

# An Application of Spatiotemporal Modeling to Finite Population Abundance Prediction

Matt Higham<sup>a</sup>, John Smith<sup>b</sup>, Dominik Leutnant<sup>c,†,‡</sup>

<sup>a</sup>St. Lawrence University Canton, NY 13617; <sup>b</sup>Institut für Informatik, Albert-Ludwigs-Universität, Freiburg, Germany; <sup>c</sup>Muenster University of Applied Sciences - Institute for Infrastructure, Water, Resources, Environment, Correnstr. 25, 48149 Münster, Germany

## ARTICLE HISTORY

Compiled August 16, 2021

## ABSTRACT

Insert abstract here.

## KEYWORDS

spatial; temporal; kriging;

## 1. Introduction

### 1.1. *Motivation*

Moose surveys in Alaska and western Canada are often 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 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 many examples involving prediction in a finite region with spatial sites that are surveyed regularly.

---

CONTACT Matt Higham. Email: [mhigham@stlawu.edu](mailto:mhigham@stlawu.edu), John Smith. Email: [john.smith@uni-freiburg.de](mailto:john.smith@uni-freiburg.de), Dominik Leutnant. Email: [leutnant@fh-muenster.de](mailto:leutnant@fh-muenster.de)

## 1.2. Background

- 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. 2021). 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 current year of the survey, so that, for example, the prediction variance is zero if all sites in the current year are sampled.

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 site abundance, 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, and we then use the spatiotemporal model with a finite population correction factor to give a Best-Linear-Unbiased-Predictor (BLUP) and its variance for total abundance in a given year.

### 2.1. Spatiotemporal Model

Let  $Y(\mathbf{s}_i, t_j)$ ,  $i = 1, 2, \dots, n_{sp}$  and  $j = 1, 2, \dots, n_t$  be a random variable, where the vector  $\mathbf{s}_i$  contains the coordinates for the  $i^{th}$  spatial site location and where  $t_j$  is the  $j^{th}$  time point. With each spatial location at each time point, the total number of data points (observed and unobserved) is  $n_{sp} \cdot n_t \equiv N$ . Then, a model for  $\mathbf{y}(\mathbf{s}_i, t_j)$ , a vector of the  $Y(\mathbf{s}_i, t_j)$  is

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

where  $\mathbf{X}$  is a design matrix for 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). In particular, a sum-with-error linear mixed model for response vector  $\mathbf{y}(\mathbf{s}_i, t_j)$  is

$$\mathbf{y}(\mathbf{s}_i, t_j) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{sp}\boldsymbol{\delta} + \mathbf{Z}_{sp}\boldsymbol{\gamma} + \mathbf{Z}_t\boldsymbol{\tau} + \mathbf{Z}_t\boldsymbol{\eta} + \boldsymbol{\nu}. \quad (2)$$

$\mathbf{Z}_{sp}$  is an  $N \times n_{sp}$  matrix where the values in a row corresponding to an observation in location  $\mathbf{s}_i$  are a 1 in the  $i^{th}$  column and 0's in all other columns. Similarly,  $\mathbf{Z}_t$  is an  $N \times n_t$  matrix where the values in a row corresponding to an observation at time

point  $\mathbf{t}_j$  are a 1 in the  $j^{th}$  column and 0's in all other columns. We assume that  $\boldsymbol{\delta}$ ,  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\tau}$ ,  $\boldsymbol{\eta}$ , and  $\boldsymbol{\nu}$  all are all mean  $\mathbf{0}$  vectors of length  $n_{sp}$ ,  $n_{sp}$ ,  $n_t$ ,  $n_t$ , and  $N$ , respectively, with the covariances

$$\begin{aligned} cov(\boldsymbol{\delta}) &= \sigma_{\delta}^2 \mathbf{R}_{sp} \\ cov(\boldsymbol{\gamma}) &= \sigma_{\gamma}^2 \mathbf{I}_{sp} \\ cov(\boldsymbol{\tau}) &= \sigma_{\tau}^2 \mathbf{R}_t \\ cov(\boldsymbol{\eta}) &= \sigma_{\eta}^2 \mathbf{I}_t \\ cov(\boldsymbol{\nu}) &= \sigma_{\nu}^2 \mathbf{I}_N, \end{aligned}$$

