

Spatiotemporal Abundance Prediction

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 use a spatial statistical model to find a prediction for the total abundance.

Though these surveys are annual, each survey is analysed completely independently of surveys from previous years. For example, a survey conducted in the year 2019 only uses counts on sites that were sampled in that year. All data from previous years is ignored by the model.

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 whose sites are surveyed annually.

Model Development

The development of the following model combines standard spatiotemporal kriging with spatial finite population block kriging.

Spatiotemporal Model Let $Y(\mathbf{s}_i, \mathbf{t}_j)$, $i = 1 \dots n_{sp}$ and $j = 1 \dots n_t$ be a latent, unobserved spatiotemporal surface with n_{sp} as the number of spatial sites and n_t as the number of time points. If every spatial location is observed at every time point, then the total number of data points is $n_{sp}n_t \equiv N$. The underlying spatiotemporal surface is

$$Y(\mathbf{s}_i, \mathbf{t}_j) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}(\mathbf{s}_i, \mathbf{t}_j), \quad (1)$$

where \mathbf{s}_i references the i^{th} spatial location, \mathbf{t}_j references the j^{th} time point, \mathbf{X} is a design matrix for fixed effects, and $\boldsymbol{\epsilon}$ is a parameter vector of fixed effects.

The error $\boldsymbol{\epsilon}(\mathbf{s}_i, \mathbf{t}_j)$ 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 $\mathbf{y}(\mathbf{s}_i, \mathbf{t}_j)$ is

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

\mathbf{Z}_{sp} is an $N \times n_{sp}$ matrix where the values in a general row k in \mathbf{Z}_{sp} at 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 general row k 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

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

where \mathbf{R}_{sp} is a spatial correlation matrix and \mathbf{R}_t is a temporal correlation matrix. σ_{δ}^2 and σ_{γ}^2 are the spatial partial sill and spatial nugget, σ_{τ}^2 and σ_{η}^2 are the temporal partial sill and temporal nugget, and σ_{ν}^2 is spatiotemporal independent error variance.

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

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

A common form of \mathbf{R}_{sp} is exponential. For observations at locations i and i' at $h_{ii'}$ distance apart,

$$\mathbf{R}_{sp,ii'} = \exp(-h_{ii'}/\phi), \quad (4)$$

where ϕ is the range parameter.

A common form of \mathbf{R}_t is also exponential, or, AR(1). For observations at time points j and j' that are $m_{jj'}$ units apart,

$$\mathbf{R}_{t,jj'} = \exp(-m_{jj'}/\rho), \quad (5)$$

where ρ 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.

Finite Population Block Kriging The above model for \mathbf{Y} is a model for N observations at n_{sp} sites and n_t time points. But, we typically do not have the resources to sample every 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. Let the subscript c denote observations in the current year that we are interested in predicting, and let the subscript p denote observations in the past that we are not interested in predicting. Finally, let the subscript a denote all of the observations. Then we can partition \mathbf{y}_a into

$$\mathbf{y}_a = [\mathbf{y}_{up}, \mathbf{y}_{sp}, \mathbf{y}_{uc}, \mathbf{y}_{sc}]'.$$

so that \mathbf{y}_{up} is the response vector at unsampled sites in the past, \mathbf{y}_{sp} is the response vector at sampled sites in the past, etc.

Our primary goal is to use the model developed in the previous section to predict values for $\tilde{\mathbf{y}}_{uc}$ from the observed data in $\tilde{\mathbf{y}}_{sp}$ and $\tilde{\mathbf{y}}_{sc}$, where $\tilde{\mathbf{y}}_{\mathbf{a}}$ denotes the realized values from the spatial process generating \mathbf{y}_a .

