boldsymbol{\Sigma}_{o,u} \\ \mathbf{X}'_o & \mathbf{X}'_u \end{pmatrix} \begin{pmatrix} \mathbf{b}_o \\ \mathbf{b}_u \end{pmatrix}, \tag{11}$$

where $\mathbf{m}$ is a p-length column vector of the Lagrange multipliers due to the unbiasedness constraint and the subscripts o and u denote observed and unobserved data points. For example, $\boldsymbol{\Sigma}_{o,o}$ denotes the $n_o \times n_o$ submatrix of $\boldsymbol{\Sigma}$ (from equation 8) corresponding only to rows and columns of observed data points and $\boldsymbol{\Sigma}_{u,o}$ denotes the $(N - n_o) \times n_o$ submatrix of $\boldsymbol{\Sigma}$ 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

$$\boldsymbol{\lambda}'_o = \mathbf{b}'_o + \mathbf{b}'_u \left[(\boldsymbol{\Sigma}_{u,o}\boldsymbol{\Sigma}^{-1}_{o,o}) - (\boldsymbol{\Sigma}_{u,o}\boldsymbol{\Sigma}^{-1}_{o,o})\mathbf{X}_o\mathbf{W}^{-1}_o\mathbf{X}'_o\boldsymbol{\Sigma}^{-1}_{o,o} + \mathbf{X}'_u\mathbf{W}^{-1}_o\mathbf{X}_o\boldsymbol{\Sigma}^{-1}_{o,o}\right], \tag{12}$$

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

$$\widehat{\mathbf{b}'_a\mathbf{y}_a} = \boldsymbol{\lambda}'_o\mathbf{y}_o, \tag{13}$$

which is equivalent to

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

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

The finite-population-corrected- variance of the BLUP in equation 13 is

$$E((\boldsymbol{\lambda}_o'\mathbf{y}_o - \mathbf{b}_a'\mathbf{y}_a)^2) = \boldsymbol{\lambda}_o'\boldsymbol{\Sigma}_{o,o}\boldsymbol{\lambda}_o - 2\mathbf{b}_a'\boldsymbol{\Sigma}_{a,o}\boldsymbol{\lambda}_o + \mathbf{b}_a'\boldsymbol{\Sigma}_{a,a}\mathbf{b}_a. \tag{14}$$

We call the predictor in equation 13 with $\boldsymbol{\Sigma}$ in equation 8 the ST-FPBK predictor.

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

$$\mathbf{b}_a = [\mathbf{b}_{up}', \mathbf{b}_{uc}', \mathbf{b}_{op}', \mathbf{b}_{oc}']' = [\mathbf{0}', \mathbf{1}', \mathbf{0}', \mathbf{1}']', \tag{15}$$

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

Though we are interested in predicting $\mathbf{b}_a'\mathbf{y}_a$ in the development above so that we are predicting a single quantity of interest and obtaining a single finite-population-corrected-prediction variance, we can also predict for multiple quantities of interest and obtain a prediction covariance matrix with $\mathbf{B}'\mathbf{y}_a$, where $\mathbf{B}$ is an $N \times n_{pred}$ matrix and n_{pred} is the number of different quantities we wish to predict. One setting where such development is potentially useful is predicting the total abundance for each time point in the study, as is done in Section 3 in Figure 3.

2.3 Estimation

In practical applications, the covariance matrix $\boldsymbol{\Sigma}$ in equation 8 that is partitioned into the various sub-matrices in equations 13 and 14 needs to be estimated from the observed data $\mathbf{y}_o$. The spatio-temporal model in equation 7 does not have any distributional assumptions: we only need to specify the mean and variance of $\mathbf{y}_o$. Restricted Maximum Likelihood (REML) can be used to estimate the covariance parameters in $\boldsymbol{\Sigma}$, 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 Thompson 1971; Harville 1977). Even if $\mathbf{y}_a$ is not multivariate normal, the REML estimator for the parameter vector $\boldsymbol{\theta}$ is formed from a set of unbiased estimating equations (Heyde 1994; Cressie and Lahiri 1993).

However, REML estimation can be computationally burdensome, particularly for large spatio-temporal data sets with many observed sites and time

points. To speed up estimation of $\boldsymbol{\theta}$ when every site is observed at every time point, Dumelle et al. (2021) iteratively apply the Sherman-Morrison-Woodbury formula (Sherman and Morrison 1950; Woodbury 1950) on the Stegle eigen-decomposition (Stegle et al. 2011) of $\boldsymbol{\Sigma}$ to more quickly obtain $\boldsymbol{\Sigma}^{-1}$ during the optimization process. When every site is not observed at every time point, Helmert-Wolf blocking (Wolf 1979) is also incorporated to retain computational efficiency during estimation. To give a rough idea of the computational efficiency gained, Dumelle et al. (2021) found that the computation time for a single matrix inversion for a particular $15,000 \times 15,000$ spatio-temporal covariance matrix took about 80 seconds with their methodology, compared to over 1500 seconds for an inversion computed with the standard Cholesky decomposition. Note, however, that in REML optimization, many matrix inversions take place during the optimization process and the total number of iterations during optimization can vary from setting to setting. We use the developments from Dumelle et al. (2021) in the application in Section 3, the simulations described in Section 4, and the accompanying `R` package to speed up estimation of $\boldsymbol{\theta}$.

3 Application

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

3.1 Data Description

The Taylor Corridor in the east-central region of Alaska is a popular area for moose hunters. Within the Taylor Corridor, abundance surveys for moose are performed annually so that biologists can assess annual abundance and monitor the moose population size. In particular, surveys were conducted from 2014 through 2020 in every year except 2016, during which sightability of moose was low due to a lack of snow cover in the area. The spatial sampling frame for our study area consists of 381 sites. There are a total of 7 unique time points represented in the data, including the missing year of 2016. Therefore, N is 2667.