where  $\mathbf{R}_{sp}$  is a spatial correlation matrix and  $\mathbf{R}_t$  is a temporal correlation matrix.  $\sigma_{\delta}^2$  and  $\sigma_{\gamma}^2$  are the spatial partial sill and spatial nugget,  $\sigma_{\tau}^2$  and  $\sigma_{\eta}^2$  are the temporal partial sill and temporal nugget, and  $\sigma_{\nu}^2$  is spatiotemporal independent error variance parameter.

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

$$var(\mathbf{y}) \equiv \boldsymbol{\Sigma} = \sigma_{\delta}^2 \mathbf{Z}_{sp} \mathbf{R}_{sp} \mathbf{Z}_{sp}' + \sigma_{\gamma}^2 \mathbf{Z}_{sp} \mathbf{I}_{sp} \mathbf{Z}_{sp}' + \sigma_{\tau}^2 \mathbf{Z}_t \mathbf{R}_t \mathbf{Z}_t' + \sigma_{\eta}^2 \mathbf{Z}_t \mathbf{I}_t \mathbf{Z}_t' + \sigma_{\nu}^2 \mathbf{I}_N. \quad (3)$$

There are many common parameterizations of  $\mathbf{R}_{sp}$ , but one form is the exponential. For observations at locations  $i$  and  $i'$  at  $h_{ii'}$  distance apart, row  $i$  and column  $i'$  of  $\mathbf{R}_{sp}$  is equal to

$$\exp(-h_{ii'}/\phi), \quad (4)$$

where  $\phi$  is the range parameter.

There are also many common parameterizations of  $\mathbf{R}_t$ , but one form is the exponential, which is equivalent to an AR(1) process in time series if the time points are equally spaced and the correlation parameter is greater than zero. For observations at time points  $j$  and  $j'$  that are  $m_{jj'}$  units apart, row  $j$  and column  $j'$  of  $\mathbf{R}_t$  is equal to

$$\exp(-m_{jj'}/\rho), \quad (5)$$

where  $\rho$  is the autocorrelation parameter.

If we assume that  $\mathbf{y}$  is multivariate normal with mean  $\mathbf{X}\boldsymbol{\beta} \equiv \boldsymbol{\mu}$  and variance  $\boldsymbol{\Sigma}$ , then all parameters can be estimated with Maximum Likelihood or Restricted Maximum Likelihood.

Note that we will explore other covariance structures, including an extension of the sum-with-error model and different spatial and temporal correlation structures.

## 2.2. Finite Population Kriging

The model in equation 2 is used for all  $N$  observations at  $n_{sp}$  sites and  $n_t$  time points. But, we typically do not have the resources to sample every spatial site in every year.

Let the subscript  $s$  denote observations that were sampled (both past and present), and let the subscript  $u$  denote observations that were unsampled. Then, we can re-

order the response vector so that

$$\mathbf{y} = [\mathbf{y}'_u, \mathbf{y}'_s]'. \quad (6)$$

Let  $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}'_u, \tilde{\mathbf{y}}'_s]'$  denote the fixed, realized values of the response variable for one data-generating process. Our primary goal is to use the model developed in the previous section to predict values for  $\tilde{\mathbf{y}}_u$  from the observed data in  $\tilde{\mathbf{y}}_s$ . That is, we want to find optimal weights  $\mathbf{a}'$  to apply to the sampled data  $\mathbf{a}'\tilde{\mathbf{y}}_s$ , such that  $\mathbf{a}'\mathbf{y}_s$  is the Best Linear Unbiased Predictor (BLUP) for  $\mathbf{b}'_a\mathbf{y}_a$ . If we are interested in the total abundance across all years, then  $\mathbf{b}_a$  is a column vector of 1's. so that we are adding up all values of the response for our predictor of total abundance.