Additionally, we are commonly interested in predicting the total abundance for the most recent year in the set of surveys. That is, we want to find optimal weights \mathbf{a}' to apply to the sampled data $\mathbf{a}'\tilde{\mathbf{y}}_s$, where $\tilde{\mathbf{y}}_s \equiv [\tilde{\mathbf{y}}_{sp}, \tilde{\mathbf{y}}_{sc}]'$, such that $\mathbf{a}'\tilde{\mathbf{y}}_s$ is the Best Linear Unbiased Predictor (BLUP) for $\mathbf{b}'_c\tilde{\mathbf{y}}_c$. If we are interested in the total abundance in the current year, then \mathbf{b}_c is a vector of 1's, so that we are adding up all values of $\tilde{\mathbf{y}}_c$.

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

The kriging weights are then found by finding λ such that

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

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

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

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 Σ corresponding only to rows and columns of sampled observations and $\Sigma_{u,s}$ denotes the $(N - n) \times n$ submatrix of Σ corresponding to rows of observations that were not sampled and columns of observations that were sampled.

Then, λ is an n by 1 vector. If we are interested in predicting the abundance in the most current year, the vector \mathbf{b}_s is a vector of 1's and 0's, with 1's for the sampled sites in the current year and 0's for the sampled sites in the past years. The vector \mathbf{b}_u is also a vector of 1's and 0's, with 1's for the unsampled sites in the current year and 0's for the unsampled sites in the past years.

Then, we can solve for the prediction weights as

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

Our prediction is then:

$$\lambda_s'\tilde{\mathbf{y}}_s,$$

which is equivalent to

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

where $\hat{\mathbf{y}}_u = \Sigma_{u,s}\Sigma_{s,s}^{-1}(\mathbf{y}_s - \hat{\boldsymbol{\mu}}_s) + \hat{\boldsymbol{\mu}}_u$ with $\hat{\boldsymbol{\mu}}_s = \mathbf{X}_s\hat{\boldsymbol{\beta}}_{GLS}$, $\hat{\boldsymbol{\mu}}_u = \mathbf{X}_u\hat{\boldsymbol{\beta}}_{GLS}$. $\hat{\boldsymbol{\beta}}_{GLS}$ is the generalized least squares estimator $(\mathbf{X}_s'\Sigma_{s,s}^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\Sigma_{s,s}^{-1}\mathbf{y}_s$.

The prediction has a prediction variance

$$E((\lambda_s'\mathbf{y}_s - \mathbf{b}'_a\mathbf{y}_a)(\lambda_s'\mathbf{y}_s - \mathbf{b}'_a\mathbf{y}_a)) =$$

$$\lambda_s'\Sigma_{s,s}\lambda_s - 2\mathbf{b}'_a\Sigma_{a,s}\lambda_s + \mathbf{b}'_a\Sigma\mathbf{b}_a,$$

Note that, again, if we are only interested in the total abundance in the current year, then unsampled sites in the past do not contribute to λ_s nor to the prediction variance because \mathbf{b}_{up} is $\mathbf{0}$. Therefore, λ_s can be rewritten as

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

Additionally, sampled sites in the past do not contribute to the prediction variance to the second and third terms in the prediction variance because \mathbf{b}_{sp} is $\mathbf{0}$. The prediction variance can then be rewritten as

$$\boldsymbol{\lambda}_s' \boldsymbol{\Sigma}_{s,s} \boldsymbol{\lambda}_s - 2 \mathbf{b}_c' \boldsymbol{\Sigma}_{c,s} \boldsymbol{\lambda}_s + \mathbf{b}_c' \boldsymbol{\Sigma}_{c,c} \mathbf{b}_c,$$

Move this: Finally, our site-by-site predictions for the unsampled sites come from the usual block kriging formulae:

$$\hat{\mathbf{z}}_u = \boldsymbol{\Sigma}_{u,s} \boldsymbol{\Sigma}_{s,s}^{-1} (\mathbf{z}_s - \hat{\boldsymbol{\mu}}_s) + \hat{\boldsymbol{\mu}}_u,$$