In each year of the survey, a team of biologists stratifies all of the spatial sites into a "High" stratum and a "Low" stratum based on wildlife biologist knowledge of moose density in the region and counts from previous surveys, which sites have land cover more suitable to moose habitat, and, in some years, stratification flights done on a portion of the region of interest prior to that year's survey (Figure 1). The stratification scheme therefore can change from year to year, though the majority of spatial sites are classified into the same stratum in every year. Biologists then randomly select some of the 381 sites to survey, and subsequently select a few additional sites non-randomly in such a way that there are no large areas in the region of interest without a sampled site. The non-randomly selected sites are a small proportion of the overall sample

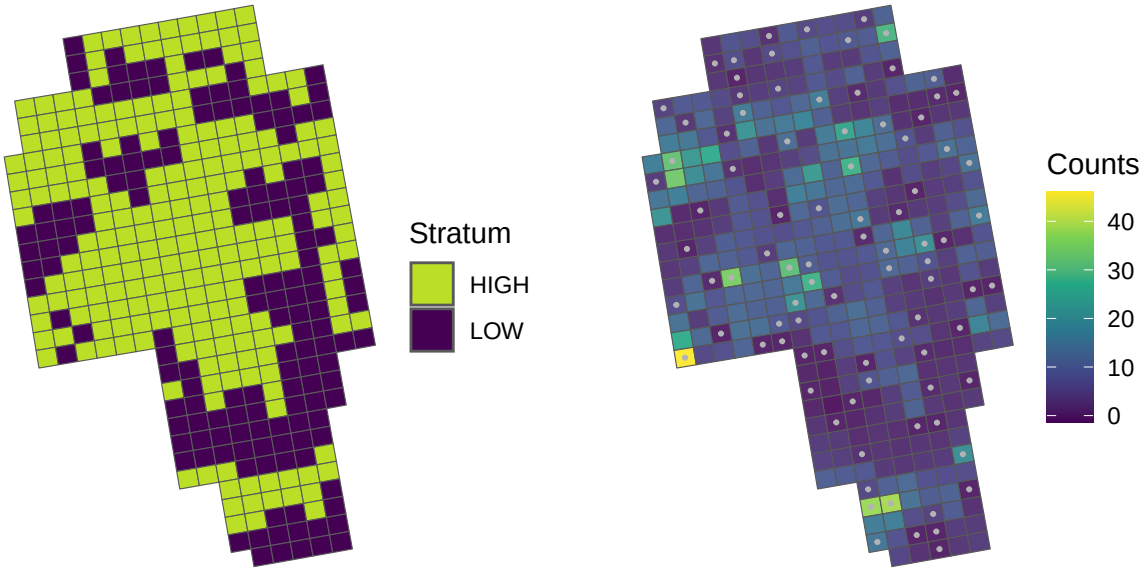

Figure 1: A map of the sites composing the Taylor corridor in eastern-central Alaska. Each site is roughly 4 kilometers in length and roughly 3.5 kilometers in width so that the centroids of two horizontally adjacent sites are about 4 kilometers apart and the centroids of two vertically adjacent sites are about 3.5 kilometers apart. (a). A map of the stratification for the sites in the year 2020. (b). A map of the predictions of sites in 2020 from the spatio-temporal model. A site with a grey dot in the center means that the site was sampled in 2020.

size (about 10-15% of the sites are selected non-randomly), and these sites are not selected preferentially by the number of expected moose or observed moose counts in previous survey years. The operations manual by Kellie and DeLong (2006) provides more details about how sites are selected for a survey in a particular year. The total number of sites that were selected varies from a low of 76 in the year 2019 to a high of 90 in the year 2020. Throughout the 7 unique years, some sites were sampled as many as five different times while others were never sampled at all. The number of units sampled throughout all survey years, n, was 487 units. Figure 1 and all remaining figure graphics are constructed with the `ggplot2 R` package (Wickham 2016).

The goal of the following analysis is to predict the total abundance of moose across all sites in the year 2020, the most recent year of the survey, using stratum as a covariate in the spatio-temporal model.

3.2 Model Fitting

We fit the product-sum covariance model defined in equation 7 using REML with stratum as a covariate in the design matrix. To select the spatial and temporal correlation structures, we examined the AIC values for each of the nine crossed combinations of the exponential, spherical, and gaussian spatial correlation structures (Cressie 2015) and the exponential, spherical, and gaussian temporal correlation structures to assess model fit. The combination with the lowest AIC was the "gaussian spatial - exponential temporal" model. However, in this application, we opt to use the "exponential spatial - exponential temporal" model, with the exponential spatial correlation structure defined in equation 5 and the exponential temporal correlation structure defined in equation 6 so that the application correlation structure matches that of the simulations in Section 4 (which are easier to conceptualize when both the spatial and temporal correlations decay according to the same function). Additionally, the AIC value for this "exponential spatial - exponential temporal" model was within four points of the best model, indicating that the quality of the model fits are not too different anyway.

Table 1 gives the estimated parameters from the model fit with the exponential correlation structures.

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

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

To help interpret what some of these fitted covariance parameter estimates mean, we can construct a fitted covariance plot (Figure 2). As the spatial distance between the centroids of two sites increases (dark colour to light colour), the covariance of two random errors decreases to 0, with the $\hat{\phi}$ parameter estimate controlling the rate of decay. In fact, the model estimates the covariance to be nearly 0 when the centroids of two sites are 20 or more kilometers apart, no matter what the temporal distance is. The covariance between two errors that are six years apart is still estimated to be positive if the two errors come from the same site or from adjacent sites.

The estimated vector of fixed effects, using "High" as the reference group, is $\hat{\boldsymbol{\beta}}' = (\hat{\beta}_0, \hat{\beta}_1) = (9.62, -4.55)$. The standard error for $\hat{\beta}_0$ is 1.01 while the standard error for $\hat{\beta}_1$ is 0.93. Therefore, the overall mean for sites in the "High" stratum is estimated to be 9.62 moose while the overall mean for sites in the "Low" stratum is estimated to be 5.07 moose.

3.3 Prediction

We now use the fitted spatio-temporal model with the BLUP from equation 13 and weights given in equation 15 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 13 and 14 and letting elements of $\mathbf{b}_a$ be equal to 1 for data points in 2020 and equal to 0 otherwise, we obtain a prediction of 3001 moose and a standard error (the square root of the prediction variance) of 217 moose. The prediction for the total and the prediction variance were fairly robust to other combinations of correlation functions to model the spatial and the temporal correlation, with the prediction never deviating from 3001 by more than 100 moose for any of the other 8 spatial and temporal correlation function combinations discussed at the beginning of Subsection 3.2.