Unbiasedness implies that  $E(\mathbf{a}'\mathbf{y}_s) = E(\mathbf{b}'_a\mathbf{y}_a)$  for all  $\beta$ . So, denoting  $\mathbf{X}_s$  as the design matrix for sampled sites,  $\mathbf{a}'\mathbf{X}_s\beta = \mathbf{b}'_a\mathbf{X}_a\beta$  for every  $\beta$ , implying that  $\mathbf{a}'\mathbf{X}_s = \mathbf{b}'_a\mathbf{X}_a$ .

The kriging weights are then found by finding  $\lambda$  such that ]

$$E\{(\mathbf{a}'\mathbf{y}_s - \mathbf{b}'_a\mathbf{y}_a)(\mathbf{a}'\mathbf{y}_s - \mathbf{b}'_a\mathbf{z}_a)\} - E\{(\lambda'\mathbf{y}_s - \mathbf{b}'_a\mathbf{z}_a)(\lambda'\mathbf{y}_s - \mathbf{b}'_a\mathbf{z}_a)\} \quad (7)$$

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

$$\begin{pmatrix} \Sigma_{s,s} & \mathbf{X}_s \\ \mathbf{X}'_s & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ m \end{pmatrix} = \begin{pmatrix} \Sigma_{s,s} & \Sigma_{s,u} \\ \mathbf{X}'_s & \mathbf{X}'_u \end{pmatrix} \begin{pmatrix} \mathbf{b}_s \\ \mathbf{b}_u \end{pmatrix}, \quad (8)$$

where again the subscripts  $s$  and  $u$  denote sampled and unsampled observations. For example, letting  $n$  denote the number of sampled observations,  $\Sigma_{s,s}$  denotes the  $n \times n$  submatrix of  $\Sigma$  corresponding only to rows and columns of sampled observations and  $\Sigma_{u,s}$  denotes the  $(N - n) \times n$  submatrix of  $\Sigma$  corresponding to rows of observations that were not sampled and columns of observations that were sampled. Then,  $\lambda$  is an  $n \times 1$  vector.

Then, we can solve for the prediction weights as

$$\lambda_s = \mathbf{b}'_s + \mathbf{b}'_u(\Sigma_{u,s}\Sigma_{s,s}^{-1}) - \mathbf{b}'_u(\Sigma_{u,s}\Sigma_{s,s}^{-1})\mathbf{X}_s(\mathbf{X}'_s\Sigma_{s,s}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}'_s\Sigma_{s,s}^{-1} + \mathbf{b}'_u\mathbf{X}'_u(\mathbf{X}'_s\Sigma_{s,s}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s\Sigma_{s,s}^{-1}. \quad (9)$$

Our prediction for the total abundance across all years of the survey is then

$$\lambda'_s\tilde{\mathbf{y}}_s, \quad (10)$$

with a prediction variance of

$$E((\lambda'_s\mathbf{y}_s - \mathbf{b}'_a\mathbf{y}_a)(\lambda'_s\mathbf{y}_s - \mathbf{b}'_a\mathbf{y}_a)) = \lambda'_s\Sigma_{s,s}\lambda_s - 2\mathbf{b}'_a\Sigma_{a,s}\lambda_s + \mathbf{b}'_a\Sigma\mathbf{b}_a. \quad (11)$$

However, we often are not interested in predicting total abundance across multiple years and instead would want a prediction of the total abundance in the most recent year of the survey. Therefore, for this goal,  $\mathbf{b}_a$  should not be a vector of 1's and should instead take a value of 1 if the observation is in the current year of the survey and a 0 if the observation is not in the current year of the survey:

$$\mathbf{b}_a = [\mathbf{b}'_{up}, \mathbf{b}'_{uc}, \mathbf{b}'_{sp}, \mathbf{b}'_{sc}]' = [\mathbf{0}', \mathbf{1}', \mathbf{0}', \mathbf{1}']', \quad (12)$$

where the subscripts *up*, *uc*, *sp*, and *sc* denote unsampled sites in past years, unsampled sites in current years, sampled sites in past years, and sampled sites in current years, respectively.