A 90% normal-based prediction interval for the total abundance in 2020 (with the exponential spatial and temporal correlation) is (2644, 3357) moose. Note that, though the response in this example is a count, a normal-based prediction interval for the total is still appropriate through an application of the central limit theorem for dependent data (Smith 1980). Sitewise predictions for

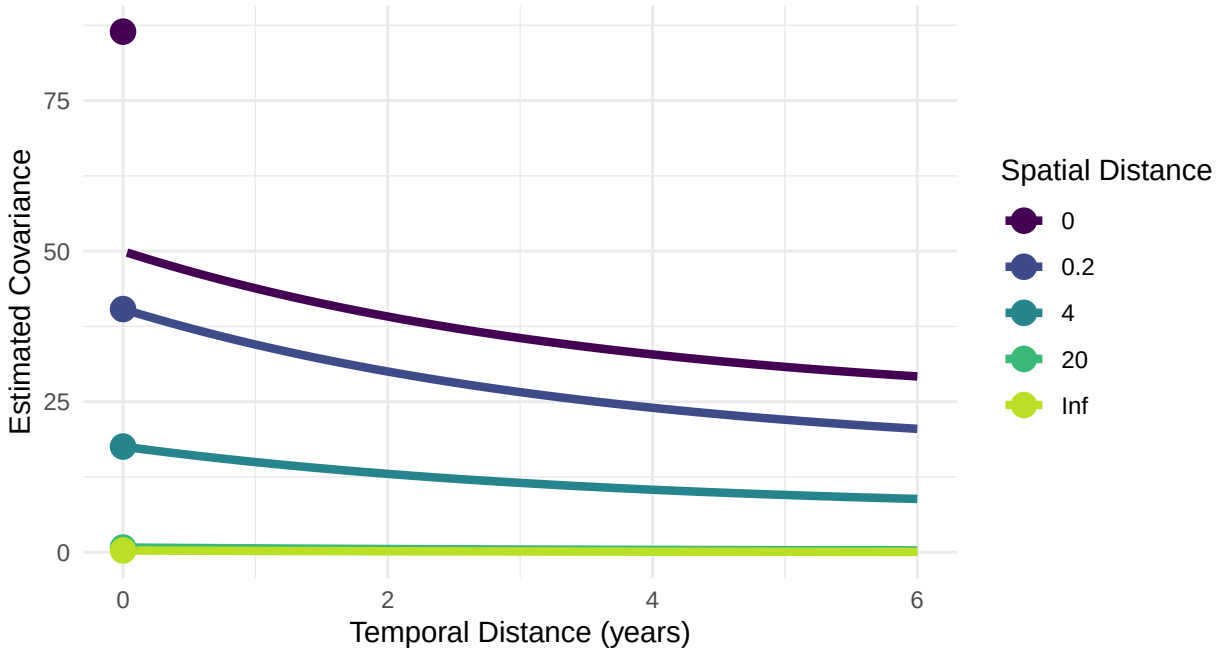

Figure 2: Estimated covariance of the errors from the estimated parameters in a spatio-temporal product-sum model. Distance between two sites is calculated from the site centroids; the centroids of two sites directly adjacent to one another are about 3.5 to 4 kilometers apart.

sites in 2020 are given in the map in Figure 1.

For comparison, we use the spatial `sptotal` package (Higham, Ver Hoef, Frank, et al. 2021) to compute the spatial FPBK prediction (Ver Hoef 2008) for the total abundance of moose in the year 2020 with stratum as a covariate. Note that the widely used GSPE software for moose surveys allows for the strata to have different covariance parameters and does not treat stratum as a covariate (DeLong 2006). For the application of moose abundance prediction, analyzing each stratum individually often results in better precision. The separate-stratum analysis is discussed in more detail in the Supplementary Material in the Appendix.

We also use the stratified random sampling design-based estimator

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

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

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

where s_i^2 is the sample variance of the observed data points in 2020 in the i^{th} stratum and n_i is the number of observed data points in 2020 in the i^{th} stratum. Both the purely spatial model fit with `sptotal` and the stratified random sampling design-based estimator use data only from 2020. Note that the stratified random sampling estimator is not actually appropriate for this application because not all sites are randomly selected within each stratum. However, the number of sites that are not randomly selected is small, so we still include the stratified random sampling estimator for a baseline comparison estimator.

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

In addition to making a prediction for the abundance in the most recent survey, we can also use the spatio-temporal model to backcast predictions for the abundance in past survey years, interpolate predictions for years during which a survey was not completed, and forecast predictions for future years. For example, in the Taylor Corridor surveys, there was no survey conducted

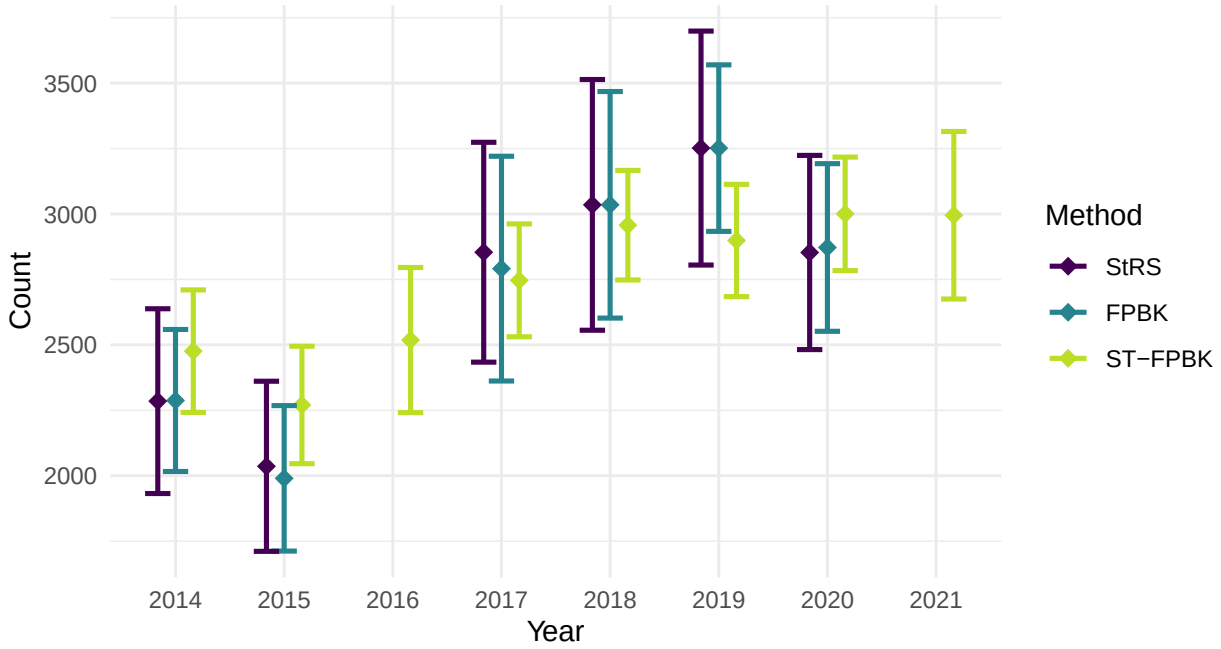

Figure 3: Moose abundance predictions for the Taylor Corridor from 2014 through 2021 with the stratified random sampling (StRS) estimator, the spatial FPBK predictor, and the ST-FPBK predictor. Predictions are given with a diamond symbol; the bars surrounding each prediction are standard error bars. Because surveys were not conducted in 2016 and 2021, there is no StRS estimator or spatial FPBK predictor for those years. Also, the standard errors for the ST-FPBK predictor for those years is larger than the standard errors in the other years. The stratification scheme used for 2016 and 2021 in the ST-FPBK analysis was the same scheme used in 2015 and 2020, respectively.

in the year 2016 because of insufficient snow cover. Leveraging the temporal structure of the ST-FPBK predictor, we can still construct a prediction and corresponding standard error though, as expected, this standard error is larger than the standard errors of years where a survey was completed (Figure 3). Also, in Figure 3, we see a forecasted prediction and corresponding standard error for the abundance in 2021. Again, the standard error associated with the forecasted prediction is larger than the standard errors for the years with completed surveys.

4 Simulation

4.1 Description

To further evaluate performance of the ST-FPBK predictor, we conduct a simulation study. We simulate a response vector $\mathbf{y}$ of length $N = 1000$ on a 10×10 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. The random vector $\mathbf{y}$ is multivariate normal with mean $\mathbf{0}$ and product-sum covariance matrix $\mathbf{\Sigma}$ defined in equation 8 with the covariance parameters given in Table 2.

Table 2: Covariance parameters used to simulate data. σ^2_δ, σ^2_γ, and ϕ are the spatial dependent error variance, independent error variance, and range parameters, respectively. σ^2_τ, σ^2_η, and ρ are the temporal dependent error variance, independent error variance, and range parameters, respectively. σ^2_ω and σ^2_ν are the spatio-temporal dependent error variance and spatio-temporal independent error variance. Note that both ϕ (and ρ) appear in $\mathbf{R}_{st}$; therefore, their values can change the underlying covariance even when σ^2_δ (and σ^2_τ) are equal to 0.

	Spatial			Temporal			Spatio-temporal	
scenario	σ^2_δ	σ^2_γ	ϕ	σ^2_τ	σ^2_η	ρ	σ^2_ω	σ^2_ν
all-dev	0.5	0.17	0.47	0.5	0.17	0.33	0.50	0.17
t-iev	0.0	0.00	0.47	0.0	1.50	0.00	0.25	0.25
spt-iev	0.0	0.00	0.00	0.0	0.00	0.00	0.00	2.00

The three scenarios in Table 2 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 $\sigma^2_\delta, \sigma^2_\tau$, and σ^2_ω; (2) **t-iev**: a scenario where there the overall variance is dominated by the temporal independent error variance parameter, σ^2_η; and (3) **spt-iev**: a scenario where all of the variability comes from σ^2_ν 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 ϕ and ρ as the range parameters in equations 5 and 6. The values 0.471 and 0.3333 are chosen for ϕ and ρ, respectively, so that the effective ranges, 3ϕ and 3ρ, are equal to the maximum possible distance between two data points in space ($\sqrt{2} = 1.414$) and the maximum possible distance between two data points in time (1). A value of 0 for ϕ (or ρ) sets the $\mathbf{R}_s$ (or the $\mathbf{R}_t$) matrix to the identity matrix. Figure 4 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 is chosen from the 1000 total data points.

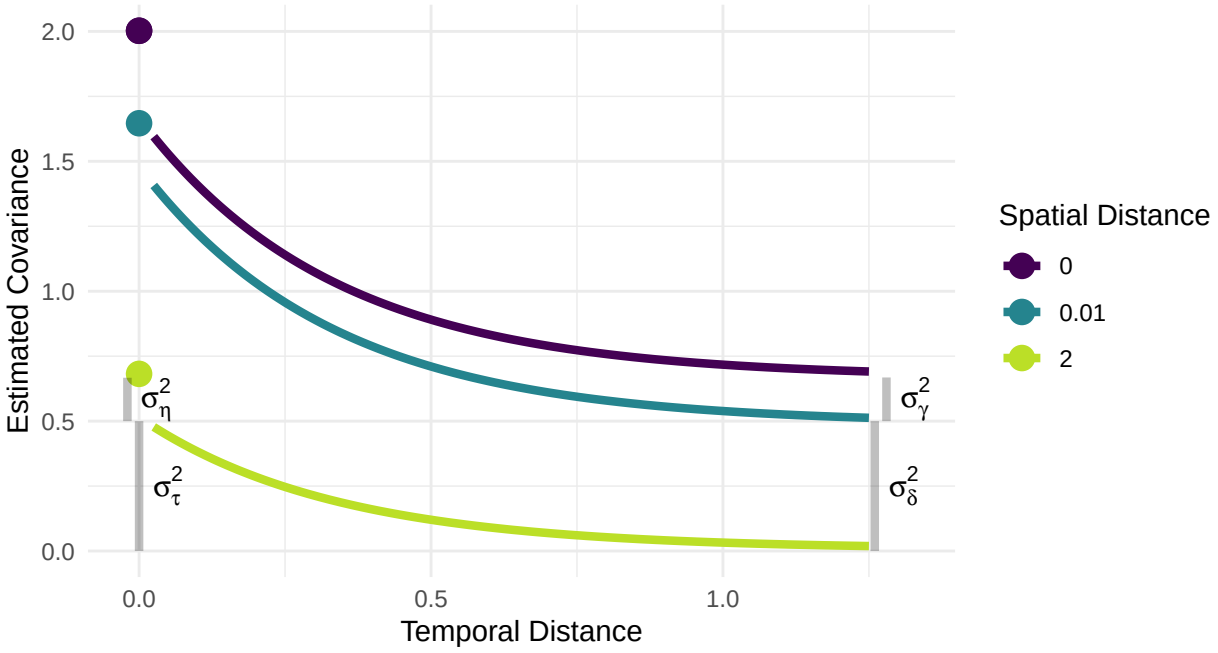

Figure 4: The model covariance used in the simulations for the spatio-temporal scenario. Covariance is approximately 0 for errors from data points that are $\sqrt{2}$ distance units apart in space and 1 distance unit apart in time. The spatial dependent error variance (σ^2_δ), spatial independent error variance (σ^2_γ), temporal dependent error variance (σ^2_τ), and temporal independent error variance (σ^2_η) are shown with grey lines.

Finally, the simulation experiment is repeated for a continuous skewed response variable and for a skewed response variable of discrete counts. To create the continuous skewed response variable for the setting called "skewed," a normally-distributed response is simulated according to the parameters given in Table 2, except that each of the variance parameters (not including ϕ and ρ) 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 that is not used in the model fitting. To create the skewed response variable of discrete counts for the setting called "poisson," for each response value, we take a Poisson draw with the continuous skewed response value as the mean (conditional on the mean, each Poisson draw is done independently of all other Poisson draws).

Therefore, the simulation study has 18 total settings coming from a $3 \times 2 \times 3$ (scenario $\times$ sample size $\times$ distribution) 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 for the most current time point quantity" the "current total."