$\lambda_s$  can then be rewritten as

$$\lambda_s = \mathbf{b}'_s + \mathbf{b}'_{uc}(\Sigma_{uc,s}\Sigma_{s,s}^{-1}) - \mathbf{b}'_{uc}(\Sigma_{uc,s}\Sigma_{s,s}^{-1})\mathbf{X}_s(\mathbf{X}'_s\Sigma_{s,s}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}'_s\Sigma_{s,s}^{-1} + \mathbf{b}'_{uc}\mathbf{X}'_{uc}(\mathbf{X}'_s\Sigma_{s,s}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s\Sigma_{s,s}^{-1}. \quad (13)$$

The prediction variance can then be rewritten as

$$\lambda'_s\Sigma_{s,s}\lambda_s - 2\mathbf{b}'_c\Sigma_{c,s}\lambda_s + \mathbf{b}'_c\Sigma_{c,c}\mathbf{b}_c, \quad (14)$$

where *c* denotes observations in the current year.

### 3. Application

Abundance surveys are performed in the TOK region of Alaska annually. In particular, surveys were conducted every year from 2014 through 2020, except for the year 2016, during which there was not sufficient snow cover to perform a survey. Before each survey, the sites are stratified into a High stratum and a Low stratum. In the high stratum, there are 230 unique spatial locations in the sampling frame of each year while in the low stratum, there are 151 unique spatial locations. For both strata, there are 7 unique time points, including the year 2016.

The strata are fit separately with a sum-with-error covariance, an exponential spatial correlation structure, and an AR(1) temporal correlation structure outlined in 2. The predicted total abundance is 2811 moose with a 95% prediction interval of (2683, 2940) moose.

Moose surveys in the TOC region were historically analyzed without explicitly using any data from surveys in past years. Therefore, we compare the spatiotemporal prediction and prediction interval with a spatial model fit with the `sptotal` package using only the data from the year 2020. The prediction total abundance is 2888 moose with a 95% prediction interval of (2306, 3469) moose. We see that the predictions are somewhat similar, but that, because the strictly spatial analysis ignores information from past years, the prediction interval for the spatiotemporal analysis is more narrow.

### 4. Simulation

- possibly include `sptotal` (on current year only) and SRS (on current year only) as reference comparisons.

For a preliminary simulation, we use a grid of 100 spatial sites and 5 time points. The spatiotemporal process is simulated as a sum-with-error model with an exponential spatial correlation structure and an exponential temporal correlation structure with the following parameters

The mean is

- $\beta = 10$ .

The spatial parameters are

- $\sigma_\delta^2 = 0.9$ ,
- $\sigma_\gamma^2 = 0.1$ ,
- $\phi = 5$ .

The temporal parameters are

- $\sigma_\tau^2 = 0.7$ ,
- $\sigma_\eta^2 = 0.3$ ,
- $\rho = 0.8$ .

And, the spatiotemporal nugget is

- $\sigma_\nu^2 = 0.4$ .

The sample size  $n$  is 100 (of the 500 total data points). For 100 iterations, the percentage of 90% prediction intervals that covered the true total was 81%.

There are a few plausible reasons for this low coverage:

- there is something incorrect about the method or code used.
- the sample size is only 100 sites, meaning that, on average, only 20 sites get selected per year. This small sample size may mean that the 8 parameters cannot be estimated accurately.
- only 100 simulations were done.

Hampering investigation of this is the fact that the simulations take a long time to run. The `stmodel` package will be very helpful for this, as we can replace the slow code for fitting the spatiotemporal model with faster code from the package.

## 5. Discussion

- mention substantial reduction of se in the application (and, presumably, the simulations).
- mention normal-based-related limitations
- mention Bayesian approach, and its potential flaws
- mention possible extension to imperfect detection
- mention forecasting potential
- take-home message: monitoring programs that use regularly surveys might consider incorporating time into their analysis to improve precision of predictors (e.g. NARS for lake assessments).

## 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.
- 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, Lisa Madsen, and Andy Aderman. 2021. "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.
- 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.
- Ver Hoef, Jay M. 2008. "Spatial methods for plot-based sampling of wildlife populations." *Environmental and Ecological Statistics* 15 (1): 3–13.