The first method uses the ST-FPBK predictor in equation 13 with the spatio-temporal model covariance in equation 8. REML estimation with the observed data $\mathbf{y}_o$ is used to obtain estimates for the covariance parameter vector $\boldsymbol{\theta}$. The second method is the FPBK spatial model fit with the `sptotal R` package (Higham, Ver Hoef, Frank, et al. 2021) that only uses data from the most current time point.

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

The SRS method gives an estimator, not a predictor, and a corresponding confidence interval, not a prediction interval, because the SRS design-based 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 and tables.

For each method, we calculate the root-mean-squared-prediction-error (rMSPE) as $\sqrt{\frac{1}{1000}(\sum_{i=1}^{1000}(T_i - \hat{T}_i)^2)}$, where T_i and $\hat{T}_i$ are the realized and predicted current totals, respectively, in the i^{th} iteration. Bias is recorded as

$\frac{1}{1000}\sum_{i=1}^{1000}(T_i - \hat{T}_i)$. We also create a normal-based 90% prediction interval for the realized current total and 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 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 A1, A2, and A3 in the Appendix give the rMSPE, bias, and interval coverage of the three methods in all 18 simulation settings. In Figure 5, we see that the ST-FPBK predictor outperforms both the purely spatial FPBK 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 σ^2_η, the temporal independent error variance. In this scenario, the mean of the response, conditional on the random effects, can fluctuate drastically from time point to time point. Therefore, in a model without any fixed effects, the realized total is susceptible to time point to time point increases and decreases more than the realized total is in the other scenarios. As expected, the ST-FPBK predictor performs no better than a purely spatial model or the SRS design-based estimator for the "t-iev" scenario because the information from data in other time points is not as useful. 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 A2 shows that the bias of each method is small compared to the squares of the rMSPE values given in Table A1.

The normal-based prediction intervals (Smith 1980) for the abundance in the most recent time point from the ST-FPBK method maintain close to appropriate coverage (90%) for all of the simulation settings used, including the scenarios where the response is skewed right and the covariance model is mis-specified and the scenarios where the response is both discrete and skewed right (Table A3). 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 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

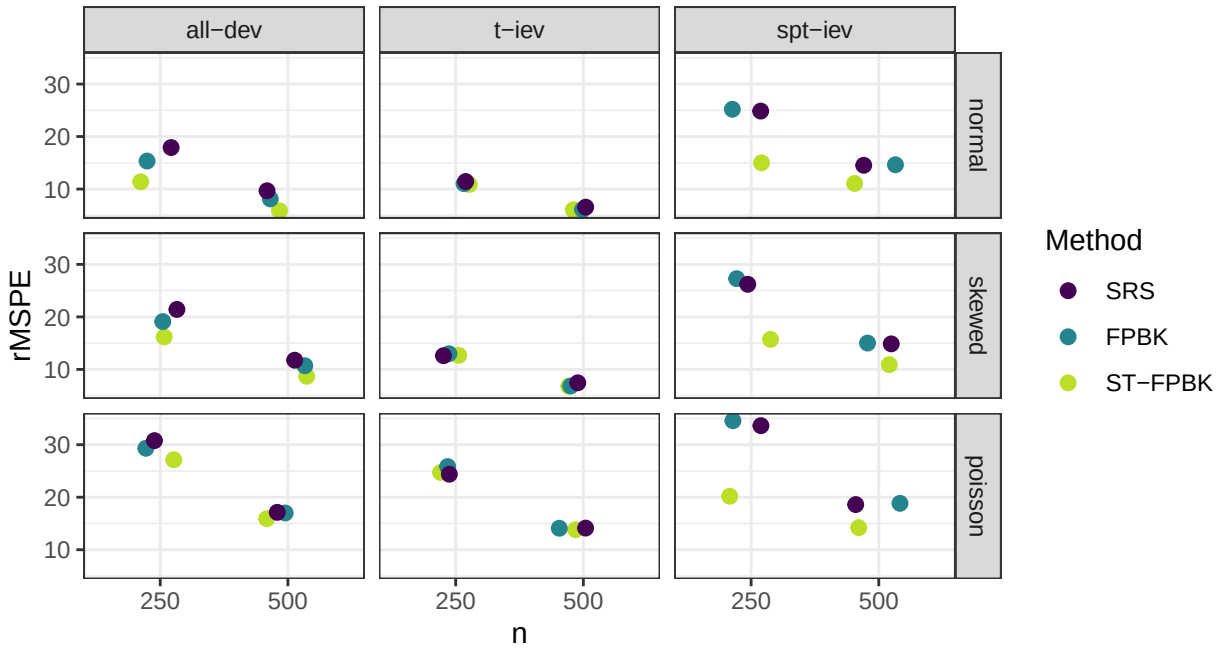

Figure 5: root-mean-squared-prediction-error (rMSPE) for all simulation settings. The ST-FPBK predictor has the smallest rMSPE in all of the 'all-dev' and 'spt-iev' scenarios while the three methods perform similarly in all of the 't-iev' scenarios.

in Section 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 or the response is discrete.

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. In Figure 3, we see the forecasted prediction for abundance in the year 2021. While there is a (presumed) loss in precision by constructing a prediction for a year that has no observed samples, the prediction could still be useful to wildlife managers for decision-making before a survey from that year is completed and analyzed. Constructing a prediction for years or time points at which a survey is not completed can be applied to other contexts as well, including temporal interpolation (e.g., the year 2016 in Figure 3).

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

We would also like to give our perception of the benefits and drawbacks of our approach with using a Bayesian hierarchical model. For example, Schmidt et al. (2022) use a Bayesian hierarchical model with spatial radial basis functions that are estimated per year and with time as a trend component in the fixed effects to make predictions for moose abundance. We argue that our approach is both simpler for practitioners and less likely to yield an unreasonable abundance prediction, compared to a Bayesian hierarchical model. In general, fitting a Bayesian hierarchical model with a complex covariance structure on the link scale requires careful thought in formulating the model and the prior distributions used for all of the covariance parameters (Conn et al. 2015).

Conn et al. (2015) also note that the need to tailor the model used to the richness of the particular data set at hand is especially important when using the log link function, as abundance estimates can become unrealistically large when back-transforming. Indeed, Ver Hoef et al. (2021) fit a Bayesian hierarchical model to marine mammal counts using a truncated normal distribution for the random errors on the log scale instead of the typical normal distribution. Applying a Bayesian hierarchical model to this data set with normally distributed random errors on the link scale and non-informative prior distributions resulted in predicted counts well above what any biologist familiar with the region would consider reasonable. Conn et al. (2014) and Ver Hoef and Jansen (2007) provide more evidence that applying a hierarchical model with a log link function could result in unrealistically high predictions, particularly when a small proportion of the region of interest is sampled. All of these examples indicate that a Bayesian hierarchical model may need significant adjustment based on the particular data

set at hand.

Another benefit of our approach is a faster fitting time, as there is no need to construct and implement the time-consuming Markov chain Monte Carlo sampler. The moose application model in Section 3 takes about 10 minutes to fit. The shorter fitting time allows us to fit and compare models with different underlying spatial and temporal covariances and/or with different fixed effects structures, which would be much more time-consuming in the Bayesian framework. For both our approach and a Bayesian approach, there is a trade-off here between how many surveys to incorporate into the model (the Alaska Department of Fish and Game has done surveys with this structure since the late 1990's) and how long the model will take to fit. We expect there to be diminishing returns in precision when incorporating older surveys, though the rate at which the returns diminish is dependent upon the application at hand.

Additionally, with the shorter fitting time and the supplementary `R` package provided to fit the models, integrating this approach with the current GSPE software would be more reasonable and would also allow scientists other than statisticians to perform the analysis. Finally, our approach is easier to assess in a simulation study, which would be too time-prohibitive for the Bayesian model. Biometricians could use simulation with our approach to answer various questions given proposed values of covariance parameters like how much efficiency would drop if a survey was only conducted every other year.

Bayesian hierarchical models, including the model by Schmidt et al. (2022), however, offer features that would be harder to implement in our approach. These models allow for incorporation of more levels in the model structure, allowing, for example, for imperfect detection of animals from a separate detectability survey. Additionally, the Bayesian hierarchical model can use a Poisson or negative binomial model for the counts. Therefore, an appropriate prediction interval for the response on one particular site could be constructed. On the other hand, for our approach, we rely on the central limit theorem for dependent data to form a prediction interval for the total. Therefore, constructing a prediction interval for a single response on one unobserved site would not be sensible with our approach.

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

Future work in this area includes developing a frequentist model for which imperfect detection of units through time is incorporated into the predictor (Higham, Ver Hoef, Madsen, et al. 2021) or how best to select sites to sample for future surveys given proposed values for the spatio-temporal covariance parameters. Additionally, for moose surveys in particular, updating the GSPE software to include analysis for spatio-temporal data could be useful for practitioners. Though we recognize that doing so would be a substantial undertaking,

the `R` package that we provide could be a useful starting point for the integration.

6 Declarations

Conflicts of Interest

The authors declare no conflict of interest.

Data and Code Availability

The Alaska Department of Fish and Game collected and provided the moose survey data used in this study. This manuscript has a supplementary `R` package that contains all of the data and code used in its creation, with the exception of the shapefile used to make the maps in some of the figures (which cannot be released due to Alaska Department of Fish and Game policy). The supplementary `R` package, along with the data used in the application, is hosted on GitHub and can be found at (link not provided because repository would identify at least one author).

The data set is also available on Zenodo at `https://doi.org/10.5281/zenodo.7636130`.

Acknowledgements

The views expressed in this manuscript are those of the authors and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency or the National Oceanic and Atmospheric Administration. Any mention of trade names, products, or services does not imply an endorsement by the U.S. government, the U.S. Environmental Protection Agency, or the National Oceanic and Atmospheric Administration. The U.S. Environmental Protection Agency and National Oceanic and Atmospheric Administration do not endorse any commercial products, services, or enterprises.

Appendix

A.1: Simulation Tables

Table A1: root-mean-squared-prediction-error (rMSPE) for the ST-FPBK predictor, the FPBK predictor, and the SRS estimator for each of the 18 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	SRS	FPBK	ST-FPBK
spt-iev	250	normal	24.85	25.19	14.99
t-iev	250	normal	11.44	11.01	10.88
all-dev	250	normal	17.91	15.33	11.38
spt-iev	500	normal	14.53	14.63	11.06
t-iev	500	normal	6.55	6.10	6.05
all-dev	500	normal	9.68	8.13	5.87
spt-iev	250	skewed	26.21	27.28	15.73
t-iev	250	skewed	12.60	12.98	12.69
all-dev	250	skewed	21.42	19.10	16.19
spt-iev	500	skewed	14.86	15.01	10.91
t-iev	500	skewed	7.45	6.84	6.80
all-dev	500	skewed	11.75	10.71	8.68
spt-iev	250	poisson	33.63	34.57	20.19
t-iev	250	poisson	24.39	25.84	24.73
all-dev	250	poisson	30.78	29.32	27.14
spt-iev	500	poisson	18.60	18.84	14.21
t-iev	500	poisson	14.13	14.09	13.82
all-dev	500	poisson	17.12	16.99	15.90

Table A2: Bias (Realized Current Total - Predicted Current Total) for the ST-FPBK predictor, the FPBK predictor, and the SRS estimator for each of the 18 simulation settings. In all settings, all methods appear fairly unbiased.

Simulation Setting			Bias		
scenario	n	Response Type	SRS	FPBK	ST-FPBK
spt-iev	250	normal	-0.40	-0.63	0.34
t-iev	250	normal	-0.28	-0.48	-0.38
all-dev	250	normal	-0.30	-0.45	-0.36
spt-iev	500	normal	-0.30	-0.35	0.05
t-iev	500	normal	-0.20	-0.21	-0.20
all-dev	500	normal	0.11	-0.08	-0.19
spt-iev	250	skewed	-0.58	-1.72	-0.14
t-iev	250	skewed	-0.49	-1.02	-0.82
all-dev	250	skewed	-0.19	-0.74	-0.47
spt-iev	500	skewed	-0.44	-0.69	-0.15
t-iev	500	skewed	-0.23	-0.33	-0.30
all-dev	500	skewed	0.18	-0.08	-0.01
spt-iev	250	poisson	-0.72	-1.79	-0.03
t-iev	250	poisson	-0.70	-1.25	-1.01
all-dev	250	poisson	-0.52	-1.34	-0.67
spt-iev	500	poisson	-0.28	-0.43	0.07
t-iev	500	poisson	-0.61	-0.69	-0.61
all-dev	500	poisson	0.57	0.30	0.61

Table A3: Prediction interval coverage for the ST-FPBK predictor, the FPBK predictor, and the SRS for each of the 18 simulation settings. All intervals are normal-based and have a nominal coverage level of 0.90.

Simulation Setting			Coverage		
scenario	n	Response Type	SRS	FPBK	ST-FPBK
spt-iev	250	normal	0.89	0.88	0.91
t-iev	250	normal	0.88	0.87	0.89
all-dev	250	normal	0.88	0.87	0.90
spt-iev	500	normal	0.89	0.88	0.90
t-iev	500	normal	0.89	0.88	0.89
all-dev	500	normal	0.91	0.88	0.90
spt-iev	250	skewed	0.86	0.83	0.90
t-iev	250	skewed	0.88	0.86	0.90
all-dev	250	skewed	0.86	0.86	0.89
spt-iev	500	skewed	0.86	0.86	0.90
t-iev	500	skewed	0.89	0.88	0.92
all-dev	500	skewed	0.89	0.86	0.90
spt-iev	250	poisson	0.86	0.84	0.91
t-iev	250	poisson	0.89	0.87	0.90
all-dev	250	poisson	0.87	0.86	0.85
spt-iev	500	poisson	0.88	0.87	0.90
t-iev	500	poisson	0.89	0.88	0.91
all-dev	500	poisson	0.87	0.86	0.88

A.2: Supplementary Analysis

As mentioned in Section 3, moose surveys in Alaska are often stratified into "High" and "Low" sites. When using stratum as a covariate in a spatio-temporal (or spatial, if performing a purely spatial analysis) model, we assume that all errors in the model are generated from the same underlying spatio-temporal (or spatial) parameters. However, for many moose surveys, it is more reasonable to allow the sites in the High stratum to have a different set of spatio-temporal (or spatial) parameters than the sites in the Low stratum.

If we allow the strata to have different covariance parameters, then, to construct the ST-FPBK predictor, we simply fit the model once for each stratum. If we assume that there is no cross-covariance (i.e. errors from sites in different strata are not correlated), then the BLUP for $\mathbf{b}_a'\mathbf{y}_a$ is

$$\widehat{\mathbf{b}_a'\mathbf{y}_a} = \boldsymbol{\lambda}_{o,l}'\mathbf{y}_{o,l} + \boldsymbol{\lambda}_{o,h}'\mathbf{y}_{o,h}, \tag{16}$$

where $\boldsymbol{\lambda}_{o,l}$ and $\boldsymbol{\lambda}_{o,h}'$ are the kriging weights for the Low and High strata, respectively (equation 13), and $\mathbf{y}_{o,l}$ and $\mathbf{y}_{o,h}$ are the vectors of observed responses for

the Low and High strata, respectively.

Again assuming that there is no cross-covariance, the prediction variance is simply the sum of the prediction variances of $\boldsymbol{\lambda}'_{o,l}\mathbf{y}_{o,l}$ and $\boldsymbol{\lambda}'_{o,h}\mathbf{y}_{o,h}$ using equation 14.

We can use the purely spatial model and FPBK as well as the spatio-temporal model and ST-FPBK to predict the total moose abundance in 2020, using separate covariance models for the strata in the moose data set in Section 3. Table A4 shows the results.

Table A4: Prediction and standard error for total abundance in 2020 using a model that allows errors in separate strata to be modeled with different covariance parameters. For reference, the prediction and standard error from the models with stratum as a covariate are also given.

method	Prediction	SE
FPBK Sep. Strat.	2900	297
ST-FPBK Sep. Strat.	2867	242
FPBK	2870	319
ST-FPBK	3001	217

The spatio-temporal predictors still have a smaller standard error than their purely spatial model counterparts. Interestingly, the purely spatial FPBK predictor has a slightly lower standard error when fitting strata separately while the ST-FPBK predictor has a slightly lower standard error when using stratum as a covariate. Whether it makes more sense for stratum to be a covariate or for the strata to be fit separately is application dependent.

For the moose application data, fitting separate covariance models to each stratum is probably the better choice, as the errors for sites in the high stratum have much more overall variability than the errors in the low stratum. However, we chose to have the separate-strata model in the supplementary materials for two reasons. First, the method can be applied to any data set with spatio-temporal covariance and a finite number of sites, and applications in other domains may not have stratification at all. Second, the syntax in the development of the ST-FPBK predictor is much cleaner when stratum is treated as a covariate than when the strata are fit separately. Using the model with stratum as a covariate allows for a better focus on the proposed method itself.

References

Adde, Antoine, Marcel Darveau, Nicole Barker, and Steven Cumming. 2020. "Predicting Spatiotemporal Abundance of Breeding Waterfowl Across Canada: A Bayesian Hierarchical Modelling Approach." *Diversity and Distributions* 26 (10): 1248–63.

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): 314–27.

Breidt, F Jay, and Jean D Opsomer. 2017. "Model-Assisted Survey Estimation with Modern Prediction Techniques."

Breivik, Olav Nikolai, Fredrik Aanes, Guldborg Søvik, Asgeir Aglen, Sigbjørn Mehl, and Espen Johnsen. 2021. "Predicting Abundance Indices in Areas Without Coverage with a Latent Spatio-Temporal Gaussian Model." *ICES Journal of Marine Science* 78 (6): 2031–42.

Brus, Dick J. 2021. "Statistical Approaches for Spatial Sample Survey: Persistent Misconceptions and New Developments." *European Journal of Soil Science* 72 (2): 686–703.

Chen, Wanfang, Marc G Genton, and Ying Sun. 2021. "Space-Time Covariance Structures and Models." *Annual Review of Statistics and Its Application* 8: 191–215.

Conn, Paul B, Devin S Johnson, Jay M Ver Hoef, Mevin B Hooten, Joshua M London, and Peter L Boveng. 2015. "Using Spatiotemporal Statistical Models to Estimate Animal Abundance and Infer Ecological Dynamics from Survey Counts." *Ecological Monographs* 85 (2): 235–52.

Conn, Paul B, Jay M Ver Hoef, Brett T McClintock, Erin E Moreland, Josh M London, Michael F Cameron, Shawn P Dahle, and Peter L Boveng. 2014. "Estimating Multispecies Abundance Using Automated Detection Systems: Ice-Associated Seals in the Bering Sea." *Methods in Ecology and Evolution* 5 (12): 1280–93.

Cressie, Noel. 2015. *Statistics for Spatial Data - Revised Edition.* John Wiley & Sons.

Cressie, Noel, and Soumendra Nath Lahiri. 1993. "The Asymptotic Distribution of REML Estimators." *Journal of Multivariate Analysis* 45 (2): 217–33.

Cressie, Noel, and Christopher K Wikle. 2015. *Statistics for Spatio-Temporal Data*. John Wiley & Sons.

Davy, Christina M, Kelly Squires, and J Ryan Zimmerling. 2021. "Estimation of Spatiotemporal Trends in Bat Abundance from Mortality Data Collected at Wind Turbines." *Conservation Biology* 35 (1): 227–38.

De Cesare, Luigi, DE Myers, and D Posa. 2001. "Product-Sum Covariance for Space-Time Modeling: An Environmental Application." *Environmetrics: The Official Journal of the International Environmetrics Society* 12 (1): 11–23.

De Iaco, Sandra, Donald E Myers, and Donato Posa. 2001. "Space–Time Analysis Using a General Product–Sum Model." *Statistics & Probability Letters* 52 (1): 21–28.

De Iaco, S, Donald E Myers, and D Posa. 2002. "Nonseparable Space-Time Covariance Models: Some Parametric Families." *Mathematical Geology* 34: 23–42.

De Iaco, S, M Palma, and D Posa. 2015. "Spatio-Temporal Geostatistical Modeling for French Fertility Predictions." *Spatial Statistics* 14: 546–62.

DeLong, Robert A. 2006. *Geospatial Population Estimator Software User's*

Guide. Alaska Department of Fish; Game, Division of Wildlife Conservation.

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): 2018–29.

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." University of Alaska. Institute of Arctic Biology.

Gneiting, Tilmann, Marc G Genton, and Peter Guttorp. 2006. "Geostatistical Space-Time Models, Stationarity, Separability, and Full Symmetry." *Monographs On Statistics and Applied Probability* 107: 151.

Hamilton, James Douglas. 2020. *Time Series Analysis*. Princeton university press.

Harville, David A. 1977. "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association* 72 (358): 320–38.

Heyde, CC. 1994. "A Quasi-Likelihood Approach to the REML Estimating Equations." *Statistics & Probability Letters* 21 (5): 381–84.

Higham, Matt, Jay Ver Hoef, Bryce Frank, and Michael Dumelle. 2021. *Sptotal: Predicting Totals and Weighted Sums from Spatial Data*. `https://highamm.github.io/sptotal/index.html`.

Higham, Matt, Jay Ver Hoef, Lisa Madsen, and Andy Aderman. 2021. "Adjusting a Finite Population Block Kriging Estimator for Imperfect Detection." *Environmetrics* 32 (1): e2654.

Kellie, Kalin A, Kassidy E Colson, and Joel H Reynolds. 2019. *Challenges to Monitoring Moose in Alaska*. Alaska Department of Fish; Game, Division of Wildlife Conservation Juneau ....

Kellie, Kalin A, and Robert A DeLong. 2006. "Geospatial Survey Operations Manual." Alaska Department of Fish; Game.

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.

Lohr, Sharon L. 2021. *Sampling: Design and Analysis*. Chapman; Hall/CRC.

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): 2874–89.

Montero, José-María, Gema Fernández-Avilés, and Jorge Mateu. 2015. *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. John Wiley & Sons.

Patterson, H Desmond, and Robin Thompson. 1971. "Recovery of Inter-Block Information When Block Sizes Are Unequal." *Biometrika* 58 (3): 545–54.

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): 639–49.

Porcu, Emilio, Reinhard Furrer, and Douglas Nychka. 2021. "30 Years of Space–Time Covariance Functions." *Wiley Interdisciplinary Reviews: Computational Statistics* 13 (2): e1512.

Posa, D. 1993. "A Simple Description of Spatial-Temporal Processes." *Computational Statistics & Data Analysis* 15 (4): 425–37.

Ross, Beth E, Mevin B Hooten, and David N Koons. 2012. "An Accessible Method for Implementing Hierarchical Models with Spatio-Temporal Abundance Data." *PLoS One* 7 (11): e49395.

Rouhani, Shahrokh, and Timothy J Hall. 1989. "Space-Time Kriging of Groundwater Data." In *Geostatistics: Proceedings of the Third International Geostatistics Congress September 5–9, 1988, Avignon, France*, 639–50. Springer.

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.

Sauer, John R, and William A Link. 2011. "Analysis of the North American Breeding Bird Survey Using Hierarchical Models." *The Auk* 128 (1): 87–98.

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 Management*, e22220.

Sherman, Jack, and Winifred J Morrison. 1950. "Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix." *The Annals of Mathematical Statistics* 21 (1): 124–27.

Smith, Tony E. 1980. "A Central Limit Theorem for Spatial Samples." *Geographical Analysis* 12 (4): 299–324.

Stegle, Oliver, Christoph Lippert, Joris M Mooij, Neil Lawrence, and Karsten Borgwardt. 2011. "Efficient Inference in Matrix-Variate Gaussian Models with\iid Observation Noise." *Advances in Neural Information Processing Systems* 24.

Stein, Michael L. 2005. "Space–Time Covariance Functions." *Journal of the American Statistical Association* 100 (469): 310–21.

Stock, Brian C, Eric J Ward, Tomoharu Eguchi, Jason E Jannot, James T Thorson, Blake E Feist, and Brice X Semmens. 2020. "Comparing Predictions of Fisheries Bycatch Using Multiple Spatiotemporal Species Distribution Model Frameworks." *Canadian Journal of Fisheries and Aquatic Sciences* 77 (1): 146–63.

Urquhart, N Scott. 2012. "The Role of Monitoring Design in Detecting Trend in Long-Term Ecological Monitoring Studies." *Design and Analysis of Long-Term Ecological Monitoring Studies*, 151–73.

Ver Hoef, Jay M. 2008. "Spatial Methods for Plot-Based Sampling of Wildlife Populations." *Environmental and Ecological Statistics* 15 (1): 3–13.

Ver Hoef, Jay M, and John K Jansen. 2007. "Space—Time Zero-Inflated Count

Models of Harbor Seals." *Environmetrics: The Official Journal of the International Environmetrics Society* 18 (7): 697–712.

Ver Hoef, Jay M, Devin Johnson, Robyn Angliss, and Matt Higham. 2021. "Species Density Models from Opportunistic Citizen Science Data." *Methods in Ecology and Evolution* 12 (10): 1911–25.

Ver Hoef, Jay M, Josh M London, and Peter L Boveng. 2010. "Fast Computing of Some Generalized Linear Mixed Pseudo-Models with Temporal Autocorrelation." *Computational Statistics* 25: 39–55.

Wang, Zhonglei, and Zhengyuan Zhu. 2019. "Spatiotemporal Balanced Sampling Design for Longitudinal Area Surveys." *Journal of Agricultural, Biological and Environmental Statistics* 24: 245–63.

Wickham, Hadley. 2016. "Data Analysis." In *Ggplot2*, 189–201. Springer.

Wikle, Christopher K, Andrew Zammit-Mangion, and Noel Cressie. 2019. *Spatio-Temporal Statistics with r*. Chapman; Hall/CRC.

Wolf, Helmut. 1979. "The Helmert Block Method and Its Origin." In *Proceedings: Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks, Held at the Marriott Hotel, Arlington, Virginia, April 24 to 28, 1978*, 55:319. Department of Commerce, National Oceanic; Atmospheric Administration ....

Woodbury, Max A. 1950. *Inverting Modified Matrices*. Department of Statistics, Princeton University.

Xu, Jiaqi, and Hong Shu. 2015. "Spatio-Temporal Kriging Based on the Product-Sum Model: Some Computational Aspects." *Earth Science Informatics* 8 (3): 639–48.